Emulating the first billion years of the Universe

Classe di Scienze
Scuola Normale Superiore
Pisa, a.a. 2kx/2kx

Daniela Breitman
Rel.: Prof. Andrei Mesinger

Abstract

The first detection of the 21-cm signal is now within reach with the upcoming Square Kilometre Array (SKA). The SKA is expected to deliver tomographic maps of the 21-cm signal across more than half of the observable Universe, providing transformative constraints on the astrophysical and cosmological processes that shaped its evolution. At present, however, such measurements remain out of reach, primarily due to challenges in data analysis and interpretation. Instead, we only have upper limits on the 21-cm power spectrum, reported by several SKA precursors. In this thesis, I employ machine learning to facilitate and enhance the interpretation of both current and upcoming cosmic dawn and epoch of reionisation observations within a Bayesian inference framework. Current upper limits on the 21-cm power spectrum become informative only when combined with complementary observations. To enable such synergistic analyses, I developed a machine-learning emulator of six summary statistics, which accelerates inference by more than four orders of magnitude compared to direct simulation. Despite the increasing realism of simulations, current Bayesian inferences of the 21-cm power spectrum remain limited by the small simulation volumes required for computational feasibility. Small box sizes restrict the number of large-scale modes that can be probed, setting a fundamental lower bound on the precision of forward-modelled power spectra at large scales — a limitation known as sample variance. To address this limitation, I developed a simulator-independent framework that mitigates sample variance with a score-based diffusion model and thereby significantly improves the constraining power of the inference. Put together, the work presented in this thesis is aimed at squeezing the most juice out of current upper limits and upcoming first detections of the 21-cm power spectrum, thereby extracting the maximum possible amount of information out of them.

Contents

Chapter 1

Introduction

1.1 An overview

Until just about a century ago, we believed that Newtonian mechanics provided a complete and accurate description of gravity. In Newtonian mechanics, masses exert instantaneous forces on one another. Newtonian physics successfully explained many phenomena such as the orbits of planets, tides, and even falling apples for more than two centuries. As our ambitions stretched to precisely measure ever larger volumes of the Universe, Newtonian gravity began to show its failures. Phenomena such as the anomalous precession of Mercury's orbit, and the incompatibility of Newton's instantaneous action-at-a-distance with the finite speed of light established by special relativity, hinted that a deeper theory was needed. Einstein's theory of general relativity revolutionised our way of thinking about the Universe. Instead of forces acting instantaneously at a distance, gravity becomes the curvature of spacetime caused by mass and energy. This profound relationship is elegantly encapsulated in the Einstein field equation,

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.1}$$

where the stress-energy tensor $T_{\mu\nu}$ describes the distribution of matter and energy, and the Einstein tensor $G_{\mu\nu}$ encodes the curvature of spacetime. It shows that spacetime tells matter how to move and matter tells spacetime how to curve. Einstein's theory of general relativity not only resolved Newtonian shortcomings, but also predicted entirely new phenomena.

One of the earliest predictions of Einstein's theories of relativity was the bending of light around massive objects. It was confirmed observationally for the first time in 1919 during a

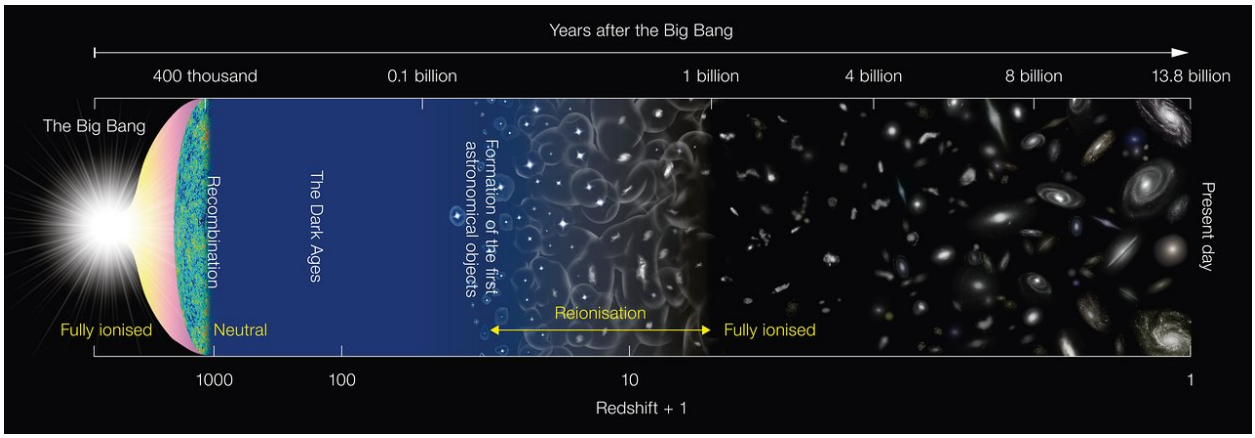

Figure 1.1: History of the Universe. Image credit: NAOJ.

solar eclipse, offering strong support for his theories. Many other predictions followed, some of which have only been confirmed just recently. Gravitational waves, predicted in 1916, were first detected a whole century later by LIGO in 2015. Black holes, hypothesized as early as the 1910s, were directly imaged for the first time by the Event Horizon Telescope (EHT) in 2019. A century on, Einstein's theory continues to prove both remarkably accurate and astonishingly visionary.

In spite of all the answers it provided, Einstein's theory of general relativity also raised many questions. For instance, Einstein realised that a Universe composed only of matter would eventually expand or collapse. To preserve the idea of a static Universe, Einstein introduces the cosmological constant Λ to counteract gravity. The notion of a static Universe, however, was soon overturned. In the 1910s, Vesto Slipher measured spectral lines from nearby galaxies. He finds that most were redshifted i.e. shifted to longer wavelengths, indicating that they were moving away from us. A decade later, Edwin Hubble extended these measurements and discovered a linear relation between the recession velocity of a galaxy v and its distance d, now known as Hubble's law: $v = Hd$. Hubble's law established the expansion of the Universe as a cornerstone of modern cosmology. In an expanding Universe, distances between galaxies grow with the cosmic scale factor $a(t)$. Instead of using physical distances or *proper* distances that change with time t, cosmologists often adopt *comoving* coordinates, which factor out the expansion of the Universe. Instead of time, cosmologists also often adopt the redshift $z = 1/(1 + a)$.

Many decades later, observations of distant type Ia supernovae at higher redshifts revealed an even more surprising fact: rather than slowing down under the pull of gravity, the expansion of the Universe is accelerating. To account for this unexpected behaviour, cosmologists revived Einstein's cosmological constant Λ, now interpreted as a mysterious form of energy inherent to space itself dubbed *dark energy*. Since its discovery in 1998, dark energy has emerged as the dominant component of the Universe, comprising about 70% of the cosmic energy budget, though its true nature remains one of the greatest mysteries in modern physics.

In 1965, Penzias and Wilson serendipitously discovered a faint background of microwave radiation. Subsequent measurements of this cosmic microwave background (CMB) by satellites such as COBE, WMAP, and Planck revealed that it is remarkably uniform, with tiny temperature fluctuations of 1 part in 100 000.

The discovery of cosmic expansion led to the idea that the Universe must have expanded from a very hot and dense state, later dubbed the Big Bang. The simplest Big Bang models, however, could not explain key observations such as the near-uniformity of the CMB and its tiny fluctuations. To solve these issues, in the early 1980s, Alan Guth and collaborators proposed the theory of cosmological inflation, which suggests a brief period of rapid expansion in the first fraction of a second following the Big Bang. After the inflationary period, the Universe continued to cool and expand more slowly, until about 380 000 years after the Big Bang, when it became cold enough for protons and electrons to combine into neutral hydrogen in an event known as *recombination*. With free electrons no longer scattering photons, the Universe became transparent, allowing photons to travel largely unimpeded. Stretched by cosmic expansion into the microwaves, these photons form the CMB which provides a direct glimpse into the Universe as it was only 380 000 years after the Big Bang.

Alongside dark energy, which drives the accelerated expansion of the Universe, another invisible component that has played an equally crucial role in shaping the Universe is dark matter. In the 1930s, Fritz Zwicky studied the motion of galaxies in the Coma cluster, and inferred the visible mass of the cluster to be too small to gravitationally bind the cluster. Decades later, Vera Rubin measured galactic rotation curves and found that stars on the outskirts of galaxies orbit far too quickly to be explained the gravitational pull of luminous matter alone. These and many other observations point to the existence of an invisible type of matter, *dark matter*, that out-

weighs ordinary, *baryonic matter* by roughly a factor of five and provides the dominant source of gravitational pull within bound structures.

The discovery of dark matter led to the development of the theory of structure formation, in which dark matter plays a central role. Tiny primordial density fluctuations imprinted in the CMB, evolve under gravity evolve until they eventually collapse into gravitationally bound structures known as *dark matter halos*. These dark matter halos form a vast cosmic web of filaments and voids, known as the large-scale structure (LSS), act as the invisible gravitational skeleton of the Universe. Baryonic matter then accretes and cools onto this gravitational blueprint to ignite the first stars and galaxies about 100 million years after the Big Bang. This period, known as the *cosmic dawn* (CD), marks the end of the dark ages with the birth of the first luminous objects. As illustrated in Figure 1.1, structure formation suggests a hierarchical evolution where small halos form first and then merge to create progressively larger galaxies and clusters. This theory of structure formation has been strongly supported since the 1980s by galaxy surveys such as the Sloan Digital Sky Survey (SDSS), which mapped millions of galaxies across a large volume of the Universe. Such galaxy surveys trace the underlying dark matter distribution and reveal the filamentary cosmic web predicted by structure formation.

Observations of the local Universe reveal an intergalactic medium (IGM) that is fully ionised, whereas inflationary theory and CMB measurements indicate that the early Universe was completely neutral. Bridging these two epochs together suggests that the first stars and galaxies must have heated and ionised the surrounding neutral IGM during an epoch of reionisation (EoR). Figure 1.1 illustrates the EoR, showing how the neutral IGM (blue) was gradually populated by ionised (black) regions from ionising radiation produced by galaxies. The CD and EoR thus mark the last great phase transition of the Universe. They are also perhaps the most *solvable* of the major mysteries in modern cosmology given our current knowledge of the Universe and observational capabilities.

We aim to understand when and how the first stars and galaxies formed, which types of galaxies drove the EoR and how individual galaxies collectively reionised the entire IGM. Many ongoing observations target the CD/EoR to answer these questions. Increasingly informative probes of CD/EoR, however, lie at increasingly higher redshifts, which are correspondingly more challenging to observe. In 1944, Dutch astronomer Hendrik van de Hulst predicts the existence of a highly forbidden line transition with a wavelength of 21-cm that arises from the hyperfine splitting of the ground state of the neutral hydrogen atom: when the proton and electron spins are aligned, the energy of the atom is $E_{21} \equiv 5.9 \times 10^{-6}$ eV (or $E_{21}/k_B = 0.068$ K) higher than when they have opposite spins. The 21-cm line was first confirmed observationally in 1951 Ewen and Purcell. While the 21-cm transition is extremely rare for a single hydrogen atom, the vast reservoir of neutral hydrogen in the IGM over cosmic time *which includes most of the observable Universe* (from $\sim 0.1 - 4$ billion years after the Big Bang in Figure 1.1) makes the cumulative signal potentially detectable. The evolution of the 21-cm signal is commonly described in terms of the *differential brightness temperature*, δT_b, which quantifies the intensity of the 21-cm signal relative to the CMB background:

$$\delta T_{\rm b}(z) \approx 35\,(1+\delta) x_{\rm HI} \left(1 - \frac{T_\gamma}{T_{\rm S}}\right) \left(\frac{\Omega_b h^2}{0.023}\right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10}\right)^{1/2} \left(\frac{H(z)}{H(z) + {\rm d}v_r/{\rm d}r}\right) {\rm mK} \quad (1.2)$$

The remaining of the introduction is aimed at unpacking all the necessary tools to produce and understand this equation. In red, we have terms governed by cosmology. To understand them, in Section 1.2, we develop a theory for the evolution Universe composed of dark matter, baryonic matter, radiation, and dark energy, expanding with rate $H(z)$. In green, we have baryonic matter

overdensity. Since dark matter forms the cosmic web onto which baryonic matter accretes, in Section 1.3, we first develop a theory of structure formation that describes how dark matter particles evolve in an expanding Universe and collapse to form dark matter halos. In Section 1.3.4, we extend dark matter structure formation to baryonic matter and describe the processes governing the formation of the first stars and galaxies. In blue, we have terms describing the ionisation state and temperature of the IGM gas which we develop in Section 1.3.5. While the 21-cm line a *direct* probe of the amount of neutral hydrogen in the IGM across cosmic time (blue terms), it is also a valuable cosmological probe (red terms) during the cosmic dawn that has the power to constrain dark matter models. All of these ingredients come together in Section 1.4, where we describe CD/EoR observables, ending with this very equation for the 21-cm signal. The goal of this thesis is to improve the interpretation of modern CD/EoR observations, with a focus on the 21-cm line. The most robust way to interpret observations is with Bayesian inference, as we shall see in Section 1.5. In this thesis, we improve Bayesian analyses with the introduction of machine learning applications, as shall be introduced in Section 1.7.

1.2 Cosmology

Modern cosmological models are based on the *cosmological principle*, which states that, on sufficiently large scales ($\gtrsim 100$ Mpc), the Universe is: (i) *homogenous* - it looks the same from any location; and (ii) *isotropic* – it looks the same in every direction. Think about how these two properties are independent: a Universe could have a preferred direction and be homogenous without being isotropic, or it could be isotropic from a single point without being homogenous. However, if a space is isotropic at every point, then it must be homogeneous. More importantly, if space is isotropic at one point and homogeneous everywhere, then it will be isotropic everywhere. Observations of the CMB support the idea that the Universe is isotropic from Earth, while galaxy surveys reveal large-scale homogeneity. Together, these observations combine to support the assumption of a Universe that is both isotropic and homogeneous, implying a maximally symmetric Universe that greatly simplifies cosmological modelling. In this section, we study the evolution of such a maximally symmetric Universe, so the next step is to define how physical systems move in such a Universe.

From the principle of stationary action, also known as Hamilton's principle, we know that the path taken by a physical system when travelling between two points is the one that makes the action stationary, often corresponding to the shortest path between the two points. In flat Euclidean space, this path is just a straight line. Recall, however, that in Einstein's perspective, mass-energy *curves* space-time, and the shortest paths, called *geodesics*, are no longer straight lines, but curves determined by the geometry of space-time. Motion in curved space-time with curvature κ in a maximally-symmetric Universe under the cosmological principle is described by the Friedmann–Lemaître–Robertson–Walker (FLRW) metric:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-\kappa r^2} + r^2\mathrm{d}^2\Omega\right], \tag{1.3}$$

which describes the motion of a particle at (r, θ, ϕ) through a space-time that expands or contracts at time t with scale factor $a(t)$ in spherical co-ordinates, where $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$ corresponds to the angular separation between the two points defining the path of the particle through space-time.

Having established how particles move through curved space-time, we now turn to the question of how space-time itself evolves in the presence of matter and energy. The relation between

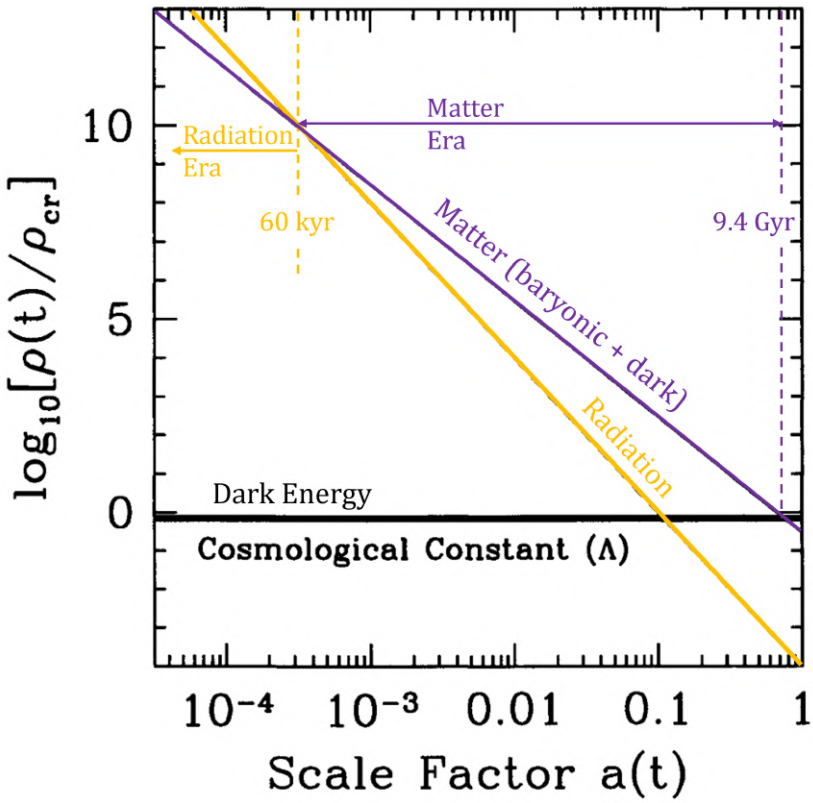

Figure 1.2: Density parameter Ω for radiation (yellow), matter (purple), and dark energy (black) as a function of the scale factor a. This plot shows that the matter era ended around 9.4 billion years after the Big Bang, and that today ($a = 1$), we live in a dark energy dominated Universe. Adapted from Dodelson and Schmidt 2020.

space-time and mass-energy is described by the Einstein field equation 1.1. Substituting the FLRW metric into the Einstein tensor $G_{\mu\nu}$ yields the *Friedmann equation*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho + \Lambda}{3} - \frac{\kappa}{a^2}, \tag{1.4}$$

where ρ is the energy density of the Universe, and Λ is the cosmological constant accounting for the accelerated expansion of the Universe. Looking at the left hand side, we can define the **Hubble parameter** $H(t) \equiv \frac{\dot{a}}{a}$. The present-day value of the Hubble parameter $H(t_0) \equiv H_0 \approx 0.67$ is known as the *Hubble constant* as its value has been inferred by Edwin Hubble in 1929 when he discovered the Hubble Law.

The Friedmann equation shows that the mass-energy density ρ is directly linked to the curvature of space-time κ. We can define the *critical density* required for a flat Universe with $\kappa = 0$: $\rho_c = \frac{3H^2}{8\pi G}$. It is convenient to express any component of the energy density ρ_x as a fraction of this critical density, defining the *density parameter* $\Omega_x \equiv \frac{\rho_x}{\rho_c}$. Introducing the density parameter into the first Friedmann equation yields:

$$\Omega - 1 = \Omega_\kappa, \tag{1.5}$$

where $\Omega_\kappa \equiv \frac{\kappa}{a^2H^2}$. This equation explicitly shows how the curvature of the Universe is determined by the total density parameter Ω: $\Omega = 1$ corresponds to a flat Universe, $\Omega > 1$ to a closed Universe, and $\Omega < 1$ to an open Universe. We can further break down the total density Ω into its main constituents: radiation r, matter (baryonic and dark) m, dark energy Λ. The Friedmann

equation then becomes $1 = \Omega + \Omega_\kappa = \Omega_r + \Omega_m + \Omega_\Lambda + \Omega_\kappa$, where $\Omega_\Lambda \equiv \frac{\rho_\Lambda}{\rho_c}$, where $\rho_\Lambda = \frac{\Lambda}{8\pi G}$. Assuming an equation of state for each component, we obtain:

$$\left(\frac{H}{H_0}\right)^2 = \frac{\Omega_{0,r}}{a^4} + \frac{\Omega_{0,m}}{a^3} + \frac{\Omega_{0,\kappa}}{a^2} + \Omega_{0,\Lambda} \tag{1.6}$$

$$\approx \frac{\Omega_{0,m}}{a^3} + \Omega_{0,\Lambda} \text{ for a flat Universe at late times,} \tag{1.7}$$

where today we observe $\Omega_{0,m} \approx 0.3$ and $\Omega_{0,\Lambda} \approx 0.7$, $\Omega_{0,r} \approx 0$, and $\Omega_{0,\kappa} \approx 0$. This equation tells us that in a flat Universe, at early times $a \to 0$, radiation is the dominant component, followed by the matter-dominated era, and concluding with the dark energy era. We illustrate these different eras in Figure 1.2.

1.3 Structure formation

In order for cosmic structure to grow, the Universe requires inhomogeneities in the primordial density field, since a perfectly uniform matter distribution would remain smooth under gravity. These primordial seeds are observed as anisotropies in the CMB temperature map, at the level of one part in 10^5. These fluctuations provide the necessary density contrasts that, amplified by gravity, evolved into the large-scale structure we observe today.

Matter in the Universe consists of two main components: baryonic matter and dark matter, with the latter outweighing the former by about a factor of five. Both components interact through gravity, but unlike baryons, dark matter does not interact with photons and thus remains invisible to electromagnetic observations. This property of dark matter has crucial consequences for structure formation: without it, we would not exist today! Unaffected by radiation, dark matter can begin collapsing into halos as soon as the Universe becomes matter-dominated at $z \sim 3400$. Baryonic matter, on the other hand, remains tightly coupled to radiation through scattering until recombination at $z \sim 1100$, after which it can finally fall into the potential wells created earlier by dark matter. This establishes structure formation as a two-step process: dark matter first provides the gravitational blueprint, and baryons subsequently amplify it to form the cosmic web we observe today.

The first step in structure formation is the gravitational clustering of dark matter particles. In Section 1.3.1, we develop a simple linear model appropriate for describing the growth of tiny perturbations such as those in the primordial overdensity field of the Universe, where we define the overdensity as $\delta = \frac{\rho - \overline{\rho}}{\overline{\rho}} \lesssim 1$ relative to the mean density $\overline{\rho}$. We then move to the non-linear regime in Section 1.3.2 necessary to accurately describe the evolution of larger perturbations of the matter density field $\delta \gtrsim 1$ at later times. We then proceed to develop a statistical description of halo populations through the *halo mass function*, which predicts the abundance of halos of different masses across cosmic time.

In Section 1.3.4, we describe how baryonic matter traces the underlying dark matter distribution, igniting the first stars and galaxies during the cosmic dawn. These first stars and galaxies subsequently drive the epoch of reionisation whose bulk properties are described in Section 1.3.5. By the end of this section, we will have painted a coherent picture of structure formation, from the growth of the earliest dark matter overdensities to the formation of the first galaxies, ending with the final phase transition of the Universe - the EoR.

1.3.1 Linear Evolution

During matter domination after $z \sim 3400$, the Universe is nearly uniform, with only small dark matter perturbations $\delta \ll 1$ seeded by the CMB. At this stage, gravity is too weak to cause collapse, and the perturbations grow only with the expansion of the Universe. In this section, we describe the linear growth of such small perturbations only affected by the expanding Universe. The linear approximation breaks down once $\delta \sim 1$, at which point a non-linear treatment is required to capture their evolution, as we shall see in Section 1.3.2. Throughout, we assume that dark matter particles behave as a cold (i.e. non-relativistic), collisionless and pressureless fluid. Under these assumptions, Newtonian theory is sufficient to describe the evolution of the density ρ and peculiar velocity $\mathbf{v}_p$ of the fluid under the influence of a gravitational field with potential ϕ.

Locally (i.e. in proper coordinates, denoted by subscript p), the dark matter fluid satisfies three fundamental relations:

- **Mass conservation:** $\frac{\partial \rho}{\partial t} = -\nabla_p \cdot (\rho\, \mathbf{v}_p)$, where $\mathbf{v}_p(\mathbf{r}, t)$ is the proper velocity of dark matter particles located at position $\mathbf{r}$ and time t.
- **Euler equation:** $\frac{\partial \mathbf{v}_p}{\partial t} + (\mathbf{v}_p \cdot \nabla_p)\mathbf{v}_p = -\nabla_p \phi$, where ϕ is the scalar gravitational potential field generated by the dark matter particles.
- **Poisson equation:** $\nabla_p^2 \phi = 4\pi G \rho$, where G is Newton's constant.

To understand the evolution of these perturbations in an expanding FLRW universe, it is more convenient to transform these three equations from proper to comoving coordinates, with the proper distance being related to comoving distance by $\mathbf{r}_p = a(t)\mathbf{V}$. From this definition, we can compute the corresponding proper velocity $\frac{\mathrm{d}\mathbf{r}_\mathrm{p}}{\mathrm{dt}} \equiv \dot{\mathbf{r}}_p$:

$$\mathbf{V} = \dot{\mathbf{R}} = a^{-1}\dot{\mathbf{r}}_p - \dot{a}\, a^{-2}\mathbf{r}_p \tag{1.8}$$

$$\mathbf{v}_p = \dot{\mathbf{r}}_p = \dot{a}\mathbf{R} + a\dot{\mathbf{R}} \equiv \dot{a}\mathbf{R} + \mathbf{v}_\mathrm{pec}, \tag{1.9}$$

where a is the scale factor, and $\mathbf{R}$ is the comoving distance. Applying this transformation to all three equations in proper coordinates above, we obtain them in comoving coordinates:

- **Conservation of mass:** $\dot{\delta} + a^{-1}\nabla \cdot ((1+\delta)\, \mathbf{v}_\mathrm{pec}) = 0$
- **Euler equation:** $\dot{\mathbf{v}}_\mathrm{pec} + H\, \mathbf{v}_\mathrm{pec} + a^{-1}\nabla\phi = 0$, where $H \equiv \frac{\dot{a}}{a}$ is the Hubble parameter.
- **Poisson equation:** $\nabla_p^2 \phi = 4\pi G \overline{\rho}_0 a^{-1}\delta$.

Keeping only first order terms in δ, we can combine these three equations into one:

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G \overline{\rho}_0 \delta. \tag{1.10}$$

This equation describes the evolution of cold, collisionless and pressureless dark matter overdensities δ sourced by gravity (RHS) and damped by the Hubble flow (second term on the LHS) in the linear regime. In the previous Section 1.2, we found that during matter domination $a \propto t^{2/3}$ and $\dot{a} \propto t^{-1/3}$. Substituting these into the equation above, we obtain:

$$\delta \propto C_1 t^{-1} + C_2 t^{2/3}. \tag{1.11}$$

Imposing initial conditions (ICs) with $\dot{\delta}(t_0) = 0$ and $\delta(t_0) = \delta_0$, we can solve for the constants and obtain $C_1 = \frac{2}{5}\delta_0 t_0$ and $C_2 = \frac{3}{5}\delta_0 t_0^{-2/3}$. Substituting these back in, we obtain:

$$\delta(t) = \frac{2}{5}\delta_0 \frac{t_0}{t} + \frac{3}{5}\delta_0 \left(\frac{t}{t_0}\right)^{2/3}. \tag{1.12}$$

The first term on the right-hand side (RHS), known as the *decaying mode*, dominates at very early times. It reflects the expansion of the Universe damping the growth of overdensities due to the rapid expanion of the Universe. The second term on the RHS, the *growing mode*, eventually takes over. It describes gravity amplifying the overdensities once expansion slows enough to allow for structure to form. The key result is that, during matter domination in the linear regime, overdensities grow with the scale factor:

$$\boxed{\delta \propto t^{2/3} \propto a} \tag{1.13}$$

1.3.2 Non-linear evolution

In the previous section, we used linear perturbation theory to show that, during the matter-dominated era, dark matter perturbations grow linearly with the scale factor as long as they remain small with $\delta \ll 1$. We now turn to the non-linear regime, where perturbations reach $\delta \gtrsim 1$ and collapse into gravitationally bound structures known as dark matter halos. In this regime, linear theory breaks down, and the evolution of the density field can no longer be captured by a simple growth rate. Since exact analytic solutions are generally unavailable, the standard approach is to follow the full non-linear dynamics with *N-body simulations*, in which the matter distribution is represented by a large number of particles whose trajectories evolve under mutual gravitational interactions. These simulations reveal the emergence of structure on scales ranging from individual halos to the cosmic web. Alongside them, analytic models based on idealised assumptions offer simplified but powerful insights into non-linear collapse. One of the most instructive is the *spherical collapse model*, later generalised by the ellipsoidal collapse model, which relaxes the assumption of spherical symmetry. In the following section, we focus on the spherical collapse model, which distils the complexity of non-linear structure formation into a tractable framework that provides physical intuition for the key processes driving non-linear structure formation.

1.3.2.1 Spherical collapse model

In this section, we examine the non-linear evolution of a single spherical overdensity embedded in an otherwise homogeneous universe closely following Mo et al. 2010. Picture the overdensity as an onion: it has a dense core of mass $M = \frac{4\pi r^3}{3}\overline{\rho}(1+\delta)$ surrounded by a thin spherical shell. We assume the motion of the mass shell is non-relativistic and can thus be treated with Newtonian mechanics. The evolution of the shell is governed by the conservation of energy per unit mass:

$$\mathrm{KE} + \mathrm{PE} + \mathrm{DE} = \mathrm{E_{tot}}, \tag{1.14}$$

where E_{tot} is constant for the shell. The first two terms represent the kinetic and gravitational potential energy, respectively, while the final term on the LHS stands for dark energy (DE). In a universe with a non-zero cosmological constant Λ, dark energy behaves as a uniform energy density with negative pressure. In Section 1.1, we found that $\rho_\Lambda = \frac{\Lambda}{8\pi G}$ and $p_\Lambda = -\rho_\Lambda$. The dark energy within a spherical volume of radius r has an effective mass $M_\Lambda = \rho_\Lambda \frac{4}{3}\pi r^3$, whcih

produces a repulsive force $F = -\frac{GM_\Lambda}{r^2} = -\frac{4\pi G\rho_\Lambda}{3}r$. This is equivalent to a harmonic oscillator with negative restoring force, yielding a potential energy contribution $\mathrm{DE} = \frac{1}{12}\Lambda \mathrm{r}^2$. Substituting the classical expressions for potential and kinetic energy, the conservation equation becomes:

$$\frac{1}{2}\dot{r}^2 - \frac{GM}{r} - \frac{\Lambda}{6}r^2 = E_{\mathrm{tot}}. \tag{1.15}$$

This equation shows that the fate of the perturbation depends on the competition between self-gravity and the repulsive effect of dark energy. If $E_{\mathrm{tot}} > 0$, the perturbation will expand forever, while if $E_{\mathrm{tot}} < 0$, it will eventually collapse. To determine the condition for collapse, we first need to evaluate the constant E_{tot}. At early times $t = t_i$, we assume the shell is simply following the Hubble expansion. Under this assumption, the initial kinetic energy of the shell is:

$$\mathrm{KE_i} = \frac{1}{2}\dot{\mathrm{r}}_\mathrm{i}^2 = \frac{1}{2}\left(\mathrm{H_i\, r_i}\right)^2. \tag{1.16}$$

We write the gravitational potential energy assuming the spherical overdensity has mass $M_i = \frac{4}{3}\pi r_i^3 \rho_i$. Substituting $\rho_i = \overline{\rho}_i(1+\delta_i)$, we obtain:

$$\mathrm{PE_i} = -\frac{\mathrm{GM_i}}{\mathrm{r_i}} = -\frac{\mathrm{G}}{\mathrm{r_i}}\frac{4}{3}\pi \mathrm{r}_\mathrm{i}^3 \overline{\rho}_\mathrm{i}(1+\delta_\mathrm{i}) = -\frac{1}{2}\Omega_\mathrm{i}\mathrm{r}_\mathrm{i}^2\mathrm{H}^2(1+\delta_\mathrm{i}), \tag{1.17}$$

where in the last step we used the definition of the density parameter $\Omega_i = \rho_i/\rho_c$. Finally, for the dark energy term:

$$\mathrm{DE_i} = -\frac{\Lambda}{6}\mathrm{r}_\mathrm{i}^2. \tag{1.18}$$

Combining kinetic, gravitational, and dark energy contributions, the total energy per unit mass of the shell is:

$$E_{\mathrm{tot}} = \frac{1}{2}H_i^2 r_i^2\left(1 - \Omega_i(1+\delta_i) - \frac{\Lambda}{12H^2}\right). \tag{1.19}$$

Since overdensities can collapse when $E_{\mathrm{tot}} < 0$, in order to collapse, an overdensity must satisfy:

$$\frac{\Lambda}{12H_i^2} + 1 < \Omega_i(1+\delta_i). \tag{1.20}$$

In the absence of dark energy, all overdensities eventually collapse as $t \to \infty$ in a flat or closed universe since $\Omega_i \approx \Omega_m \approx 1$ in the early matter-dominated universe. As expected, a non-zero cosmological constant raises the critical overdensity threshold required for collapse.

The time for an overdensity to reach its turnaround radius r_{turn}, where it stops expanding and begins to collapse, can be found by setting $\dot{r} = 0$ in Equation 1.15. At turnaround, the total energy per unit mass is $E_{\mathrm{tot}} = -\frac{GM}{r_{\mathrm{turn}}} - \frac{\Lambda}{6}r_{\mathrm{turn}}^2$ Solving equation 1.15 and introducing the dimensionless variables $\zeta \equiv \frac{\Lambda r_{\mathrm{turn}}^3}{6GM}$, $x = r/r_{\mathrm{turn}}$, we obtain a convenient integral expression for the time:

$$t = \frac{1}{H_0}\sqrt{\frac{\zeta}{\Omega_{\Lambda,0}}}\int_0^{r/r_{\mathrm{turn}}} \mathrm{dx}\left[\frac{1}{\mathrm{x}} - 1 + \zeta(\mathrm{x}^2 - 1)\right]^{-1/2}, \tag{1.21}$$

where $\Omega_{\Lambda,0} = \frac{\Lambda}{3H_0^2}$ as we had from Section 1.2. This equation allows us to compute the collapse times of spherical overdensities in a ΛCDM Universe, explicitly showing the effect of dark energy in delaying or preventing collapse. Integrating this equation from 0 to 1 by setting $r = r_{\mathrm{turn}}$), we solve for the time until the shell reaches its turnaround radius, t_{turn}:

$$t_{\mathrm{turn}} = \frac{1}{H_0}\sqrt{\frac{\zeta}{\Omega_{\Lambda,0}}}\int_0^{1} \mathrm{dx}\left[\frac{1}{\mathrm{x}} - 1 + \zeta(\mathrm{x}^2 - 1)\right]^{-1/2}. \tag{1.22}$$

After reaching the turnaround radius $r_{\rm turn}$ at time $t_{\rm turn}$, the spherical shell begins to collapse, and its evolution is then described by:

$$t = t_{\rm turn} + \frac{1}{H_0}\sqrt{\frac{\zeta}{\Omega_{\Lambda,0}}}\int_{r/r_{\rm turn}}^{1} {\rm dx}\left[1 - \frac{1}{\rm x} - \zeta({\rm x}^2 - 1)\right]^{-1/2}. \tag{1.23}$$

Approximating the integral for early times t_i, when $r_i \ll r_{\rm turn}$, the spherical shell grows as:

$$t_i = \frac{2}{3H_0}\sqrt{\frac{\zeta}{\Omega_{\Lambda,0}}}\left(\frac{r_i}{r_{\rm turn}}\right)^{3/2}\left[1 + \frac{3}{10}\frac{r_i}{r_{\rm turn}}(1+\zeta)\right]. \tag{1.24}$$

This expression can be further simplified for a flat Universe with $\Omega_{m,0} + \Omega_{\Lambda,0} = 1$ at early times $t_i \ll t_0$ by substituting $\Omega_{\Lambda,i}/\Omega_{m,i} = \Omega_{\Lambda,0}/\Omega_{m,0}(1+z_i)^{-3}$:

$$t_i = \frac{2}{3H_0}\sqrt{\frac{\Omega_{m,i}^{-1} - 1}{\Omega_{\Lambda,0}}} \tag{1.25}$$

Taking the limit $r_i/r_{\rm turn} \ll 1$ and keeping only first order terms, Equation 1.24 simplifies to:

$$r_i/r_{\rm turn} \approx \left(\frac{\Omega_{\Lambda,i}}{\Omega_{m,i}\zeta}\right)^{1/3}. \tag{1.26}$$

Substituting this back into Equation 1.24 by replacing the LHS with Equation 1.25 and the RHS $r_i/r_{\rm turn}$ with the approximation and keeping only first-order terms, we obtain a more accurate approximation for the initial-to-turnaround radius ratio:

$$\frac{r_i}{r_{\rm turn}} \approx \left(\frac{\Omega_{\Lambda,i}}{\Omega_{m,i}\zeta}\right)^{1/3}\left[1 - \frac{1}{5}(1+\zeta)\left(\frac{\Omega_{\Lambda,i}}{\Omega_{m,i}\zeta}\right)^{1/3}\right]. \tag{1.27}$$

Re-introducing $\zeta \equiv \frac{\Lambda r_{\rm turn}^3}{6GM}$ and using the expression for the mass of the spherical overdensity $M = \frac{4}{3}\pi r_i^3\rho_i = \frac{4}{3}\pi r_i^3(1+\delta_i)\Omega_{m,i}\rho_c$, we obtain an expression for the initial overdensity of the shell:

$$\delta_i = \frac{3}{5}(1+\zeta)\left(\frac{\Omega_{\Lambda,i}}{\Omega_{m,i}\zeta}\right)^{1/3}. \tag{1.28}$$

It is more convenient to write this expression in terms of the present-day overdensity δ_0, which is obtained by linearly extrapolating δ_i with the growth factor g and its corresponding scale factor a:

$$\delta_0 = \frac{a_0 g_0}{a_i g_i}\delta_i = \frac{3}{5}g_0(1+\zeta)\left(\frac{\Omega_{\Lambda,0}}{\Omega_{m,0}\zeta}\right)^{1/3}, \tag{1.29}$$

where we take $g_i \to 1$ for $a_i \ll a_0$ and $a_0 = 1$ by definition. Assuming that the spherical shell collapses at time $t_{\rm coll} = 2t_{\rm turn}$, we can relate δ_0 to the collapse time using the integral expression for $t_{\rm turn}$ from Equation 1.22 yielding the critical overdensity at the time of collapse:

$$\delta_c(t_{\rm coll}) = \frac{3}{5}g(t_{\rm coll})(1+\zeta)\left(\frac{\Omega_{\Lambda,t_{\rm coll}}}{\Omega_{m,t_{\rm coll}}\zeta}\right)^{1/3}, \tag{1.30}$$

which is often approximated by:

$$\delta_c(t_{\rm coll}) = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3}[\Omega_m(t_{\rm coll})]^{0.0055} \approx 1.686, \tag{1.31}$$

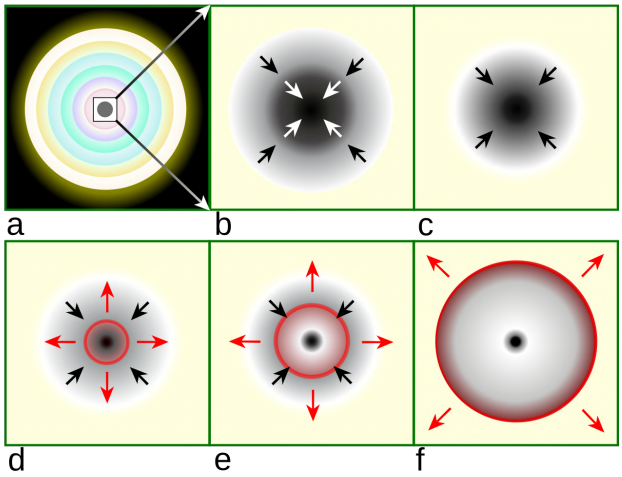

Figure 1.3: Illustration of the spherical collapse model. In plot a, we show the onion model of the spherical overdensity with a dense core surrounded by thin spherical shells. In plot b, we show the spherical shell collapsing toward the core. In plot c, we show that the dark matter particles slow down as they approach the centre of the overdensity but they do not actually concentrate into a single point. In plots d,e, and f, we show shell crossing as the dark matter particles fly past each other in red and expand again. They will continue oscillating about the centre of the overdensity, eventually virializing into gravitational equilibrium. Credit: R.J. Hall.

where in the second equality we drop the weak dependence on Ω_m. Equivalently, we can also express this equation as a function of $z_{\rm coll}$ at which the overdensity collapses to a point at the center of the shell and $\delta(z_{\rm coll}) \to \infty$:

$$1 + z_{\rm coll} = \frac{3}{5}\frac{\delta_i(1+z_i)}{1.686}, \tag{1.32}$$

where we assume that the growing mode dominates and use $\delta_c = \frac{3}{5}\delta_i a_{\rm coll}/a_i$ as we have obtained at the end of Section 1.3.1. This equation illustrates that perturbations require a very long time to collapse, on the order of the Hubble time. For instance, a large perturbation with $\delta_i = 0.1$ at $z_i = 100$ would collapse at $z_{\rm coll} = 2.6$, which corresponds to over 2 Gyr! Perturbations of order $10^{-5} - 10^{-6}$ as observed in the CMB around $z \sim 1000$ have not yet collapsed, and perhaps may never do so given that we are no longer in a matter-dominated universe, but in a dark energy-dominated one. This highlights the importance of starting simulations at very high redshift to allow sufficient time for structure growth. Dark matter decoupling from the CMB at $z \sim 3400$, much earlier than baryons at $z \sim 1000$, was absolutely essential as it allowed dark matter to form gravitational wells early, thus speeding up baryonic structure formation and ultimately enabling the formation of galaxies as well as our existence!

To avoid specifying both δ_i and z_i, we can re-write Equation 1.32 in terms of its linear extrapolation to the present-day overdensity $\delta_0 = \frac{3}{5}\delta_i(1+z_i)$:

$$1 + z_{\rm coll} = \frac{\delta_0}{1.686}. \tag{1.33}$$

In all of the derivations above, we assumed that the spherical shell collapses to a point at the centre of the overdensity at time $t_{\rm coll} = 2t_{\rm turn}$. While this approximation is valid during the early stages of collapse, in reality, however, the collisionless and pressureless nature of dark matter particles will have them experience *shell crossing* during which particles pass through each other and oscillate about the centre of the shell like a damped harmonic oscillator, rather

than concentrating into a single point. A great visualization of the spherical collapse model with shell crossing can be found here, but we also illustrate shell crossing in plots d, e, and f in Figure 1.3. This process is also referred to as *virialization*, as the dark matter halo reaches a gravitational equilibrium during this period of oscillation.

Although shell crossing is challenging to model analytically, we can estimate the final density of a virialized dark matter halo using the virial theorem, which states that for a system in equilibrium $\mathrm{KE} = -\frac{1}{2}\mathrm{PE}$. At the turnaround, we have $\mathrm{E_{tot}} = \mathrm{PE_{turn}} = -\frac{\mathrm{GM}}{\mathrm{r_{turn}}}$. Applying conservation of energy with the virial theorem, we obtain:

$$2\mathrm{KE_{vir}} + \mathrm{PE_{vir}} = \mathrm{E_{tot}} \tag{1.34}$$

$$-\frac{1}{2}\frac{GM}{r_{\mathrm{vir}}} = -\frac{GM}{r_{\mathrm{turn}}} \tag{1.35}$$

$$r_{\mathrm{vir}} = \frac{1}{2} r_{\mathrm{turn}}. \tag{1.36}$$

This relationship between the turnaround and virial radii implies that the virial density is $\rho_{\mathrm{vir}} = 2^3 \rho_{\mathrm{turn}}$. Using Equation 1.29 to estimate δ_{turn} and Equation 1.32 to relate to the collapse redshift, we can write the turnaround density as $\rho_{\mathrm{turn}} = (1 + \delta_{\mathrm{turn}})\overline{\rho}_0(1 + z_{\mathrm{turn}})^3$. This yields the final virial density:

$$\rho_{\mathrm{vir}} = 18\pi^2 \overline{\rho}_{z_{\mathrm{coll}}}. \tag{1.37}$$

This equation implies that a virialised halo has a density $18\pi^2 \approx 180$ times higher than its background density at the time of collapse.

The spherical collapse model may be simple, but it provides valuable insight into the evolution of a collapsing mass shell. The shell initially expands with the Hubble flow. Its expansion eventually slows as it reaches a maximum radius r_{turn} after which it begins to collapse. The mass shell is then said to have collapsed when its overdensity reaches a linearly-extrapolated value of 1.686. In reality, however, the shell does not collapse to a single point, and shell crossing begins sometime before the collapse time. Our simple calculations do not describe shell crossing. Instead, we study the spherical shell as it virializes, reaching a gravitational equilibrium during shell crossing. A virialised structure is roughly half its maximal size and has a density about 180 times higher than the background density at the time of collapse.

While the spherical collapse model offers valuable intuition for the formation of a *single* dark matter halo, the universe contains countless such perturbations. Extending this simple model to multiple independent perturbations provides some insight, but it misses the bigger picture: the goal is not only to describe individual structures, but rather to uncover the statistical laws governing structure formation in the Universe.

1.3.3 Statistics of collapsed structures

In this section, we focus on the statistical properties of dark matter halos. The most direct description of a halo field would be to specify the overdensity $\delta(\boldsymbol{x})$ at every point in real space, or equivalently its Fourier modes $\delta_{\boldsymbol{k}}$. This, however, is impractical: $\boldsymbol{x}$ is continuous and unbounded, so specifying the full density field would require infinitely many numbers. Fortunately, such a detailed description is unnecessary if we think of the Universe in statistical terms.

A useful analogy is rolling a dice. We do not care about the outcome of a particular throw—say, rolling a 2—but about the statistical law that governs the process, i.e. $P(2) = 1/6$. Our Universe is like that single roll: one realisation of an underlying random process. Instead of reconstructing the exact cosmic density field $\delta(\boldsymbol{x})$, we aim to describe the statistical process capable of generating such a field.

CMB observations suggest that the primordial density field of the Universe is well-approximated by a homogeneous and isotropic Gaussian random field. In statistical terms, a Gaussian random field is fully specified by its mean and covariance:

$$\langle \delta_{\boldsymbol{k}} \rangle = 0, \langle \delta_{\boldsymbol{k}} \delta_{\boldsymbol{j}} \rangle \equiv \Sigma_{\boldsymbol{kj}}, \tag{1.38}$$

where the first property comes from the definition of the overdensity, and the second defines the covariance matrix. Since our Universe is homogeneous (translational invariance), the covariance matrix in real space can only depend on the distance between two points, rather than the position of each point, i.e.:

$$\langle \delta_{\boldsymbol{p}_1} \delta_{\boldsymbol{p}_2} \rangle = \langle \delta_{\boldsymbol{p}_1} \delta_{\boldsymbol{p}_1 + \boldsymbol{x}_{12}} \rangle = \langle \delta^2_{\boldsymbol{x}_{12}} \rangle, \tag{1.39}$$

where $\boldsymbol{x}_{12}$ is the distance between points $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. Isotropy (rotation invariance) on the other hand implies that the covariance matrix cannot depend on $\boldsymbol{x}_{12}$ i.e. a vector pointing in some direction, but only on its magnitude $|\boldsymbol{x}_{12}| \equiv x$. Translating these two facts to Fourier space makes the covariance matrix diagonal due to homogeneity and only a function of the Fourier magnitude $|\boldsymbol{k}| \equiv k$:

$$\langle |\delta_k|^2 \rangle = \sigma_k^2, \langle \delta_k \delta_j \rangle = 0 \text{ for } k \neq j. \tag{1.40}$$

Let us now take the real space covariance matrix from above and write it using standard (inverse) Fourier transform convention and substituting equation 1.39:

$$\langle \delta_x^2 \rangle = \frac{V^{-1}}{(2\pi)^3} \int \mathrm{d^3k} \, \langle \delta_\mathrm{k} \delta_\mathrm{p}^* \rangle \mathrm{e}^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{p})\cdot\boldsymbol{x}} \tag{1.41}$$

$$= \frac{V^{-1}}{(2\pi)^3} \int \mathrm{d^3k} \, \sigma_\mathrm{k}^2 \tag{1.42}$$

$$= \frac{V^{-1}}{(2\pi)^3} \int \mathrm{dk} \, 4\pi \mathrm{k}^2 \sigma_\mathrm{k}^2 \tag{1.43}$$

$$= \frac{V^{-1}}{2\pi^2} \int \frac{\mathrm{dk}}{k} \, k^3 \sigma_k^2 \tag{1.44}$$

$$\equiv \int \mathrm{d} \log \mathrm{k} \, \Delta_\mathrm{k}^2, \tag{1.45}$$

where the third equality is obtained by integrating over all directions in spherical coordinates. The last two steps define Δ_k^2, the *spherically-averaged matter power spectrum* (PS) per logarithmic k-bin. This derivation shows that a Gaussian random field is completely specified by its PS. It's a ubiquitous quantity and we shall see it many times throughout the thesis. Note that intrinsically Δ_k^2 has the same dimensionality as the real-space field and thus is 3D quantity when calculated over a density field in real space.

While the power spectrum is defined at all scales, including arbitrarily small ones, in practice, we are only interested in computing it at some scales of interest such as the scales relevant to structure formation. We filter out the unwanted small scales by smoothing the density field to retain only scales above some threshold by applying a *window function* that acts as a low-pass filter.

A typical window function, at a given scale, will be associated with a volume and a corresponding mass. The most common window function is a spherical top-hat in real space with enclosed mass $M = V\overline{\rho} = \frac{4}{3}\pi R^3 \overline{\rho}$ at a given scale R. We can re-write our previous definition of the power spectrum now including the window function which gives the power spectrum probing scales larger than a scale of interest R:

$$\sigma_M^2(R) = \langle \delta_M^2 \rangle = \int \mathrm{d} \log \mathrm{k} \, \Delta_\mathrm{k}^2 \mathrm{W}_\mathrm{k}^2, \tag{1.46}$$

where W_k is the Fourier transform of the window function. $\sigma_M^2(R)$, as we shall see shortly, plays a central role in the statistical description of structure formation.

In the early Universe, where the density field evolves linearly, the variance of perturbations on scale k at redshift z can be expressed in terms of the growing mode $D(z,k)$ [Eisenstein and Hu, 1999]:

$$\sigma_k^2 \propto k^n T^2(k) D^2(z,k), \tag{1.47}$$

where n is the spectral index of the primordial power spectrum and $T(k)$ is the *transfer function*. The transfer function accounts for the suppression of growth at smaller scales due to processes such as radiation pressure, baryon acoustic oscillations, and the transition from radiation to matter domination. The growth factor $D(z,k)$ describes the linear growth of perturbations during the matter-dominated era and is normalised to be unit at present-time. Together, these bridge the gap between the nearly scale-invariant primordial fluctuations predicted by inflation and the matter power spectrum observed at later cosmic times.

1.3.3.1 Press-Schechter formalism

The power spectrum or equivalently the variance offer a convenient way to describe the density field in a statistical sense. In this section, we use the variance as a tool to study the statistics of dark matter halos. We start by writing the probability density function (PDF) of finding an overdensity of mass M and corresponding scale R at redshift z:

$$P(\delta_M, z) = \frac{1}{\sqrt{2\pi\sigma_M^2}} \exp\left(-\frac{\delta_M^2}{2\sigma_M^2}\right) \tag{1.48}$$

In section bla, we used the spherical collapse model to find that regions in the linear density field with density above the critical density $\delta_c(z) \approx 1.686$ have collapsed into a virialized dark matter halo. We are interested in determining the *collapsed fraction* which is the fraction of dark matter in the Universe that resides in collapsed structures of mass greater than M_h at redshift z:

$$f_{\rm coll}(>M_h, z) = \int_{1.686}^{\infty} P(\delta_M, z)\, {\rm d}\delta_{\rm M} = \frac{1}{2}{\rm erfc}\left(\frac{\delta_{\rm c}({\rm z})}{\sqrt{2}\sigma_{\rm M}({\rm M_h, z})}\right), \tag{1.49}$$

where we obtain the collapsed fraction by integrating the PDF from the previous equation over all overdensities above the critical overdensity 1.686. Here, erfc refers to the complementary error function. When Press and Schechter first derived this equation in 1974, they noticed a puzzling issue: the collapsed fraction should approach unity as $M_h \to 0$, but instead the formalism gives $f_{\rm coll}(>M_h, z) \to \frac{1}{2}$. They corrected this by inserting a fudge factor of two. This discrepancy, however, is not merely a mathematical quirk—it reflects a deeper physical issue that later became known as the cloud-in-cloud problem, illustrated in Figure 1.4.

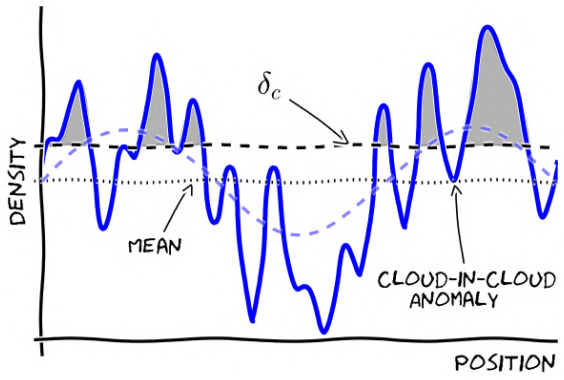

Figure 1.4: Illustration of the cloud-in-cloud problem: three smaller halos with density above / below the critical density that are actually part of one larger halo with density above the critical density. Figure from Murray 2017.

A useful analogy is to imagine a small rocky island protruding above the ocean surface. Diving below, one sees that the island is just the visible tip of a much larger underwater mountain. The island is not an isolated structure, but part of a larger whole. Similarly, an

overdensity identified on small scales may actually be embedded within a larger-scale overdensity that itself collapses into a halo. Conversely, underdense regions may lie within a larger overdense region that ultimately collapses. The original Press–Schechter formalism double-counts or misses such contributions, explaining the need for the missing factor of two.

Recognizing this issue was pivotal: it highlighted that properly describing halo statistics requires tracking how overdensities evolve across different scales. This realization directly motivated the excursion set formalism [Bond et al., 1991], which provided a more rigorous framework for resolving the cloud-in-cloud problem.

The key idea is to fix a point in space x_1 and vary the smoothing scale M_h around it. At very large scales ($M_h \to \infty$), the smoothed density field is completely uniform, and its variance is zero by construction. As we decrease the smoothing scale, the density contrast δ_M fluctuates up or down with equal probability, producing a stochastic *random walk* because the underlying field is Gaussian. Repeating this procedure for many spatial points yields an ensemble of such random walks — *the excursion set*. The solution to the cloud-in-cloud problem comes from counting only the first up-crossing of the collapse threshold δ_c: the first time a trajectory becomes dense enough to collapse as the smoothing scale is decreased. This "first-crossing" condition ensures that small-scale overdensities embedded in larger collapsed regions are not double-counted. By treating halo formation as a stochastic process governed by first-crossing statistics, the excursion set formalism naturally recovers the Press–Schechter mass function — now with the missing factor of two accounted for.

Differentiating the *corrected* Press–Schechter collapsed fraction — i.e. the version that includes the additional factor of two obtained from the excursion set solution to the cloud-in-cloud problem — with respect to halo mass, we obtain the *halo mass function*, which gives the comoving number density of halos of mass M_h at redshift z:

$$\frac{\mathrm{d}f_{\mathrm{coll}}(>M)}{\mathrm{d}M} = \sqrt{\frac{2}{\pi}}\frac{\delta_c(z)}{\sigma_M^2(M)}\left|\frac{\mathrm{d}\sigma_M}{\mathrm{d}M}\right|\exp\left(-\frac{\delta_c^2(z)}{2\sigma_M^2(M)}\right), \tag{1.50}$$

which is the fraction of collapsed mass inside halos in a mass range around M. More typically, however, we will refer to the halo mass function as a number density since $\frac{\mathrm{d}f_{\mathrm{coll}}(>M)}{\mathrm{d}M} = \frac{M}{\overline{\rho}_0}\frac{\mathrm{d}n(>M)}{\mathrm{d}M}$, giving us:

$$\frac{\mathrm{d}n(>M)}{\mathrm{d}M} = \frac{\overline{\rho}_0}{M}\frac{\mathrm{d}f_{\mathrm{coll}}(>M)}{\mathrm{d}M}, \tag{1.51}$$

which now gives the number of halos per unit volume in some mass bin M. The halo mass function is the key connecting models of structure formation such as the Press-Schechter formalism to observations such as galaxy surveys for example...

When comparing the Press-Schechter halo mass function to numerical simulations, we find that they agree remarkably well over a wide range of scales.

1.3.4 Cosmic dawn of the first stars and galaxies

So far, we only discussed how dark matter particles evolve. But what about normal matter aka baryons? Baryons follow potential wells formed by dark matter to eventually form stars and galaxies. There are two key properties of baryonic matter that dark matter does not have: (i) be pressure supported against gravitational contraction; and (ii) interact with radiation which allows it to cool and heat.

Following the same steps as in Section 1.3.1, we obtain the linear evolution of non-relativistic

and collisionless baryonic matter perturbations (in Fourier space) due to gravity and pressure:

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k = \left[4\pi G\overline{\rho} - \frac{k^2 c_s^2}{a^2}\right]\delta_k. \tag{1.52}$$

Note how this equation is exactly the same as equation 1.10 with the addition of the pressure term on the RHS, where c_s is the speed of sound.

Allowing for the particles to be pressure-supported lets us introduce the Jeans quantities that correspond to the threshold where pressure support turns on and prevents further gravitational collapse. Setting the RHS of the above equation to zero, we derive the Fourier scale k at which pressure support is equal to gravitational attraction, also known as the Jeans scale:

$$\frac{k_J^2 c_s^2}{a^2} = 4\pi G\overline{\rho} \tag{1.53}$$

$$k_J = \frac{a}{c_s}\sqrt{4\pi G\overline{\rho}}. \tag{1.54}$$

Only perturbations with k < kJ can grow. Converting the Jeans scale into real space, we define the characteristic proper Jeans length $\lambda_J \equiv \frac{2\pi}{k_J}$:

$$\lambda_J = c_s\sqrt{\frac{\pi}{G\overline{\rho}}}. \tag{1.55}$$

This is the distance a sound wave can travel in one gravitational free-fall time, $t_{\rm ff} = 1/\sqrt{G\overline{\rho}}$. If $\lambda > \lambda_J$, gravity dominates over pressure and perturbations collapse; if $\lambda < \lambda_J$, pressure support prevents collapse, leading to free-streaming damping. In the collisional case, $\lambda < \lambda_J$ instead leads to acoustic oscillations rather than complete erasure.

The Jeans mass is then the characteristic mass above which structures can collapse:

$$M_J \equiv \frac{4\pi}{3}\left(\frac{\lambda_J}{2}\right)^3 \overline{\rho}. \tag{1.56}$$

We can write the sound speed of an ideal gas as $c_s^2 = \frac{5k_B T}{3\mu m_p}$, where T is the gas temperature, μ is the mean molecular weight, and m_p is the proton mass. From the previous equation, we see that the Jeans mass is therefore sensitive to both the gas temperature and the background density.

As the Universe evolves, both T and $\overline{\rho}$ change with redshift, and so does the Jeans mass. At very high redshifts ($z \sim 100$), the baryons are still thermally coupled to the CMB, yielding a very high Jeans mass: $M_J \sim 10^5 - 10^6, M_\odot$. This is far too massive for small-scale perturbations to collapse, preventing star formation.

After thermal decoupling at $z \sim 150$, the baryon temperature begins to cool adiabatically, while the density still scales as $\overline{\rho} \propto (1+z)^3$. The Jeans mass therefore decreases rapidly with time. By $z \sim 30$, the Jeans mass falls to $M_J \sim 10^3, M_\odot$, finally allowing gas in sufficiently overdense regions to collapse.

In the early Universe $3400 \lesssim z \lesssim 1100$, baryons are still tightly coupled to photons. Such hot gas can be modelled as a relativistic fluid with speed of sound $c_s = c/\sqrt{3}$. Since $M_J \propto c_s^3$, we end up with a very high Jeans mass $M_J \approx 10^{18} M_\odot$ that is much larger than the mass of a typical galaxy! Recall, however, that in the meantime, dark matter that does not interact with photons is free to start forming structure.

While the Jeans mass controls baryonic structure formation on large scales, on smaller scales, baryons do not evolve on their own: they follow the potential wells already dug by dark matter.

Baryons accrete onto already formed dark matter halos and thus get gravitationally heated to the virial temperature of a dark matter halo. Applying the virial theorem KE $= -\frac{1}{2}$ PE (see end of Section 1.3.2), we find that an ideal gas heated by falling into a gravitational potential PE $= -\frac{GM_{\rm vir}}{r_{\rm vir}}$ has virial temperature:

$$T_{\rm vir} \approx 10^4 {\rm K} \frac{\mu}{0.6} \left(\frac{\rm M_h}{10^8 {\rm M}_\odot}\right)^{2/3} \left(\frac{1+{\rm z}}{10}\right) \left[\frac{\Omega_{\rm m,0}}{0.3} \frac{1}{\Omega_{\rm m}({\rm z})} \frac{\Delta_{\rm c}}{18\pi^2}\right], \tag{1.57}$$

where $\Delta_c = \rho/\overline{\rho}$ is the mean density of a dark matter halo of mass M_h and μ is the mean molecular weight of the baryonic gas that depends on its ionization state: $\mu \sim 0.6$ for fully ionized hydrogen gas and goes up to $\mu \sim 1.2$ for fully neutral hydrogen gas. The virial temperature is thus a much more relevant quantity than the mean temperature of the intergalactic medium (IGM) when considering the conditions for the formation of the first stars. For star formation to occur, baryonic gas must be able to radiate away the heat it acquires as it collapses within the gravitational potential of a halo. This highlights a key distinction between dark matter and baryons: while dark matter is collisionless and unable to shed energy, baryonic gas can cool radiatively, allowing it to collapse to much higher densities.

Efficient cooling is therefore the critical ingredient that enables star formation. Primordial gas, composed almost entirely of hydrogen and helium, relies on these elements for its cooling pathways. Atomic hydrogen becomes inefficient at temperatures below $T \sim 10^4$ K (e.g. Barkana and Loeb 2001), which, according to equation 1.57, corresponds to halos of mass $\sim 10^8 M_\odot$ at $z \sim 10$. The very first galaxies must therefore have relied on an additional cooling mechanism: molecular hydrogen (H_2). Conveniently, H2 is an efficient coolant at $T \lesssim 10^4$ K, precisely where atomic hydrogen and helium cooling shut off, and can cool gas further down to $T \sim 10^3$ K (e.g. Barkana and Loeb 2001, Abel et al. 2002, Bromm and Larson 2004). This corresponds to halos with masses $\sim 10^6 - 10^7 M\odot$. However, molecular hydrogen is fragile, and its abundance is easily reduced by environmental effects. In particular, H_2 is readily destroyed by radiation in the Lyman–Werner (LW) band ($\sim$ 10.2-13.6 eV), which is produced once the first stars form. As a result, the formation of stars in neighbouring halos could be strongly regulated or even entirely suppressed by LW feedback. Understanding when and how H_2 ceases to be an effective coolant remains an active research area today (e.g. see Holzbauer and Furlanetto 2011, Fialkov et al. 2013, Wolcott-Green et al. 2017, Schauer et al. 2021, Muñoz et al. 2022). Key uncertainties include the precise H_2 formation rates in different primordial environments and the impact of LW feedback, which can dissociate H_2 molecules and thereby suppress star formation in nearby halos.

When H_2 cooling is effective, the gas can continue collapsing and radiating away thermal energy until it becomes compact enough to form the first generation of stars. These so-called *molecularly-cooled galaxies* (MCGs) host Population III (Pop III) stars, which are thought to be massive, short-lived, and metal-free. Their nuclear burning rapidly synthesises heavier elements — metals such as carbon (C), nitrogen (N), and oxygen (O)—which are dispersed into the interstellar medium (ISM) and eventually the intergalactic medium (IGM) through energetic winds from supernova (SN) explosions. Metal enrichment fundamentally changes star formation. Metals and dust grains introduce efficient new cooling pathways, enabling the transition to the next generation of stars in atomic-cooling galaxies (ACGs), also known as Population II (Pop II) stars. These stars are less massive and longer-lived than their Pop III progenitors, and their properties more closely resemble stellar populations observed in present-day galaxies. Observationally, this transition from Pop III to Pop II star formation remains one of the central frontiers in high-redshift galaxy studies. Spectroscopic surveys with instruments such as JWST probe the metal content of high-redshift galaxies, tracing the build-up of heavy elements and their relation to stellar mass and star formation rate (SFR). Early results indicate a relatively tight relation between stellar

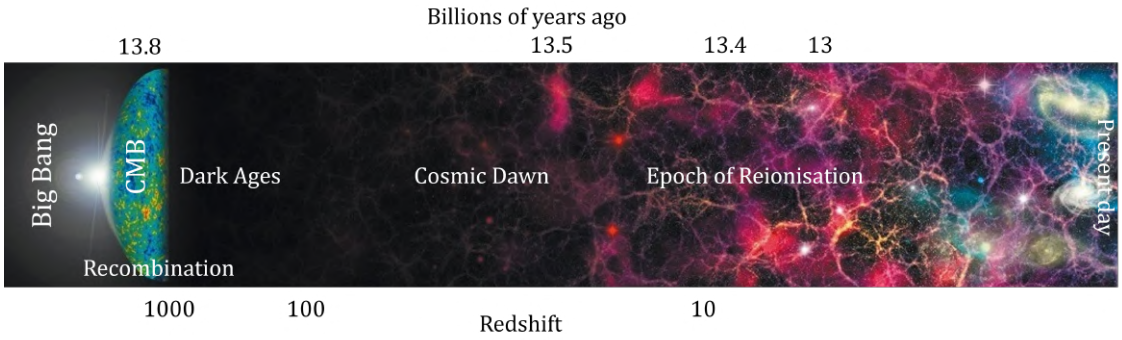

Figure 1.5: History of the Universe, updated from Figure 1.1 with modern simulations. Neutral gas is shown in black. The cosmic web formed by dark matter is in orange, while ionised bubbles carved by the first stars and galaxies are in pink. Credit: Adapted from NASA/WMAP Science Team; R. Ellis (Caltech)

mass, metallicity and SFR (e.g. see Mannucci et al. 2010) extending up to $z \sim 14$, providing critical clues about the timing and efficiency of metal enrichment in the first galaxies.

1.3.5 Epoch of reionisation

Only a few percent of all baryons are in stars and galaxies, while most of them are actually in the diffuse IGM between the galaxies. The first stars form from primordial gas in the IGM.

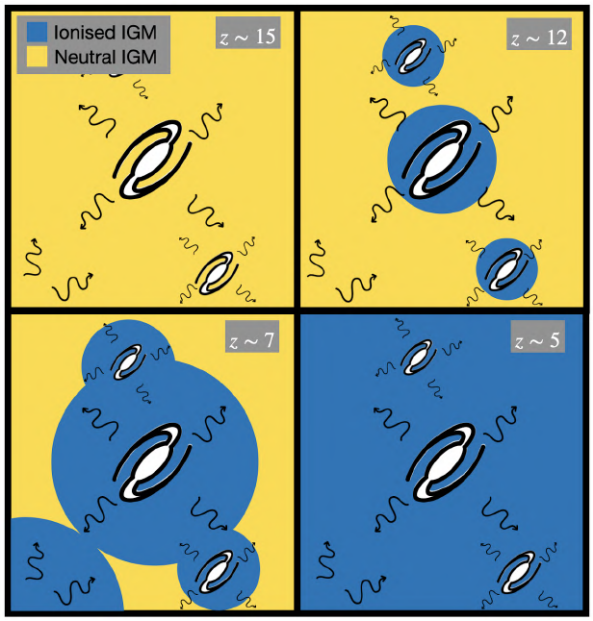

Figure 1.6: Slice of the Universe as it evolves during the EoR. At $z \sim 15$, the first stars and galaxies form in a completely neutral IGM shown in yellow. These galaxies heat the surrounding IGM, creating ionised bubbles shown in blue. The ionised bubbles grow and merge during the EoR $z \sim 12 - 7$, until a completely ionised IGM is observed by $z \sim 5$. Credit: Benedetta Spina.

As the Universe forms more and more galaxies, the gravitational heat they all radiate away begins to heat and ionise the neutral IGM gas around them. These ionised bubbles surrounding galaxies expand and merge over time as they occupy an ever-growing fraction of the volume of the Universe. This so-called *epoch of reionization* (EoR) is the final phase transition of the Universe as the neutral IGM becomes completely ionized as we observe it to be today. We can describe the IGM by tracing its volume-averaged ionisation state and temperature.

A simple model for the volume-averaged ionisation history depends on (i) the properties of the ionising sources i.e. galaxies; and (ii) the properties of the intergalactic gas such as its clumpiness which can cause ionised gas to recombine and thus become neutral again. The mean free path of ionizing photons through the IGM is very short. As a consequence, to first order, the ionization state of the IGM is well-described by highly ionised regions mixed with gas that is mostly neutral. We can thus write the volume-averaged ionization fraction $Q_{\rm HII} \equiv \langle x_{\rm ionised} \rangle$ as competition be-

tween photoionization (first term on RHS) and recombinations (second term on LHS) following Lidz 2016:

$$\frac{\mathrm{d}Q_{\mathrm{HII}}}{\mathrm{d}t} = \frac{\mathrm{d}\left(N_{\mathrm{ion}}/\langle n_{\mathrm{H}}\rangle\right)}{\mathrm{d}t} - \frac{Q_{\mathrm{HII}}}{\bar{t}_{\mathrm{rec}}}, \tag{1.58}$$

where $\bar{t}_{\mathrm{rec}}$ is the average recombination time for ionised hydrogen in the IGM. We further unpack each of these two terms by relating them to galaxy and IGM properties, respectively. We relate the source term to galaxy properties by assuming that each dark matter halo above mass M_{min} hosts a galaxy which emits ionising photons at a certain efficiency ζ. We can thus write the *cumulative* number of ionising photons emitted per hydrogen atom as:

$$N_{\mathrm{ion}} = \zeta f_{\mathrm{coll}}(> M_{\mathrm{min}}) = 40 \left(\frac{f_{\mathrm{esc}}}{0.1}\right)\left(\frac{f_*}{0.1}\right)\left(\frac{N_{\gamma/b}}{4000}\right) f_{\mathrm{coll}}(> M_{\mathrm{min}}), \tag{1.59}$$

where we model the ionizing efficiency ζ as a product of three galaxy properties, each an active research topic even today: (i) the stellar fraction f_* is the fraction of baryons within stars which depends on the star formation efficiency and is expected to be of order percent (e.g. see Dayal et al. 2014, Mutch et al. 2016, Park et al. 2019); (ii) the number of ionising photons per stellar baryon $N_{\gamma/b}$, which depends on the initial mass function (IMF). The IMF is a probability density function which describes the likelihood of a star to have a certain mass when it is formed. Pop II stars are expected to produce roughly 4000 ionising photons during their lifetime, while pop III stars, which are on average more massive than pop II stars, are expected to produce about an order of magnitude more (e.g. see Tumlinson and Shull 1999, Schaerer 2002); and (iii) the escape fraction f_{esc}, which is set by the dynamical and thermal evolution of galaxies. For example, strong winds caused by supernova feedback may eject gas out of the galaxy, thus clearing the way for ionising photons to escape. Current observations of galaxies at lower redshifts $\sim 3-4$ suggest values of order percent (e.g. see Steidel et al. 2001, Inoue et al. 2006, Shapley et al. 2006), while at higher redshifts, escape fractions are also expected to be higher due to more efficient supernova feedback in fainter galaxies (e.g. see Kimm and Cen 2014, Paardekooper et al. 2015, c.f. Ma et al. 2020).

Similarly, we can relate the recombination time to IGM properties. The rate at which ionized gas in the IGM recombines scales as density squared. The volume-averaged recombination rate thus depends on $C \equiv \frac{\langle\rho^2\rangle_{\text{ionized IGM}}}{\langle\rho\rangle^2}$, also known as the clumping factor inside the ionised IGM, which we expect to be of order unity - few during the EoR but could be larger at the earlier stages of the EoR (e.g. see Pawlik et al. 2009, Emberson et al. 2013). We can now write the average recombination time as:

$$\bar{t}_{\mathrm{rec}} = 0.93\ \mathrm{Gyr}\left(\frac{C}{3}\right)\left(\frac{1+z}{7}\right)^{-3}\left(\frac{T_0}{2\times 10^4\ \mathrm{K}}\right)^{0.7}. \tag{1.60}$$

Substituting these last two equations into equation 1.58 and simplifying, we obtain:

$$\frac{\mathrm{d}Q_{\mathrm{HII}}}{\mathrm{d}t} = f_{\mathrm{esc}} f_* N_{\gamma/b} \dot{f}_{\mathrm{coll}}(> M_{\mathrm{min}}) - \alpha_B n_H C Q_{\mathrm{HI}}. \tag{1.61}$$

Armed with a model for the ionisation state of the IGM, we can now describe its thermal evolution. We begin by modelling the IGM as a classical ideal gas in local thermal equilibrium with kinetic temperature:

$$T_K = \frac{2U}{3k_B n}, \tag{1.62}$$

where U is the total internal energy of all gas particles in the system and n is the number density. In the IGM, we have $n \approx n_b(1 + x_i)$. Taking the time derivative and applying the chain rule, we obtain the temperature evolution of the IGM temperature:

$$\dot{T}_K = \frac{2}{3k_B}\left[\frac{1}{n}\dot{U} - \frac{U}{n^2}\dot{n}\right]. \tag{1.63}$$

Applying the chain rule on our previous approximation, we have that $\dot{n} \approx \dot{n}_b(1 + x_i) + n_b\dot{x}_i$. Unpacking $\dot{U} = \dot{Q} + \dot{U}_{\rm adia}$ and $\dot{U}_{\rm adia} = \frac{5}{2}(1 + x_i)k_B T_K \dot{n}_b$. Putting everything together, we obtain [Hui and Gnedin, 1997]:

$$\dot{T}_K = -2HT_K + \frac{2T_K}{3(1+\delta)}\dot{\delta} - \frac{T_K\dot{x}_i}{1+x_i} + \frac{2\dot{Q}}{3k_B n_b(1+x_i)}, \tag{1.64}$$

where the first two terms describe adiabatic cooling/heating, the third term describes the change in energy due to change in the number of particles, and the last term describes IGM heating per unit volume by gas particles from the surrounding radiation field sourced by galaxies.

Interpreting this equation, we have that at early times $200 \lesssim z \lesssim 1100$ the baryon temperature is coupled to the CMB through Compton scattering. During the dark ages $20 \lesssim z \lesssim 200$, the Universe expands and adiabatic cooling takes over the Compton heating. With the cosmic dawn of the first stars around $10 \lesssim z \lesssim 20$, the IGM is heated by the X-rays emanating from these first galaxies. As the IGM begins to reionise during the EoR around $6 \lesssim z \lesssim 10$, ionised regions can reach temperatures of about 10^4 K while neutral regions can remain as cold as 100 K.

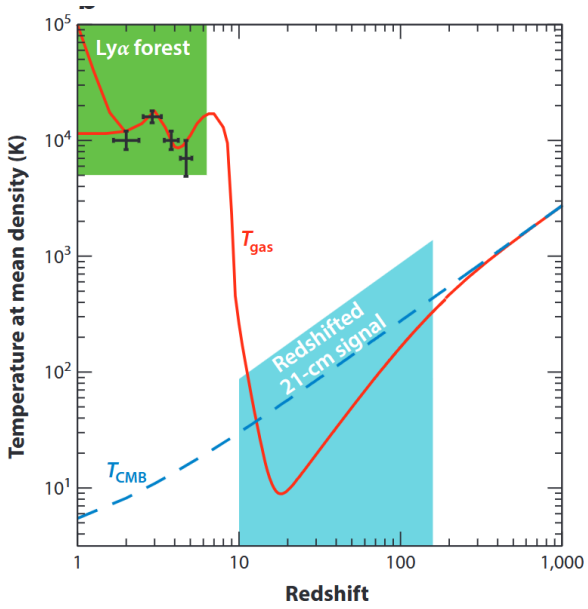

Figure 1.7: Theoretical model of the thermal history of the IGM (solid red line). At low redshift, the IGM gas becomes multiphase, hence the bifurcation of the temperature curve. Figure from McQuinn 2016.

1.4 CD/EoR Observations

In the previous sections, we described very simple models for the formation of the first stars and galaxies. We then developed a simple description of the thermal and ionisation history of the IGM and related them to galaxy and IGM properties. We know that reality is much more complex. Observations provide guidance to improve these simple theories and make them more realistic. In this section, we discuss four common observational probes of the CD/EoR that shall be used throughout the rest of the thesis.

1.4.1 Optical depth to the cosmic microwave background

CMB photons scatter away from the line of sight as the EoR produces free electrons thus increasing the ionised electron column density n_e between us and the last scattering surface at $z \sim 1100$. One of the most powerful probes of the imprints of the EoR on the CMB is the power spectrum (PS) obtained from large-scale E-mode polarization anisotropy maps as observed by

CMB experiments such Planck (e.g. see Planck Collaboration et al. 2020) and WMAP (e.g. see Bennett et al. 2013). EoR affects the CMB PS in two main ways: (i) it scatters CMB photons thus blurs primordial CMB anisotropies. It affects the CMB PS amplitude at all scales; and (ii) the re-scattering of CMB photons from free electrons released during the EoR produces a bump on the CMB polarization (so-called E-mode) PS. The position and height of this bump is determined by the EoR midpoint and duration.

In addition to the E-mode polarization PS, the effect of the EoR on the CMB can also be described via the globally-averaged *Thomson optical depth to the CMB*:

$$\tau_e = \left\langle \int_0^{z_{\rm LSS}} dz \left| \frac{cdt}{dz} \right| n_e \sigma_T \right\rangle_{\rm LOS}, \tag{1.65}$$

where σ_T is the Thomson scattering cross section and n_e is the electron number density. A higher τ_e implies earlier or more extended EoR, while a lower τ_e implies a later or shorter EoR. Calculating τ_e requires a model for n_e and therefore requires a parametrization of the ionization fraction. Typically, CMB experiments assume a simple and often non-physical shape for the reionisation history such as an instantaneous EoR or a hyperbolic tangent EoR. In Figure 1.8, we show values of the Thomson optical depth as reported by various CMB experiments over the last twenty years under the assumption of an instantaneous EoR. Under this assumption, the Universe is considered completely neutral until some reionization redshift $z_{\rm reion}$ and completely ionised after it. As such, redshifts $z > z_{\rm reion}$ do not contribute to the Thomson scattering optical depth, as there are no free electrons during that time, and $\tau_e \approx \tau_{\rm reion}$. For example, the most recent (top) value in Figure 1.8 obtained from Louis et al. 2025 assuming an instantaneous EoR measured $\tau_e = 0.0632^{+0.0055}_{-0.0066}$ (68%CL) with a corresponding reionisation midpoint at $z_{\rm reion} = 8.47^{+0.54}_{-0.61}$. Qin et al. 2020 has found that τ_e is highly model-dependent i.e. the choice of ionisation model can significantly affect the resulting value. They re-analyse the CMB PS from Planck [Planck Collaboration et al., 2020] which report $\tau_e = 0.0522 \pm 0.0080$ with the hyperbolic tangent model and instead obtain $\tau_e = 0.0569^{+0.0081}_{-0.0086}$ with a realistic EoR history.

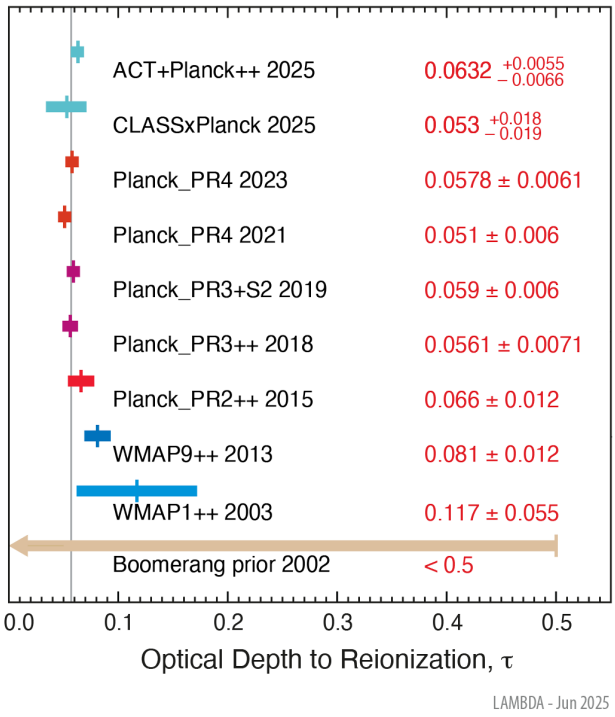

Figure 1.8: Thomson scattering optical depth to the CMB measurements from various CMB experiments over time with most recent at the top. The values of τ_e plotted here assume an instantaneous reionisation model at some redshift $z_{\rm reion}$ for the ionisation fraction where the Universe is fully neutral before $z_{\rm reion}$ and fully ionised after $z_{\rm reion}$. The gray vertical line, representing the weighted average of WMAP and Planck data points, is positioned at $\tau_e = 0.0580$. Image credit: NASA / LAMBDA Archive Team.

1.4.2 Ultraviolet luminosity functions

Galaxies, especially those with young stars that emit many ionising photons, are prime candidates for driving the EoR. In Section 1.3.5, we considered a simple model for the EoR with galaxies as our primary ionising sources. To test this model, we need to measure whether galaxies are abundant enough and bright enough to produce enough ionising photons to drive the EoR. High redshift galaxy surveys test such theories by providing the number density, luminosity, and other galaxy properties, thus linking galaxy formation to the ionization history of the IGM.

In Equation 1.59 in Section 1.3.5, we saw that the photon budget available to ionise the IGM depends on four main galaxy properties: (i) the number of galaxies; (ii) the number of stars in those galaxies; (iii) the number of ionising photons per stellar baryon; and (iv) the fraction of ionising photons that escapes into the IGM. Ionising radiation with rest-frame $\lesssim 900$ Å gets absorbed by the galaxy itself before it reaches the IGM. Galaxy surveys thus observe in the rest-frame ultraviolet (UV) band ~ 1500 Å as it is a direct tracer of young and massive stars which dominate the ionising photon production that can actually escape the galaxy into the IGM. The UV luminosity function (LF) provides the number density of galaxies per UV magnitude bin, $\mathrm{M_{UV}}$. The UV luminosity observed by galaxy surveys is related to the UV magnitude via Oke and Gunn [1983]:

$$\log\left(\frac{L_{\mathrm{UV}}}{\mathrm{erg\ s^{-1}\ Hz^{-1}}}\right) = 0.4 \times (51.63 - M_{\mathrm{UV}}). \tag{1.66}$$

The UV luminosity can then be related to the star formation rate (SFR) via:

$$\dot{M}_*(M_h, z) = \mathcal{K}_{\mathrm{UV}} \times L_{\mathrm{UV}}, \tag{1.67}$$

where $\mathcal{K}_{\mathrm{UV}}$ is often taken to be a constant $\mathcal{K}_{\mathrm{UV}} = 1.15 \cdot 10^{-28} \mathrm{M_\odot yr^{-1}\ Hz\ s\ erg^{-1}}$ (e.g., Madau and Dickinson 2014, Sun and Furlanetto 2016).

UV LF observations (e.g. Bouwens et al. 2015, 2016, Oesch et al. 2018, Harikane et al. 2023), as shown in Figure 1.9, support the theory that the dominant sources of ionizing photons during the EoR are star-forming galaxies. In particular, they find that it is the most abundant, faint and low-mass galaxies (right side on the plots in Figure 1.9) that are thought to be the main drivers of the EoR, in spite of them having lower SFRs and lower UV output individually. UV LF observations during the EoR therefore constrain the astrophysics governing ionising photon production in galaxies via e.g. the star formation rate as well as the efficiency with which these photons escape into the IGM.

1.4.3 IGM Lyman-α damping-wing absorption

While galaxy surveys constrain the global ionising photon budget, rare galaxies hosting the brightest quasars at their core probe individual lines of sight through the IGM. In Figure 1.10, we illustrate how an extremely luminous quasar at high redshift z_{emt} acts as a flash light, illuminating the IGM along its line of sight.

The quasar emits light with an intrinsic flux F_0, in particular with strong emission in the Lyα band. As its light travels through the IGM and is being redshifted, neutral hydrogen clouds along the line of sight absorb photons redshifted into the Lyα band, thus imprinting a damping wing profile at frequencies redward of Lyα. The observed flux F_{obs} at wavelength λ_{obs} for a quasar at redshift z_{emt} can be written as:

$$F_{\mathrm{obs}}(\lambda_{\mathrm{obs}}) = F_0 \left(\frac{\lambda_{\mathrm{obs}}}{1+z}\right) e^{-\tau(\lambda_{\mathrm{obs}})}, \tag{1.68}$$

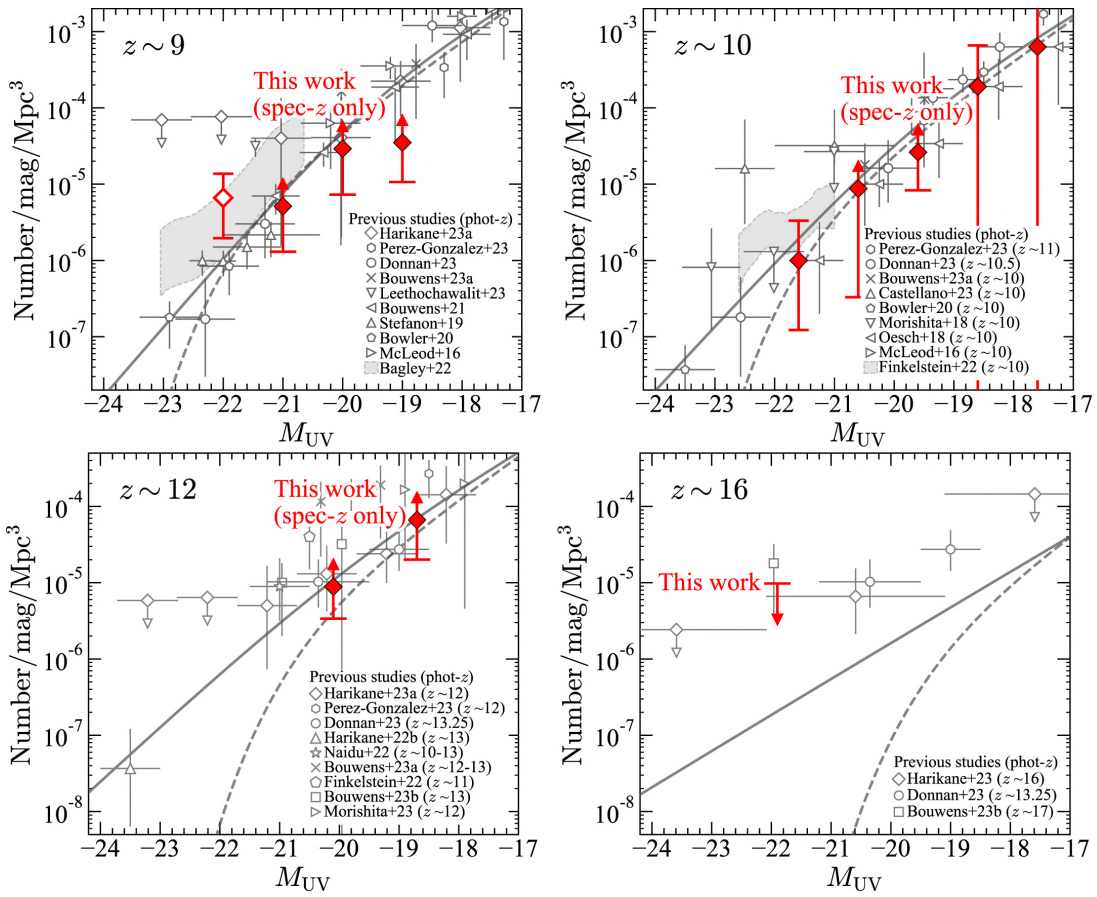

Figure 1.9: Ultraviolet luminosity functions for redshift bins centred around redshift 9, 10, 12, and 16. The points in grey have been obtained from photometric measurements, while the red points have been obtained using more robust spectroscopic measurements. Figure from Harikane et al. 2023.

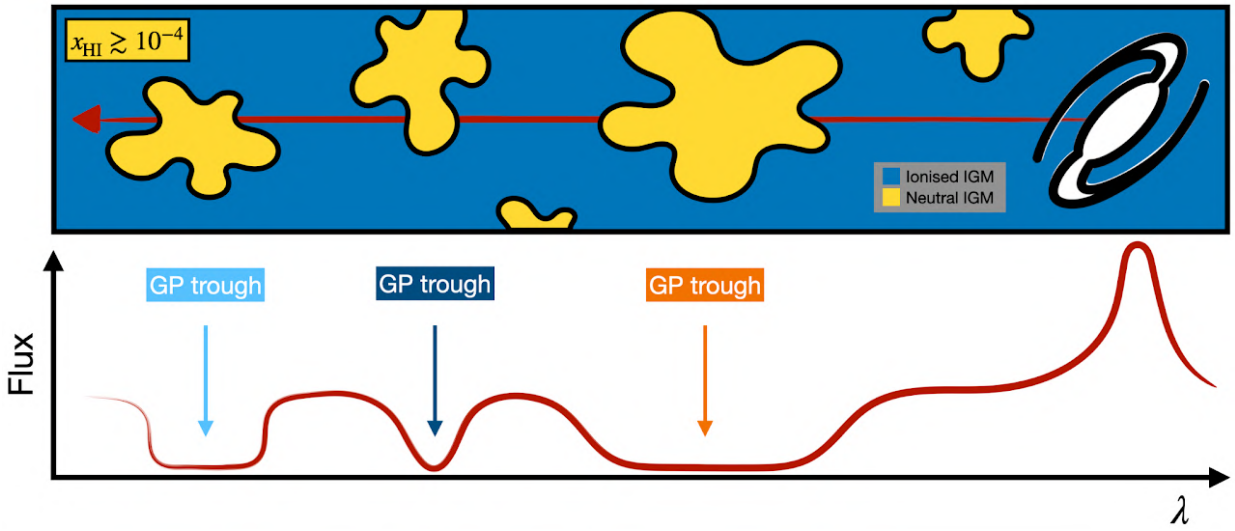

Figure 1.10: A bright quasar emits light with some intrinsic spectrum. As the spectrum redshifts toward the observer, intervening neutral regions (yellow) absorb photons at the Lyα resonance, producing Gunn–Peterson troughs in the observed spectrum. Because Lyα absorption is sensitive to even trace neutral fractions ($x_{\rm HI} \sim 10^{-4}$), quasar spectra provide strong lower bounds on the neutral hydrogen content of the IGM. Multiple lines of sight to different quasars enable tighter constraints on the EoR history. Credit: Benedetta Spina.

where the optical depth τ is defined in the same way as in Equation 1.65, except that in this case we only have access to one single line of sight and the integration is performed up to the redshift of the quasar $z_{\rm emt}$. The absorption profile of the optical depth has two main components: (i) the optical depth due to the local environment of the quasar notably including the ionised region surrounding the quasar (a few kpc in size); and (ii) the optical depth due to neutral hydrogen regions in the IGM beyond the local environment of the quasar. Each of these two components leaves a very different imprint on the observed flux of the quasar. By jointly modelling these two components, we can infer the size of the ionised bubble of the quasar, providing insight into EoR morphology, as well as the IGM neutral fraction along its LOS, which constrains the global IGM neutral fraction.

In Figure 1.11, the coloured points with error bar show constraints on the global neutral fraction from such measurements of quasar damping wing spectra.

1.4.4 The 21-cm signal of neutral hydrogen

So far, we described observations from the CMB (z=1100) and from lower redshifts (z < 15). Current observations focus on closing this observational gap as it also translates into a theoretical gap: Lack of observations in this period leave theories of the formation of the first stars and galaxies largely untested. One approach to closing this gap is by building better instruments that allow us to push observations to higher redshifts e.g. JWST. There is, however, another approach which involves turning to a completely different probe of cosmology and astrophysics: the 21-cm line of neutral hydrogen.

The 21-cm signal of neutral hydrogen is produced by the hyperfine splitting of hydrogen atom ground state: when the proton and electron spins are aligned, the energy of the atom is $E_{21} \equiv 5.9 \times 10^{-6}$ eV (or $E_{21}/k_B = 0.068$ K) higher than when they have opposite spins. When a hydrogen atom transitions from the higher state E_2 to the ground state E_1, it emits a photon

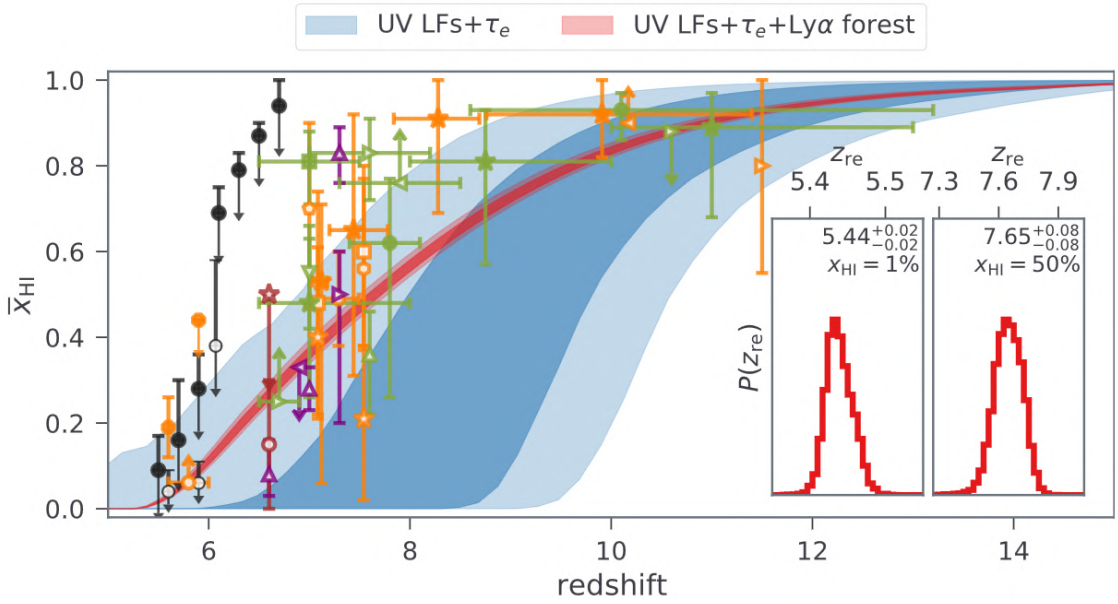

Figure 1.11: EoR history from synergistic Bayesian inferences (see Section 1.5 for more details) shown with shaded regions, while coloured points with error bars are obtained from various observational probes such as the dark pixel method (black, see e.g. McGreer et al. 2015) and Lyα damping-wing absorption from quasars (orange, see e.g. Bañados et al. 2018, Greig and Mesinger 2017, Greig et al. 2019, 2022). Figure from Qin et al. 2025.

with rest-frame wavelength 1420 MHz or 21.1 cm.

The 21-cm line of neutral hydrogen was first predicted theoretically in 1944 by Dutch astronomer Hendrik van de Hulst as a method to study the structure of our galaxy. His prediction was first confirmed observationally in 1951 by Ewen and Purcell at Harvard. The discovery revolutionized radio astronomy, enabling astronomers to map the distribution and motion of hydrogen gas in the Milky Way and other galaxies. Since then, the 21-cm line became a powerful tool for mapping galactic structure and kinematics as well as probing large-scale structure of the IGM. In 2023, the CHIME (Canadian Intensity Hydrogen Mapping Experiment) collaboration made the first statistically-significant detection of the cosmological 21-cm signal from large-scale structure at redshift $z \sim 1$ [CHIME Collaboration et al., 2023], thus validating the 21-cm line as a technique for cosmology. The real power of the 21-cm line, however, lies at higher redshifts. Recall that before the EoR, the Universe was completely neutral hydrogen. We thus expect the 21-cm line to be a powerful probe of CD/EoR as it directly measures the abundance of neutral hydrogen in the IGM. It has the potential of mapping out the structure of the first billion years of our Universe and thus fill the observational gap at $15 \lesssim z \lesssim 30$.

The *redshifted* 21-cm line from CD/EoR roughly corresponds to frequencies of 30-200 MHz, making it a perfect target for modern radio telescopes. Before commissioning a bunch of 21-cm experiments, we need to know whether it is feasible for modern telescopes to actually detect the cosmological 21-cm signal from CD/EoR. The first step to answering this question is performing a simple back-of-the-envelope estimation of the strength of the cosmological 21-cm signal we would observe on Earth.

We consider three mechanisms for 21-cm photon production as shown in Figure 1.13 (left to right): (i) absorption: a neutral hydrogen atom in the ground state absorbs an incoming photon and is promoted to the upper hyperfine level; (ii) spontaneous emission: a neutral hydrogen atom in the upper hyperfine level can spontaneously decay into the ground state and release a

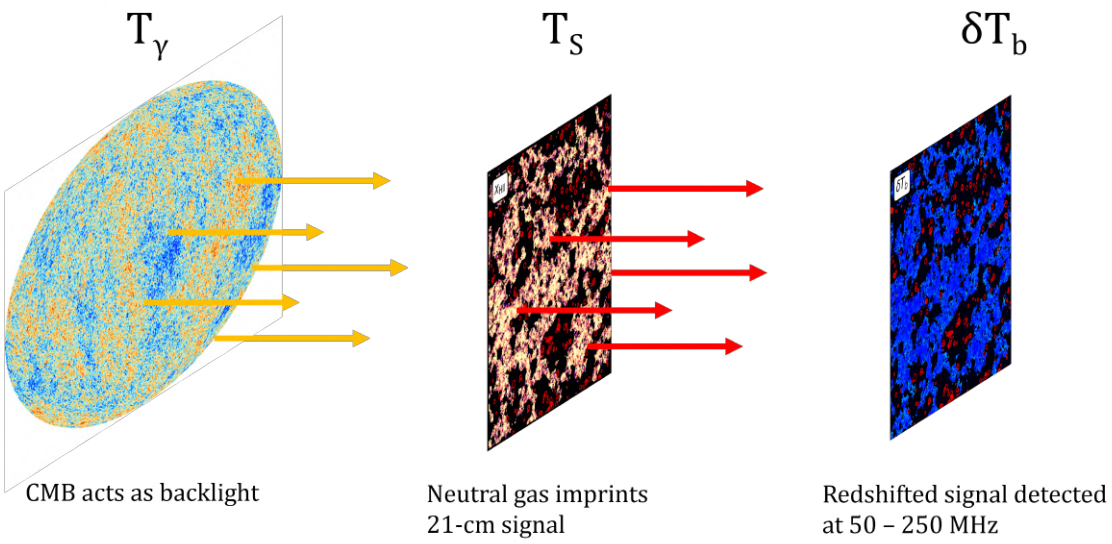

Figure 1.12: The CMB is essentially a backlight shining through the IGM composed primarily of neutral hydrogen before the EoR. The CMB photons interact with neutral hydrogen causing them to emit or absorb 21-cm photons. Adapted with figures from Davies et al. 2025 and ESA/Planck.

21-cm photon; and (iii) stimulated emission: a neutral hydrogen atom in the upper hyperfine level interacts with a photon causing it to decay into the ground state and thus release an additional 21-cm photon. The 21-cm transition has a very long lifetime of about 11 million years because the interaction involved is very weak. On average, a hydrogen atom in the excited hyperfine state will thus spontaneously emit a 21-cm photon roughly once every 11 million years which corresponds to a rate of $A_{21} \sim 1/11$ million years $= 2.85 \times 10^{-15}\mathrm{s}^{-1}$. Moreover, since a neutral hydrogen atom can interact with a photon of rest-frame wavelength of almost exactly 21 cm in order to eventually release it after those 11 million years, we expect 21-cm photons to be extremely rare. However, the IGM is very big and contains *a lot* of neutral hydrogen, so there is still a chance that the signal is strong enough to be observed at higher redshifts!

The three interaction mechanisms from Figure 1.13 suggest that we need the neutral IGM to interact with an external source of photons in order to excite the neutral hydrogen and eventually produce 21-cm photons. At high redshifts, before the formation of the first stars, the most obvious source of photons is the CMB. We thus consider a simple model illustrated in Figure 1.12 to describe the 21-cm signal characterised by its brightness temperature T_b and sourced by a background with temperature $T_\gamma(z) = T_{\mathrm{CMB}}(z) = 2.73(1+z)$ K. As shown on the left of Figure 1.12, the CMB acts as a backlight, shining photons that cross the neutral IGM. A tiny fraction of those CMB photons will have the correct wavelength of 21-cm to interact with the neutral IGM thus emitting or absorbing 21-cm photons relative to the CMB background. Our goal is therefore to write down the *differential brightness temperature* δT_b of the 21-cm signal against the CMB (for in-depth reviews, see e.g. Furlanetto et al. 2006, Pritchard and Loeb 2012).

We begin with the basic radiative transfer equation describing photons emitted with specific intensity I_ν i.e. per unit frequency ν, passing through a gas cloud (i.e. the IGM) in the absence of scattering along path s as shown in Figure 1.12:

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \tag{1.69}$$

where the first term describes photon absorption by the intervening gas while the second term describes photon emission. We drop the path s in favour of the *optical depth* at frequency ν,

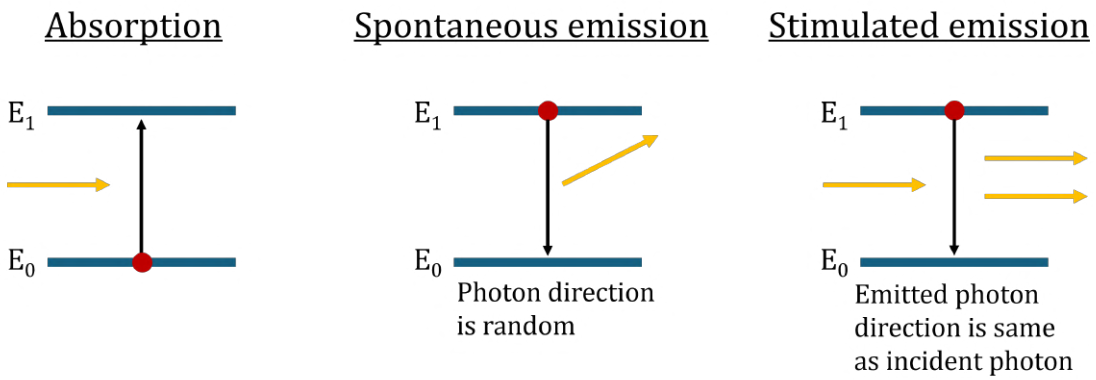

Figure 1.13: We consider three interaction mechanisms between CMB photons and neutral hydrogen atoms. The first one is absorption: the CMB photon hits a neutral hydrogen atom on the ground state (i.e. proton and electron spins are misaligned, E0) causing a spin-flip of the electron. As a result, the neutral hydrogen atom absorbs the CMB photon in favour of the excited state with aligned spins (E1). The second one is spontaneous emission: a neutral hydrogen in the excited state can spontaneously transition from E1 to E0 thus emitting a 21-cm photon. This process is very slow since the excited state has a very long lifetime ($\sim$ 11 million years). The third one is stimulated emission: a CMB photon (or from spontaneous emission) hits an excited hydrogen atom in E1, causing it to transition to the ground state and emit a 21-cm photon (with the original CMB photon unaffected).

defined as $\tau_\nu \equiv \int \alpha_\nu^{\rm eff} {\rm d}s$, where $\alpha_\nu^{\rm eff}$ is the effective absorption coefficient that includes stimulated emission. We then solve the above equation for radiative transfer over a uniform slab of IGM gas in local thermal equilibrium (LTE) with constant emissivities α_ν and j_ν. Since the 21-cm line redshifts to low frequencies, the Rayleigh-Jeans limit with $I \propto T$ is an excellent approximation to the Planck black-body curve. We thus obtain an expression for the brightness temperature of the 21-cm signal sourced by the CMB with temperature T_γ through a uniform gas with optical depth τ_ν:

$$T_b = T_{\rm S}(1 - \exp(-\tau_\nu)) + T_\gamma \exp(-\tau_\nu), \tag{1.70}$$

where T_S is the spin temperature of the IGM gas. This equation suggests that an opaque gas with $\tau_\nu \to \infty$ will have $T_b \to T_S$, while a transparent gas that does not interact with the photon background at all with $\tau_\nu \to 0$ will have $T_b \to T_\gamma$. The spin temperature of the IGM gas is defined in terms of the ratio of the number densities in each of the two hyperfine levels n_2/n_1:

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} \exp(-T_*/T_S) = 3 \exp(-0.068{\rm K}/T_S), \tag{1.71}$$

where $\frac{g_1}{g_0} = 3$ is the ratio of the statistical degeneracy factors of the upper hyperfine level (1S triplet) and and lower hyperfine level (1S singlet), respectively, and $T_* = E_{21}/k_B = 0.068$ K is the temperature difference between the two energy levels. Assuming the IGM gas is in LTE and that the emissivity is isotropic, the optical depth can be written as:

$$\tau_\nu = \int {\rm d}s \; \sigma_1(\nu) n_1 \phi(\nu) (1 - \exp(-0.068{\rm K}/T_S)), \tag{1.72}$$

where $\sigma_1(\nu) \equiv \frac{3c^2 A_{21}}{8\pi\nu_{21}^2}$ is the absorption cross-section with the Einstein coefficient $A_{21} = 2.85 \times 10^{-15}{\rm s}^{-1}$ for spontaneous decay rate of the spin-flip transition, and $\phi(\nu)$ is the normalised 21-cm

line profile such that $\int \phi(\nu)\mathrm{d}\nu = 1$. The line profile may include effects such as natural broadening due to the finite lifetime of excited states, pressure broadening due to collisions between the atoms, and bulk motion due to large-scale peculiar velocities. The line profile for the 21-cm transition is narrow and can be well-approximated by a Dirac delta function at the rest-frame frequency $\phi(\nu) \approx \delta(\nu - \nu_{21})$. Evaluating the above integral, we obtain the expression for the 21-cm optical depth at observed frequency $\nu_{\rm obs}$:

$$\begin{aligned}\tau_{\nu_{\rm obs}} &= \frac{9\hbar c^3}{32\pi\nu_{21}^2}\frac{A_{10}}{4k_B}\frac{\Omega_{b,0}h}{m_H G\sqrt{\Omega_{m,0}}}x_{\rm HI}(1+\delta)\left(\frac{1+z}{10}\right)^{1/2}\left(\frac{10}{T_S}\right)\left(\frac{H}{H+\mathrm{d}v_r/\mathrm{d}r}\right)\\ &\approx 0.003\, x_{\rm HI}(1+\delta)\left(\frac{1+z}{10}\right)^{1/2}\left(\frac{10\,\mathrm{K}}{T_S}\right)\left(\frac{H}{H+\mathrm{d}v_r/\mathrm{d}r}\right)\end{aligned} \tag{1.73}$$

Under the assumption of a small optical depth, which is valid for most of the IGM, we obtain a convenient and intuitive form for the differential 21-cm brightness temperature $\delta T_{\rm b} = T_b - T_\gamma$ from equations 1.70 and 1.73 (see e.g. Madau et al. 1997, Furlanetto et al. 2006, Pritchard and Loeb 2012):

$$\boxed{\delta T_{\rm b} \approx 35\, x_{\rm HI}(1+\delta)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1+z}{10}\right)^{1/2}\left(1-\frac{T_\gamma}{T_{\rm S}}\right)\left(\frac{H}{H+\mathrm{d}v_r/\mathrm{d}r}\right)\,\mathrm{mK}} \tag{1.74}$$

Looking carefully at this equation, we can see that it is the value of the spin temperature relative to the background temperature that determines the detectability of the 21-cm signal: if the spin temperature is equal to the background temperature, then there is no 21-cm signal. In order to quantify the detectability of the 21-cm signal we must therefore understand how the spin temperature is determined. The spin temperature is determined by three main physical processes: (i) radio background coupling: absorption/emission of 21-cm photons from/to the radio background, in our case the CMB; (ii) collisional coupling: collisions with other hydrogen atoms and electrons, effective when the IGM is dense; and (iii) Lyman-α coupling (also known as Wouthuysen-Field (WF) coupling): after the first stars and galaxies form, they emit Lyman-α photons ($\sim$ 122 nm) into the neutral IGM that can be readily scattered off the neutral hydrogen atoms. During the scattering process, the photon is absorbed and excites the neutral hydrogen atom that then relax into *either one* of the hyperfine levels as it re-emits the Lyman-α photon. This resonant scattering of Lyα photons can thus induce a spin-flip.

These processes occur much faster than the lifetime of the 21-cm transition (11 million years). The spin temperature is thus well-approximated by the combination of these effects (see e.g. Pritchard and Loeb 2012 for a detailed review):

$$T_S^{-1} = \frac{T_\gamma^{-1} + x_\alpha T_\alpha^{-1} + x_K T_K^{-1}}{1 + x_\alpha + x_K}, \tag{1.75}$$

where $T_\gamma = T_{\rm CMB}$ as before and T_K is the IGM gas kinetic temperature derived at the end of Section 1.3.5. T_α is the temperature of the Lyα radiation field at the Lyα frequency produced by the first stars and galaxies. The Lyα optical depth is very large in the early Universe due to the high abundance of neutral hydrogen. Lyα photons therefore scatter so many times in the neutral IGM that that they bring the Lyα radiation field and the IGM gas into local thermal equilibrium. We can thus approximate $T_\alpha \sim T_K$.

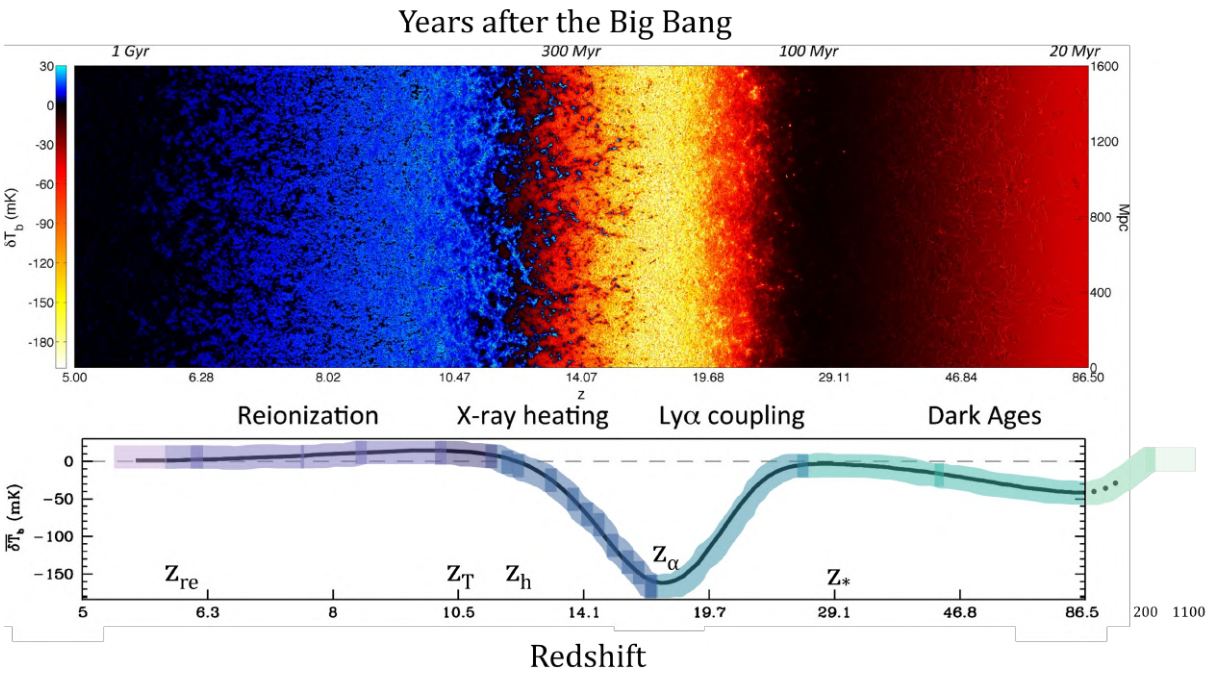

Figure 1.14: Top panel: 21-cm signal lightcone (i.e. averaged over two of the three spatial dimensions) as a function of redshift. Bottom panel: globally-averaged (i.e. over all three spatial dimensions) 21-cm signal. Each of the eight highlighted phases correspond to the eight phases described in the main body. Adapted from Mesinger et al. 2016.

Having understood the physical mechanisms governing the spin temperature, we finally have all the ingredients to round up our understanding of the evolution of the 21-cm differential brightness temperature over cosmic time as illustrated in Figure 1.14, where we follow the shaded regions of the global 21-cm signal in the bottom panel from right to left:

1. $\mathbf{200 \lesssim z \lesssim 1100}$: In the early Universe, the IGM is dense enough to collisionally couple the spin temperature to the gas kinetic temperature i.e. $T_S \sim T_K$. Simultaneously, the gas temperature is thermally coupled to the CMB via Compton scattering due to the free electrons remaining from recombination, setting $T_K \sim T_\gamma$ which yields $\delta T_\mathrm{b} \sim 0$ mK.

2. $\mathbf{80 \lesssim z \lesssim 200}$: As the Universe expands, the IGM gas cools adiabatically so eventually $T_K \propto (1+z)^2 < T_\gamma \propto (1+z)$. Collisional coupling is still effective so we have that $T_S \sim T_K < T_\gamma$, thus producing an early absorption signal. The fluctuations in the 21-cm signal at this time are induced by density fluctuations, making the 21-cm signal a powerful cosmological probe of the matter power spectrum.

3. $\boldsymbol{z_* \lesssim z \lesssim 80}$: As the Universe continues to expand and become less dense, collisional coupling becomes ineffective. The spin temperature decouples from the gas temperature and instead couples to the CMB temperature such that $T_K < T_S \sim T_\gamma$ (we still have adiabatic cooling ensuring $T_K < T_\gamma$). No 21-cm signal at this point.

4. $\boldsymbol{z_\alpha \lesssim z \lesssim z_*}$: The first stars form at z_* and couple $T_S \sim T_K < T_\gamma$ due to WF coupling. The 21-cm signal begins to drop again reaching its minimum around -200 mK. The fluctuations in the 21-cm signal at this time are dominated by density fluctuations and fluctuations in the Lyα radiation field providing an indirect probe of the astrophysical processes governing

the formation of the first stars and galaxies. As more and more stars form, WF coupling will eventually saturate such that $T_S = T_K$ by redshift z_α.

5. $z_h \lesssim z \lesssim z_\alpha$: Fluctuations in the Ly$\alpha$ radiation field do not affect the 21-cm signal once WF coupling saturates. 21-cm signal fluctuations are driven by T_K fluctuations at this time since the IGM is heated. By redshift z_h, the IGM gas is heated everywhere such that $\overline{T}_K = T_\gamma$ and no 21-cm signal is observed except in some hotter regions where $T_K > T_\gamma$.

6. $z_T \lesssim z \lesssim z_h$: Heating continues such that $T_S > T_K > T_\gamma$ and the 21-cm signal is observed in emission. The epoch of reionisation begins as ionised bubbles begin to form around the first galaxies. At this point in time, the 21-cm signal is composed of fluctuations in ionisation, density and gas temperature.

7. $z_{\rm re} \lesssim z \lesssim z_T$: The IGM becomes so hot that the spin temperature becomes saturated at z_T with $T_S \sim T_K \gg T_\gamma$ and we can drop the dependence on T_S in Equation 3.7 yielding a 21-cm signal in emission. The saturated spin temperature regime is very useful for simplifying simulations of the 21-cm signal during the epoch of reionisation. At this point, the 21-cm signal is dominated by ionisation fluctuations, probing all processes which heat the IGM, both cosmological and astrophysical. Currently, we believe the dominant heating mechanism to be X-rays from early accreting black holes, but there are many other exotic models such as cosmic strings, etc.

8. $z \lesssim z_{\rm re}$: Once the epoch of reionisation ends, the IGM is completely ionised with residual neutral hydrogen only found in so called *self-shielded* regions of dense systems such as galaxies as detected by CHIME [CHIME Collaboration et al., 2023]. There is no detectable 21-cm signal from the ionised IGM.

1.4.4.1 Current state of 21-cm

While the theoretical motivation for the 21-cm signal is robust, as we saw above, the real challenge lies in actually detecting it and interpreting the results with our models. The 21-cm signal promises to be an exceptionally rich probe of both astrophysics and cosmology. As such, over the last decade,there have been many efforts to observe the 21-cm signal during the cosmic dawn and epoch of reionisation. To date, however, only upper limits on the 21-cm power spectrum have been reported as shown in Figure 1.15). On the other hand, the Canadian Hydrogen Mapping Experiment (CHIME) has successfully detected the 21-cm signal at $z \sim 1$. Why has no analogous detection been made at higher redshifts? Observing the 21-cm signal comes with one major challenge: foregrounds. Their strength strongly depends on the frequency of observation and becomes increasingly problematic at lower frequencies (i.e. higher redshifts). At the frequencies relevant for 21-cm CD/EoR studies, foregrounds are 2–3 orders of magnitude stronger than at the frequencies used to detect the $z \sim 1$ signal. Figure 1.16 illustrates this challenge. The cosmological 21-cm signal (top-right), as might be obtained from Equation 3.7, is dwarfed by both galactic and extragalactic foregrounds (middle panels). Combined, these are about four orders of magnitude brighter than the target signal. The bottom-left panel shows the combined signal, the starting point for all 21-cm experiments.

Despite these challenges, the 21-cm line has unmatched potential, ultimately able to provide a 3D tomographic map of more than half of our observable Universe, as expected with the upcoming Square Kilometre Array (SKA[1], e.g. Mellema et al. 2013, Koopmans et al. 2015, Mesinger 2019).

[1] https://www.skao.int/en

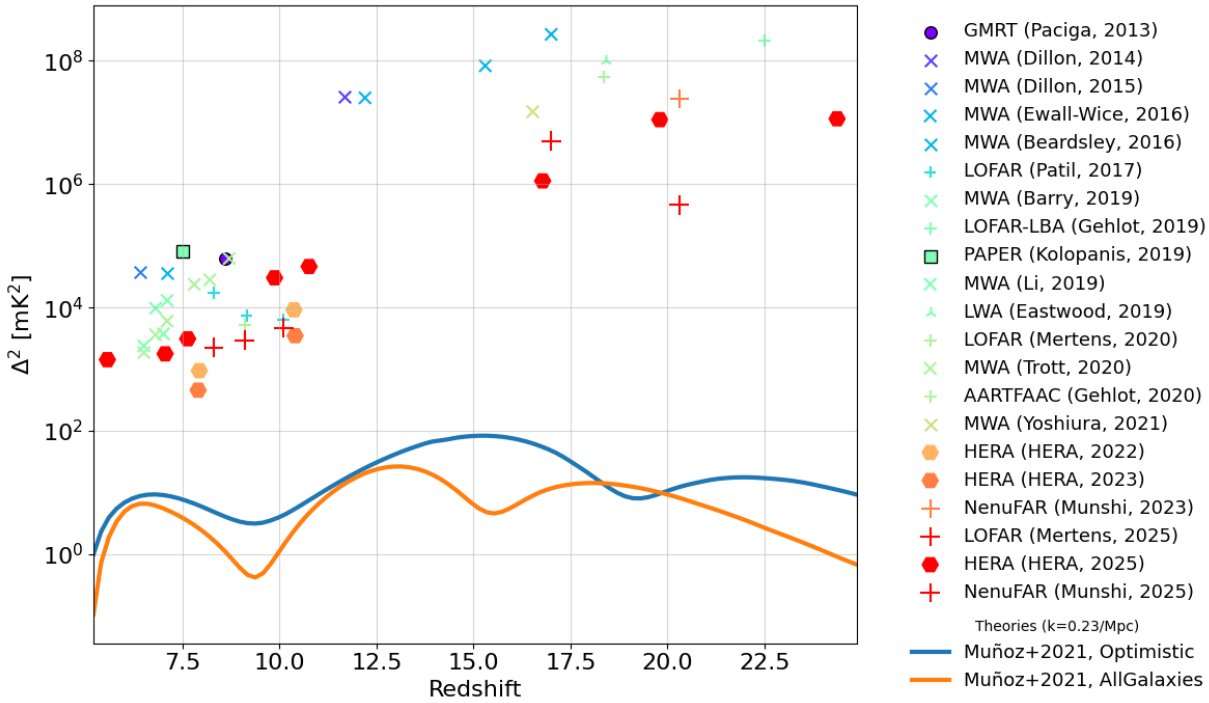

Figure 1.15: Upper limits on the 21-cm PS from current experiments. Figure courtesy of Steven Murray.

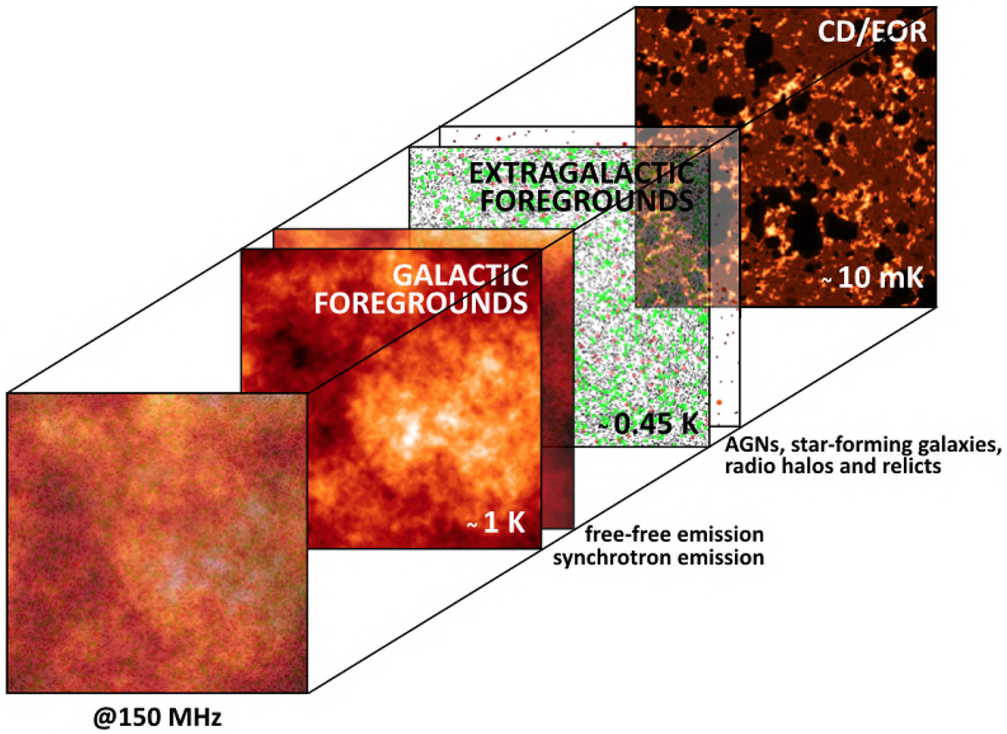

Figure 1.16: Contributions to the observed 21-cm signal. Top-right: cosmological 21-cm signal; middle panels: extragalactic and galactic foregrounds; bottom-left: total observed signal. Foregrounds exceed the cosmological signal by ~ 4 orders of magnitude. Figure from Chapman and Jelić 2019.

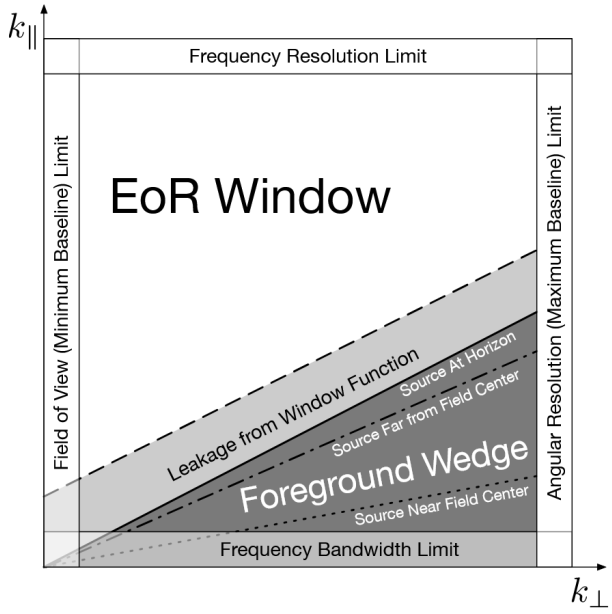

Figure 1.17: Cylindrical 21-cm PS foregrounds anatomy. Foregrounds populate the low $k_\parallel$ region. Chromaticity of the beam causes foregrounds to leak into smaller scales, creating the foreground wedge. The EoR window is the region of cylindrical PS that is considered dominated by the cosmological 21-cm signal. Figure from Kittiwisit et al. 2022.

Precursors to the SKA telescope, interferometers such as the the Murchison Widefield Array (MWA[2], Tingay et al. 2013), the Hydrogen Epoch of Reionisation Array (HERA[3], e.g. DeBoer et al. 2017), LOw Frequency ARray (LOFAR[4], e.g. van Haarlem et al. 2013), and New Extension in Nançay Upgrading loFAR (NENUFAR[5], e.g. Zarka et al. 2012) are instead focused on a first detection of the 21-cm power spectrum (PS) since Fourier space is a natural domain for interferometers. By binning modes in Fourier space, interferometers enhance their S/N. Intrinsically, the 21-cm PS Δ^2_k has the same dimensionality as the real-space field, and is therefore a 3D quantity. In practice, we frequently encounter two reduced forms: the 1D power spectrum, obtained by spherical averaging over all modes of a given magnitude $k = |\boldsymbol{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2}$, and the 2D power spectrum, obtained by cylindrical averaging into line-of-sight modes $k_\parallel = k_z$ and transverse modes $k_\perp = \sqrt{k_x^2 + k_y^2}$. In cylindrical PS space, foregrounds are restricted to low $k_\parallel$ or large line-of-sight modes because they are spectrally smooth, as illustrated by the galactic foregrounds panel in Figure 1.16. However, the chromatic response of the instrument beam causes the signal from foreground-contaminated modes to leak into higher $k_\parallel$ modes, forming the characteristic foreground wedge (e.g. see Parsons et al. 2012, Liu et al. 2014a,b) as shown in grey in Figure 1.17. As a result, clean EoR measurements are performed outside (or near the boundary) of the wedge defining the *EoR window* shown in white in Figure 1.17, where we expect the power to be dominated by the cosmological 21-cm signal.

An additional challenge arises from the instrument itself, whose response imprints non-trivial effects on the observed signal that are often difficult to model or calibrate. These effects, known as *systematics*, can significantly exacerbate the impact of foregrounds, further pushing them into the EoR window and make their removal even more challenging. While foreground contamination remains the primary obstacle for modern 21-cm interferometric experiments, instrumental systematics that amplify this problem pose an equally critical challenge for instruments such as HERA (e.g. Pascua et al. 2025, Rath et al. 2025). Accordingly, HERA is devoting substantial effort to characterizing its main sources of systematics such as mutual coupling, with the aim of mitigating their impact on the final data products.

[2] https://www.mwatelescope.org/
[3] https://reionization.org/
[4] http://www.lofar.org/
[5] https://nenufar.obs-nancay.fr/en/homepage-en/

1.5 Learning about the first billion years

As discussed in the previous section, modern CD/EoR experiments are driven by the following open questions: (i) **first stars** — how and when did the first stars form, what were their physical properties, and how did they evolve into the stellar populations we observe today? (ii) **sources driving the EoR** — which types of galaxies were primarily responsible for reionization, and how efficiently did they leak ionizing photons? (iii) **EoR morphology** — how did ionized bubbles grow, merge, and percolate through the IGM? and (iv) **EoR timing and duration** — when did reionization begin, and how long did it last? Each of these questions has corresponding observational probes that constrain it: (iv) is being constrained by the optical depth τ; (ii), and (i) to some degree, are being constrained by UV luminosity functions; and (iii) is being constrained by quasar Lyα damping wings. A map of the 21-cm signal, such as anticipated from the upcoming SKA, has the potential to constrain all of these simultaneously. However, since such measurements are not yet available,this thesis instead focuses on developing analysis techniques that extract maximal information from current upper limits on the 21-cm power spectrum (Figure 1.15). In particular, we show that exploiting the *synergy* between complementary observables allows us to jointly constrain the timeline of reionisation and the astrophysical properties of the galaxies that powered it.

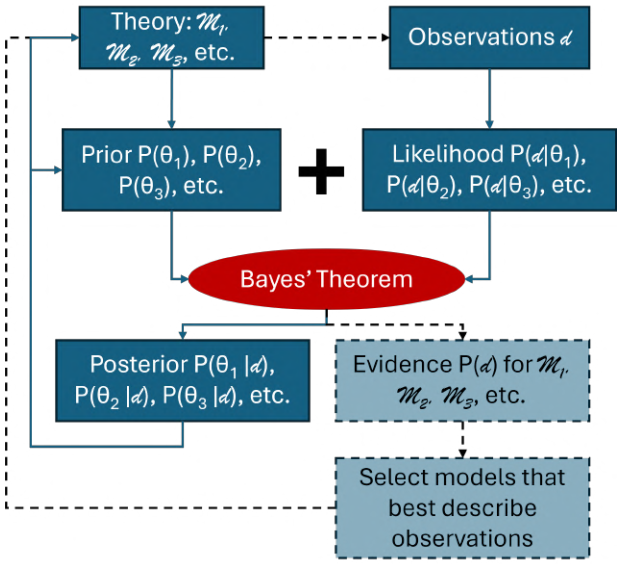

Figure 1.18: The Bayesian scientific method cyclically updates our prior knowledge of the Universe. New observations are translated into knowledge about the Universe (i.e. theory) by updating the prior via Bayes' theorem. Theory in turn informs experimental design, and the cycle continues.

In this section, we introduce Bayesian analysis: the only statistical method that can robustly interpret observations and translate them into knowledge about the Universe. In Figure 1.18, we illustrate the cycle updating our knowledge about the Universe with Bayesian statistics. Starting from the top left of the plot, we begin by encoding *prior beliefs* $P(\theta_1)$ about parameters θ_1 of some model of the Universe $\mathcal{M}_1(\theta_1)$. This prior can be anything: it can reflect previous observations, theoretical arguments, or be deliberately ignorant. Given our new observations d, we update these prior beliefs into a *posterior* $P(\theta|d)$ incorporating the new information from the observations via Bayes' theorem:

$$P(\theta|d) = \frac{P(d|\theta)\ P(\theta)}{P(d)}, \tag{1.76}$$

where $P(d|\theta)$ is the *likelihood* of measuring d given model parameters θ and $P(d) = \int d\theta\ P(d|\theta)P(\theta)$ is the model *evidence*. As shown in Figure 1.18, there may be multiple different models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ describing the same observations d. The model evidence compares them on equal footing and quantifies which model better describes the observations (pale blue in Figure 1.18). This process, known as Bayesian model selection, provides feedback to improve theoretical models. To complete the cycle, theory may inform experimental design, as was the case with the 21-cm line.

Instead of the Bayesian approach, many studies employ what is called the *frequentist approach*,

which is only concerned with the likelihood $P(d|\theta)$ of measuring the observation d with the given model. In other words, it tests whether a model is capable of producing d, but does not quantify how probable the model itself is in light of the observations. This makes it difficult to assess whether a model provides a good overall description of the Universe, or to compare different models on equal footing. Bayesian analysis, on the other hand, updates prior knowledge with new data d, yielding a posterior distribution $P(\theta|d)$ that directly encodes *how plausible different models and parameters are given the observations.*

While simple Gaussian-like posteriors can often be estimated analytically or with approximations such as the Fisher matrix, realistic problems rarely yield such simple forms. Once models become more complex and/or involve a large number of parameters θ, the resulting posteriors can be highly non-Gaussian, multimodal, degenerate, or a combination thereof. Evaluating these posteriors requires applying Bayes' theorem by sampling from the prior and computing the associated likelihood. For high-dimensional problems whose likelihood involves evaluating models with a large number of parameters θ, accurately sampling the posterior becomes extremely challenging. This is known as *the curse of dimensionality*: as the number of parameters grows, the volume of the corresponding parameter space grows exponentially and thus exponentially more samples are needed to explore it adequately.

We mitigate the curse of dimensionality by relying on sampling algorithms that can approximate the posterior efficiently and generate representative samples without brute-force grid evaluations. Markov Chain Monte Carlo (MCMC) methods, for instance, construct correlated chains of samples that eventually converge to the true posterior distribution. Variants such as Metropolis–Hastings or Hamiltonian Monte Carlo differ in how they propose new samples and balance parameter space exploration with sampling efficiency. These algorithms, however, were not built with the Bayesian evidence in mind. Nested sampling (NS), on the other hand, was designed not only to sample the posterior but also to compute the Bayesian evidence, making it valuable for model selection. Within NS itself there are also multiple implementations such as `MultiNest`[6] [Feroz et al., 2009], `UltraNest`[7] [Buchner, 2016, 2019, 2021], `Nautilus`[8] [Lange, 2023], and `PolyChord` (ref) each making different trade-offs in efficiency and robustness.

In practice, no single algorithm is universally optimal: each sampler excels in particular scenarios but may perform poorly in others. Highly multimodal posteriors, strongly degenerate parameter spaces, or extremely high-dimensional models each require different strategies to be sampled efficiently. Consequently, Bayesian inference in cosmology and astrophysics often involves benchmarking several samplers and choosing the most suitable one for the problem at hand, as we shall see in the coming chapters.

In addition to choosing an appropriate sampler, a successful inference setup also requires a carefully chosen likelihood $P(d|\theta)$. The likelihood can be derived or approximated analytically for many simple models. The more complex a physical process is, however, the more difficult it is to analytically write down its likelihood. The likelihood for the 21-cm signal is in fact analytically intractable. This is because the 21-cm signal evolves in a highly non-linear way, coupling to other complex fields throughout its evolution. Supported by CMB experiments, we often approximate the 21-cm PS likelihood with a Gaussian form.

An alternative approach that circumvents the need for an explicit likelihood is *simulation-based inference* (SBI, also known as likelihood-free inference). Instead of evaluating $P(d|\theta)$ directly, SBI relies on the fact that forward models can generate mock observations given model parameters. The posterior is then approximated directly from simulated datasets, enabling in-

[6] http://johannesbuchner.github.io/PyMultiNest/
[7] https://johannesbuchner.github.io/UltraNest/
[8] https://github.com/johannesulf/nautilus

ference even when the likelihood is analytically intractable or computationally prohibitive. SBI is currently a particularly active area of research for 21-cm analyses (e.g. see Prelogović and Mesinger 2023, Meriot et al. 2024, etc.) in preparation for the SKA. A Gaussian likelihood is sufficient as long as the observations being analysed are low S/N, such as upper limits on the 21-cm PS [Prelogović and Mesinger, 2023, Meriot et al., 2024]. However, once high-precision measurements become available, more accurate likelihood treatments—and thus SBI approaches—will be essential. In this thesis, we therefore adopt the Gaussian likelihood approximation as the most efficient choice for present-day 21-cm upper limits, while keeping in mind that SBI is a natural extension.

1.6 Simulating the first billion years

Robustly translating CD/EoR observations into theoretical constraints with Bayesian inference requires simulators that can forward-model multiple observables simultaneously. This is very challenging: simulating only the 21-cm signal is insufficient, and realistic simulations must balance box size, resolution, and physical complexity. These requirements quickly make simulations computationally prohibitive, particularly given the curse of dimensionality: the more parameters a model has, the more forward-model evaluations are needed. For weakly constraining data, such as current 21-cm power spectrum upper limits, this often amounts to $\sim 10^6$ evaluations (e.g. HERA Collaboration et al. 2023), making efficiency essential.

Driven by this computational demand, semi-numerical simulations have emerged as the practical alternative to costly hydrodynamic simulations: in exchange for a modest loss in accuracy, they provide a speed gain of many orders of magnitude, thus making Bayesian inference computationally feasible. Throughout the thesis, we use the `21cmFAST` simulator [Mesinger and Furlanetto, 2007, Mesinger et al., 2011, Murray et al., 2020, Davies et al., 2025], which achieves a runtime of about 1 core hr per simulation – over six orders of magnitude faster than typical radiative transfer simulations – in exchange for a modest $\sim 20\%$ error on the largest modes of the 21-cm power spectrum at low neutral fractions near the end of reionisation (e.g. see Zahn et al. 2011).

1.6.1 Evolving radiation fields

Figure 1.19 provides an overview of how 21cmFAST simulates the radiation fields composing the Universe. The simulation begins by using cosmological parameters to generate initial conditions of the density field, modelled as a Gaussian random field drawn from the matter power spectrum. These initial conditions are then evolved forward in time using second-order Lagrangian perturbation theory (2LPT, Scoccimarro 1998) which bypasses the need for computationally expensive N-body simulations that track the dynamics of individual particles. Dark matter halos are identified on the evolved density fields using a combination of the excursion-set formalism introduced in Section 1.3.3 and a merger-tree algorithm [Davies et al., 2025]. The merger-tree algorithm is applied backwards in time to generate a self-consistent dark matter halo population evolving in time as marked with red circles in Figure 1.19. Galaxies are assigned to the resulting population of dark matter halos by sampling a conditional probability function based on well-established empirical scaling relations (e.g. see Park et al. 2019, Davies et al. 2025) such as the SHMR (see Section ??) and the SFMS (see Section ??). This model produces a *stochastic* galaxy population. The mean and scatter of the scaling relations control the amount of stochasticity..

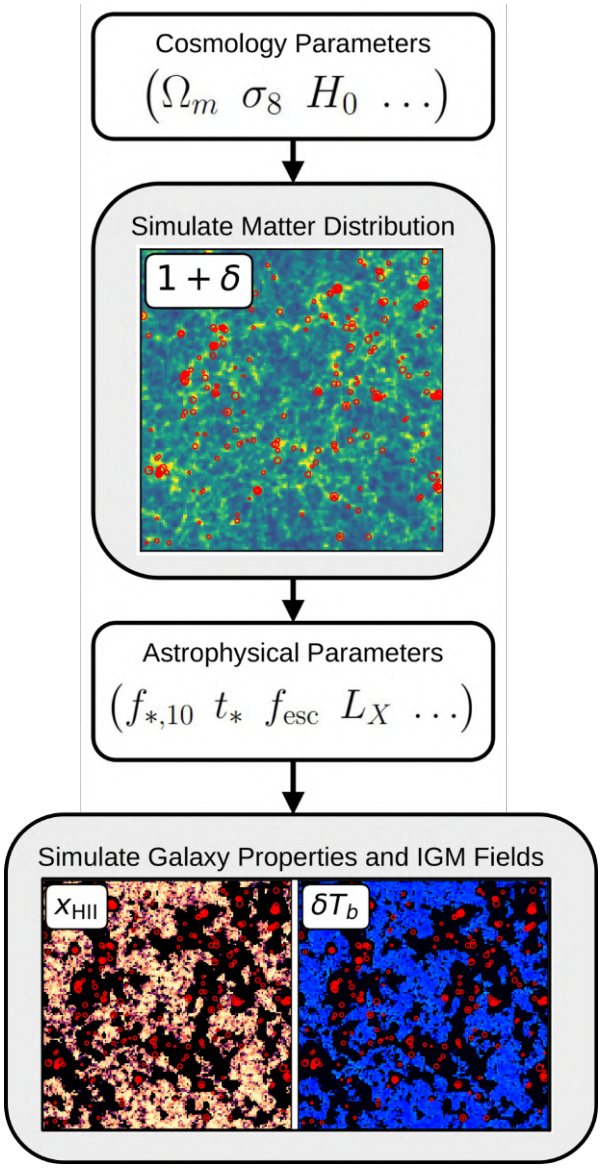

Figure 1.19: 21cmFAST simulator, adapted from Davies et al. 2025.

As in Section 1.3.5, 21cmFAST also operates under the assumption that dark matter halos host galaxies which source inhomogeneous, large-scale cosmic radiation fields. The ionizing, X-ray and soft UV cosmic radiation fields sourced by these galaxies are computed with approximate radiative transfer and evolve the thermal state of the IGM. At the bottom of Figure 1.19, we show slices of the IGM neutral fraction and the 21-cm brightness temperature at the same redshift as the density slice above. The ionization and thermal state of the IGM gas are then tracked with a set of coupled differential equations, allowing us to compute observables such as the ones introduced previously in Section 1.4. Stitching together boxes of evolved radiation fields allows to produce *lightcones* that mimic observations. In Figure 1.20, we show, from left to right, lightcones of the dark matter density, followed by the lightcones of two galaxy emissivities: the cumulative ionising photon density per hydrogen atom (Nion in Section bla) and the soft-band X-ray emissivity, followed by lightcones of two observables: the 21-cm signal lightcone and CII surface brightness density.

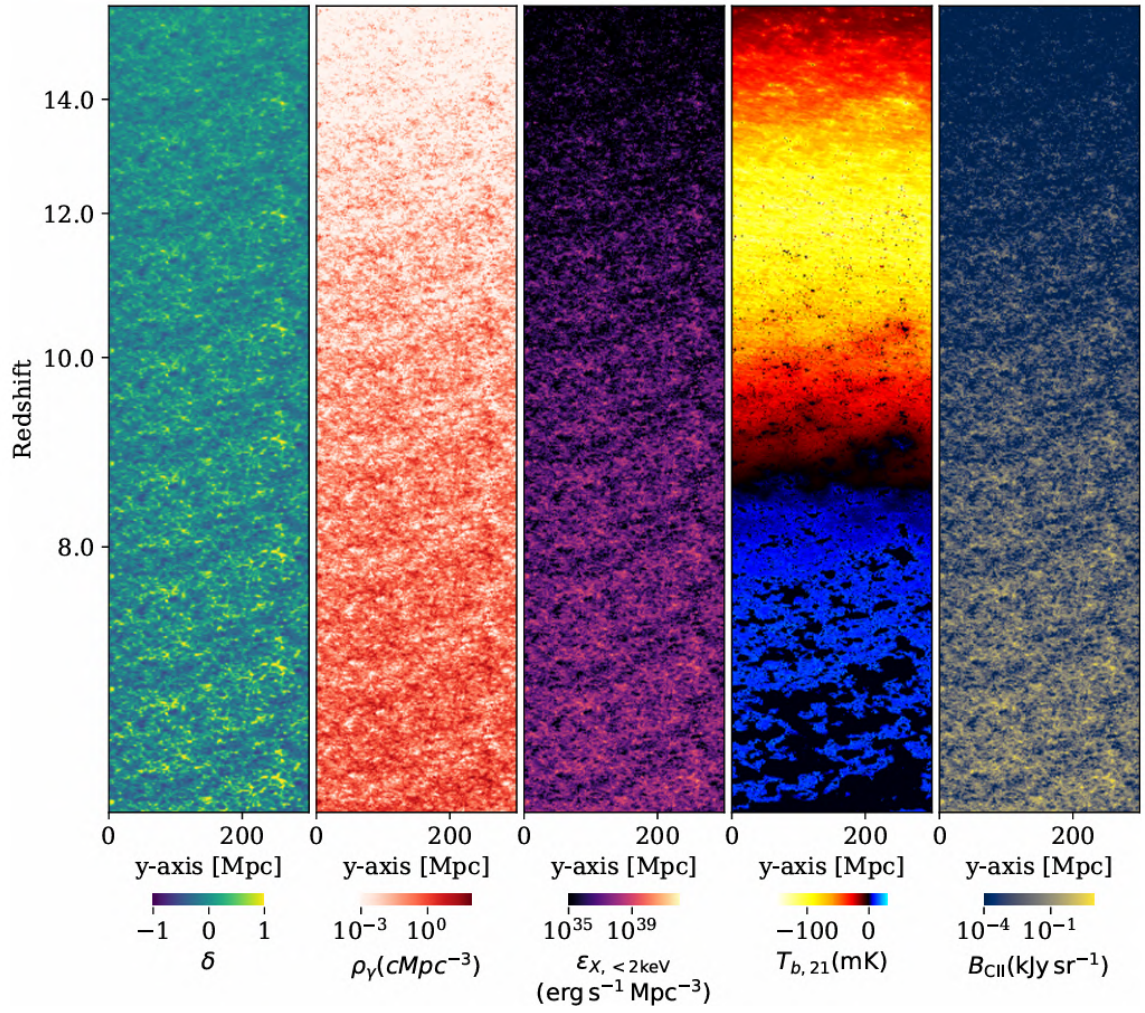

Figure 1.20: From left to right: a lightcone of the matter density, followed by two lightcones of galaxy properties: cumulative ionising photon density per hydrogen atom ($N_{\rm ion}$) and soft-band X-ray emissivity, followed by lightcones from two observables: 21-cm line, and singly-ionised Carbon line (CII) surface brightness density. Figure from Davies et al. 2025.

1.7 Machine learning

While semi-analytical simulators such as `21cmFAST` are orders of magnitude faster than full radiative transfer simulations, they remain computationally expensive for Bayesian inference: a typical Bayesian inference with $\sim 10^6$ forward model evaluations would require $\sim 10^6$ CPU core hours. With new observations released at an unprecedented rate due to the numerous CD/EoR experiments on sky, the demand for faster inferences is greater than ever. In the past few years, machine learning (ML) has been proved to be a powerful tool in across many applications, offering an attractive solution to the challenge. In this section, we introduce machine learning and focus on its use as an *emulator* that replaces the simulator during the inference and accelerates it by over four orders of magnitude, enabling full inferences in as little as a few GPU minutes.

1.7.1 Introduction to machine learning

Machine learning started out as an attempt to understand the human brain and recreate its function in a computer program. Its goal is to mimic aspects of human cognition by *learning* patterns from data and then *generalizing* these learned patterns to make accurate predictions on new, unseen data. In practice, this amounts to learning a mapping from some input to its corresponding output without explicitly prescribing the functional form of that mapping. The mapping is instead approximated by many simple building blocks called *neurons*, inspired by nervous cells in the brain. In artificial neural networks (ANNs), each neuron receives signals from incident neurons, applies a transformation to their weighted sum, and passes the result onward. This behaviour has a natural mathematical analogue: linear regression. Given an input x, the simplest artificial neuron computes a weighted sum $y_i = W_i^\top x$, where the weights W_i of a neuron i represent the strength of the connections between that neuron and its incident neurons. However, a network composed only of such linear neurons cannot capture the complex patterns present in real data. Introducing non-linear *activation functions* f significantly improves the capability of the ANN to learn complex patterns. Common activation functions include the rectified linear unit (ReLU) $f(y_i) = \max(0, y_i)$, the sigmoid $f(y_i) = \frac{1}{1+\exp(-y_i)}$, and the hyperbolic tangent $f(y_i) = \tanh(y_i) = \frac{\exp(y_i)-\exp(-y_i)}{\exp(y_i)+\exp(-y_i)}$. By connecting many such neurons into layers and stacking them into deep architectures with multiple hidden layers as shown in Figure 1.21, ANNs can capture complex patterns in data, much like networks of neurons in the brain collectively give rise to cognition.

The use of machine learning as general-purpose function approximators is rigorously supported by mathematical arguments. In particular, the *universal approximation theorem* [Hornik et al., 1989] implies that with sufficiently many neurons, the neural network can approximate any continuous function on compact domains to arbitrary accuracy. Other formulations of the theorem show that universality can also be achieved by keeping the width of the network fixed while increasing its depth. In practice, this implies that a sufficiently wide or deep feed-forward neural network can, in principle, model the complex, non-linear relationships in real-world data. While these universal approximation theorems provide an essential mathematical foundation for machine learning, it is important to note that they are *existence proofs*: they guarantee that a sufficiently large neural network with the right activation function can approximate any continuous function, but they do not prescribe how to tune the weights of the network, nor do they specify how wide or deep the network must be for a given task. In practice, identifying and training a suitable network remains a non-trivial challenge.

A neural network is *trained* by optimizing its free parameters W with some cost function that quantifies its performance. Analogous to Bayesian inference, training requires exploring

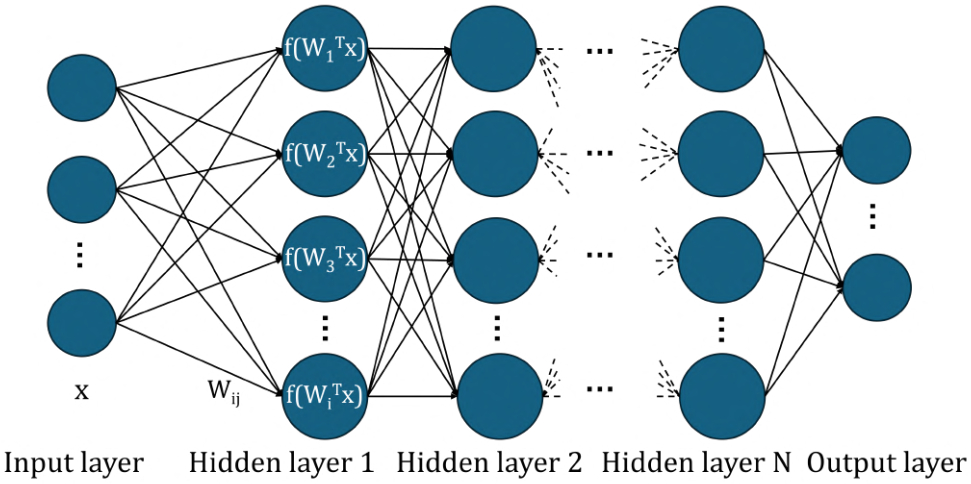

Figure 1.21: Feed-forward neural network with N hidden layers. A neuron i with weight matrix W_i outputs the weighted sum $f(W_i^\top x)$, where f is a non-linear activation function.

a very high-dimensional parameter space, often with millions of dimensions. Unlike Bayesian inference, however, NN training typically relies on simpler optimization algorithms that exploit the fact that we can efficiently compute the gradient of the cost function with respect to all trainable parameters. In practice, the cost function and its gradient are estimated on *mini-batches*, small random subsets of the training data. Mini-batches make training computationally more feasible and introduce stochasticity into the training process which can contribute to improving the NN's generalisation capabilities (ref). Generalisation can be further encouraged by adding other explicit forms of *regularisation* which prevent the NN from simply memorising the training data. The weights W are optimised iteratively on each mini-batch via stochastic gradient descent (SGD) or variants such as Adam. Each mini-batch provides an estimate of the gradient of the cost function. Analytically deriving the gradient of a complex network can be challenging, especially as its depth increases. An efficient solution is the *backpropagation* algorithm which computes the gradient of the cost function by recursively applying the chain rule. Once the gradient is obtained, the NN weights are updated in the direction of the negative gradient. This process is repeated for successive mini-batches until convergence.

While this process efficiently trains the NN, it has limitations. The high-dimensional parameter space being optimised is very complex. It has many local minima and simple optimisation algorithms can easily become stuck in a local minimum, especially when the training procedure is combined with a decaying learning rate. Nonetheless, empirical evidence suggests that most local minima encountered in practice are nearly as good as global minima (e.g. Li et al. 2018). Understanding how to shape the parameter space and design architectures that facilitate training remain active research subjects today.

1.7.2 Emulating the 21-cm signal

Relatively inexpensive semi-analytical simulators such as `21cmFAST` make it feasible to generate databases large enough to train accurate emulators of various observables, such as those discussed in Section 1.4. As detailed in Section 1.4.4.1, most CD/EoR 21-cm experiments are targeting the power spectrum, often presented in the form of upper limits on the spherically-averaged 1D PS. As such, the first emulators in the field naturally targeted the 1D PS (e.g. see Kern et al. 2017, Shimabukuro and Semelin 2017, Schmit and Pritchard 2018, Jennings et al. 2019,

Ghara et al. 2020, Mondal et al. 2022). These early emulators typically vary three parameters governing the EoR such as the ionising efficiency introduced in Section 1.3.5 and use a simple ANN architecture as the one illustrated in Figure 1.21. They emulate the 1D PS w about 10% accuracy.

Parallel to 1D PS observations, there have also been single-dish experiments targeting the globally-averaged 21-cm signal. The Experiment to Detect the Global Epoch of Reionization Signature (EDGES) collaboration reported the detection of a strong 21-cm absorption signal between redshifts $\sim 15 - 20$ (c.f. the global 21-cm signal in Figure 1.14). To interpret these surprising results, there became a demand for emulators of the global 21-cm signal (e.g. see Cohen et al. 2020, Bevins et al. 2021, Bye et al. 2022). These emulators vary seven parameters – many more parameters than previous emulators. They emulate the global signal w about 1% accuracy. Recent emulators boast much more sophisticated NN architectures such as recursive neural networks (RNNs) [Prelogović and Mesinger, 2023], and long-short-term memory (LSTM, Hochreiter and Schmidhuber 1997) [Dorigo Jones et al., 2024].

All of the emulators mentioned so far only emulate one single summary statistic. As discussed in Section 1.4 however, different CD/EoR summary observables probe complementary astrophysical and cosmological mechanisms, and their synergy is essential to obtain the most informative posterior. My thesis focuses on exploiting this synergy while employing cutting edge ML architectures. *My work provides the first emulators capable of jointly predicting six CD/EoR summary statistics to sub-percent error while varying up to twelve astrophysical and cosmological parameters and employing state-of-the-art NN architectures such as score-based diffusion models.*

1.8 Thesis overview

1.8.1 Paper contributions

In Chapter 2, we introduce `21cmEMU`, the `21cmFAST` first emulator capable of emulating six summary statistics simultaneously with sub-percent median accuracy. Given nine astrophysical parameters as input, `21cmEMU` produces all six summary statistics in under a millisecond, thus reducing the cost of Bayesian inferences by over 4 orders of magnitude in comparison to `21cmFAST`. We then test the robustness of `21cmEMU` by applying it on the same inference as was done in HERA Collaboration et al. 2022a with `21cmFAST` and find that the result with the emulator is qualitatively the same as with direct simulation. Having established `21cmEMU` as a reliable and accurate tool for Bayesian inference, we then perform a set of new inferences. First, we perform a set of inferences that demonstrate how different CD/EoR probes form a synergy and set the stage for 21-cm PS upper limits from HERA Collaboration et al. 2023 to further constrain the posterior. Next, we investigate the impact of constraints on the EoR history from Lyα forest observations and find an improved constrained on the parameters governing the ionising escape fraction. Finally, we use `21cmEMU` to forecast the constraint from HERA's 6th season phase II observations *assuming no residual systematics in the data.* We find that, optimistically, we could expect a detection of the 21-cm PS at $z \sim 6 - 7$. We also find that we can recover the EoR history down to $\sim 10\%$ accuracy, which places stronger constraints on the paramters governing the ionising escape fraction, which are almost completely unknown currently.

In Chapter 3, we present the addition of a third emulator to `21cmEMU` - `21cmEMUv3`, which is trained on more realistic `21cmFAST` simulations including a population of MCGs (pop III stars) in addition to ACGs (pop II stars). Moreover, this third emulator features cutting edge architectural elements, such as LSTM layers and score-based diffusion. In `21cmEMUv3`, LSTM layers are used to emulate five out of six summary statistics. The sixth summary statistic, the 21-cm PS, is upgraded from the spherically averaged 1D PS to the cylindrically averaged 2D PS. To emulate the 2D 21-

cm PS, we implement a score-based diffusion model. We test the robustness of this model by reproducing the result obtained by Lazare et al. 2023, who were the first to train a `21cmFAST` emulator of ACGs and MCGs of the 1D 21-cm PS. When including both ACGs and MCGs into the model, we expect the X-ray luminosity produced by one population to be lower in comparison to simulating only one galaxy population which then needs to produce more X-rays. We find that we can recover their result when using the `MultiNest` sampler, but not when using the `nautilus` sampler. We then build upon their result by introducing a new prior informed by hydrodynamic simulations which favours lower stellar masses. We find that such a prior nevertheless supports higher X-ray luminosity than the flat prior used in Lazare et al. 2023. This highlights that when the likelihood is not highly constraining, the choice of prior becomes critical, as it can strongly influence the posterior and thereby the qualitative conclusions drawn from it.

In Chapter 4, we change gears and use machine learning as a tool to improve the quality of Bayesian inferences. One major issue in modern Bayesian inferences is that they typically forward model a much smaller volume of the Universe than that probed by the observations they are supposed to reproduce. This discrepancy in volumes between the simulation and the observation is the source of two problems: (i) the sample variance of the forward model (i.e. the deviation of the simulated sample from the population mean due to finite box size) is much larger than the cosmic variance of the observation; and (ii) the forward model probes a completely different region in cylindrical k-space than the observation. (i) is an issue because, as upcoming observations become more and more sensitive at large scales, forward models remain sample variance limited at those same scales. The fact that the forward model is sample variance limited reduces the precision of the forward model at large scales and therefore prevents us from extracting the most information out of the observations. (ii) is an issue because the cylindrical 21-cm PS is anisotropic due to redshift-space distortions (RSDs), and the evolution of the lightcone along the line of sight. The anisotropy of the 2D 21-cm PS can thus introduce significant biases onto spherically averaged 1D 21-cm power spectra obtained by averaging over different regions in cylindrical k-space. In this work, we mitigate issue (i) with `21cmPSDenoiser`, a score-based diffusion model that mitigates sample variance given a model-agnostic cylindrical 21-cm PS realisation by generating the corresponding mean 2D 21-cm PS. We demonstrate that `21cmPSDenoiser` outperforms traditional sample variance mitigation techniques such as fixing and pairing. We also show that it generalises well to cylindrical PS from other simulators. We mitigate issue (ii) by introducing a cut on the Fourier modes included when spherically-averaging the 2D PS to ensure we only average over a k-space footprint that is very similar to observations. Finally, we apply `21cmPSDenoiser` and the k-space cut in an inference context and show that it produces an unbiased posterior 50% narrower in most inferred parameters in comparison to the traditional inference pipeline which produces a biased and broad posterior.

1.8.2 Package contributions

Advancing science relies on building upon the work of others. For this reason, I have always been motivated to publish my work as well-documented and well-written Python packages, ensuring they are easily accessible to the community through common installation pathways such as `pip`. While I significantly contributed to multiple existing packages such as `powerbox` [Murray, 2018], `21cmMC` [Greig and Mesinger, 2015, 2017, 2018], and `21cmSense` [Pober et al., 2013, 2014, Murray et al., 2024], I have also developed from scratch and actively maintain the following packages:

- `21cmEMU` was developed for Breitman et al. 2024 and extended with `21cmEMUv2` Cang et al. 2024 and `21cmEMUv3` in Breitman et al. 2025b. This package provides an easy to use

Python wrapper that allows the user to select which emulator they wish to use and produce summary statistics from it in just a few lines of code.

- `21cmPSDenoiser` was developed for Breitman et al. 2025a. It's a wrapper around the score-based diffusion model presented in that paper.
- The Ultimate Eor Simulation Data AnalYser `tuesday` was developed throughout all of these projects, but especially the last two, Breitman et al. 2025a and Breitman et al. 2025b. The purpose of this package is to provide the community with well-documented Python code for performing common tasks in CD/EoR analyses, such as calculating power spectra and other post-processing operations. It also serves as a centralized, maintainable repository that preserves and organizes code contributions from members of our group as they transition to new positions.

Chapter 2

21cmEMU: an emulator of 21cmFAST summary observables

Daniela Breitman[1], Andrei Mesinger[1 2], Steven G. Murray[1 3], David Prelogović[1], Yuxiang Qin[4 5], Roberto Trotta[2 6 7]

[1]Scuola Normale Superiore (SNS), Piazza dei Cavalieri 7, Pisa, PI, 56125, Italy
[2]Centro Nazionale "High Performance Computer, Big Data and Quantum Computing"
[3]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA
[4]School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
[5]ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D)
[6]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136 Trieste, Italy
[7]Imperial Centre for Inference and Cosmology (ICIC), Imperial College, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, U.K.

Abstract

Recent years have witnessed rapid progress in observations of the Epoch of Reionization (EoR). These have enabled high-dimensional inference of galaxy and intergalactic medium (IGM) properties during the first billion years of our Universe. However, even using efficient, semi-numerical simulations, traditional inference approaches that compute 3D lightcones on-the-fly can take 10^5 core hours. Here we present `21cmEMU`: an emulator of several summary observables from the popular `21cmFAST` simulation code. `21cmEMU` takes as input nine parameters characterizing EoR galaxies, and outputs the following summary statistics: (i) the IGM mean neutral fraction; (ii) the 21-cm power spectrum; (iii) the mean 21-cm spin temperature; (iv) the sky-averaged (global) 21-cm signal; (v) the ultraviolet (UV) luminosity functions (LFs); and (vi) the Thomson scattering optical depth to the cosmic microwave background (CMB). All observables are predicted with sub-percent median accuracy, with a reduction of the computational cost by a factor of over 10^4. After validating inference results, we showcase a few applications, including: (i) quantifying the relative constraining power of different observational datasets; (ii) seeing how recent claims of a late EoR impact previous inferences; and (iii) forecasting upcoming constraints from the sixth observing season of the Hydrogen Epoch of Reionization Array (HERA) telescope. `21cmEMU` is publicly-available, and is included as an alternative simulator in the public `21cmMC` sampler.

2.1 Introduction

The cosmic dawn (CD) of the first luminous objects and eventual reionization of the Universe remain among the greatest mysteries in modern cosmology. Recent years have seen a dramatic increase in observations of the CD and epoch of reionization (EoR). These include: (i) the Lyman-α forest (e.g., Fan et al. 2006, Becker et al. 2007, 2015, Bosman et al. 2018, D'Odorico et al. 2023); (ii) damping wings in quasar spectra (e.g., Bolton et al. 2011, Mortlock et al. 2011, Bañados et al. 2018, Wang et al. 2020, Yang et al. 2020); (iii) Lyman-α emission from galaxies (e.g., Ouchi et al. 2010, Clément et al. 2012, Konno et al. 2014, Drake et al. 2017, Hoag et al. 2019, Shibuya et al. 2019); (iv) large-scale polarization of the cosmic microwave background (CMB; e.g., Planck Collaboration et al. 2020, de Belsunce et al. 2021, Heinrich and Hu 2021; (v) secondary kinetic Sunaev-Zeldovich (kSZ) CMB anisotropies (e.g., Das et al. 2014, George et al. 2015, Reichardt et al. 2021); (vi) upper limits on the cosmic 21-cm power spectrum (PS; e.g., Mertens et al. 2020, Trott et al. 2020, HERA Collaboration et al. 2022b, 2023). This is set to culminate with a 3D map of HI during the first billion years, expected with the upcoming Square Kilometer Array (SKA; e.g., Mellema et al. 2013, Koopmans et al. 2015, Mesinger 2019).

In step with these observational advances, Bayesian inference techniques have been developed that allow us to forward model the observations and constrain the parameters of reionization as well as the galaxies responsible (e.g., Choudhury and Ferrara 2005, Mesinger et al. 2015, Greig and Mesinger 2015, Mason et al. 2018, Greig et al. 2019, Mondal et al. 2020, Ghara et al. 2020, Qin et al. 2021, Choudhury et al. 2021, HERA Collaboration et al. 2022a, Maity and Choudhury 2022, Nikolić et al. 2023). These rely on efficient simulators, so-called semi-numerical simulations (e.g., Mesinger and Furlanetto 2007, Thomas et al. 2009, Santos et al. 2010, Mesinger et al. 2011, Visbal et al. 2012, Ghara et al. 2015, Murray et al. 2020, Trac et al. 2022, Maity and Choudhury 2022, Schneider et al. 2022), that typically approximate computationally-expensive radiative transfer (RT) with approaches based on cheap Fast Fourier Transforms (FFTs). However, inference can be computationally expensive even with semi-numerical simulations. As an example, the recent, state-of-the-art inference using nine galaxy parameters in HERA Collaboration et al. 2022a (hereafter HERA22) took $\sim 10^5$ core hours on an HPC center: roughly 10^5 likelihood evaluations each taking ~ 1 core hour to simulate the corresponding observables.

A popular alternative approach is to use emulators (e.g., Kern et al. 2017, Schmit and Pritchard 2018, Shimabukuro and Semelin 2017, Jennings et al. 2019, Ghara et al. 2020, Mondal et al. 2022, Bye et al. 2022, Lazare et al. 2023). Once trained on a set of simulation outputs, an emulator can replace the expensive, on-the-fly simulation step in Bayesian inference: a single likelihood evaluation taking ~ 0.1 s instead of ~ 1 hour. As such, the computational cost is *amortized*, requiring only the initial database of simulations in order to perform subsequent, inexpensive inferences[2]. Of course, such amortized inference is restricted to the theoretical model that is used to train the emulator. Moreover, there is also the additional emulator error to account for, which can be non-negligible for high precision measurements and in corners of parameter space that are poorly sampled (e.g., Kern et al. 2017, Appendix B of HERA22). Nevertheless, emulators allow us to rapidly perform many inferences of the same model, testing the impact on the posterior of different likelihood choices, priors, and new data. Moreover, the emulator error is sub-dominant compared with current, relatively low S/N observations, such as the 21-cm power spectrum upper limits.

[2]Another form of amortized inference is to train neural density estimators to fit the likelihood or likelihood / evidence ratio using simulated data (e.g., Papamakarios et al. 2018, Alsing et al. 2018, Cole et al. 2022). This is referred to as simulation-based inference (SBI), and has the additional benefit of not having to specify an explicit functional form for the likelihood. SBI has recently been applied to mock 21cm observations [Zhao et al., 2022b,a, Prelogović and Mesinger, 2023, Saxena et al., 2023], with very promising results.

Here we present 21cmEMU[3] – a public emulator of several summary outputs from the semi-numerical code 21cmFAST[4]. These include (i) the volume-averaged hydrogen neutral fraction; (ii) the 21-cm power spectrum; (iii) the global 21cm brightness temperature; (iv) the neutral IGM spin temperature; (v) the ultraviolet (UV) luminosity functions (LFs); (vi) the Thomson scattering optical depth of CMB photons. Our emulator was trained on summary observables from the *withHERA* inference in HERA22, which sampled nine astrophysical parameters that characterize galaxy properties. As a result, our work presents a few important improvements over previous emulators. The unprecedented number of summary outputs allows us to include complementary, multi-wavelength probes of high-z galaxies and the EoR when computing the likelihood. Moreover, our physically-motivated galaxy parametrization [Park et al., 2019] allows us to easily motivate different choices of priors. We will periodically update 21cmEMU to include new summary outputs and astrophysical models.

We showcase our emulator by re-analysing the HERA power spectrum upper limits published in HERA22. We also perform inferences including various combinations of the data, illustrating the constraining power of each probe on the posterior. One call of 21cmEMU takes ~ 0.1 s (compared to ~ 1 hour for 21cmFAST), with a typical inference finishing in a few hours.

This paper is organized as follows. In Section 2.2, we introduce the data used to train the emulator. In Section 3.4, we introduce the network and discuss its architecture, training procedure, and performance. In Section 2.4, we showcase applications of the emulator to EoR/CD inference problems. We conclude in Section 4.7. We assume a flat ΛCDM cosmology, with $(\Omega_\Lambda, \Omega_m, \Omega_b, h, \sigma_8, n_s)$ = (0.69, 0.31, 0.049, 0.68, 0.82, 0.97), consistent with results from [Planck Collaboration et al., 2020]. Unless stated otherwise, all lengths are in comoving units.

2.2 Simulated Dataset

Our datasets for training and testing are taken from the *withHERA* inference from HERA22, using an increased number of livepoints (18k). This inference used the `Multinest` [Feroz et al., 2009] sampler in 21cmMC [5] [Greig and Mesinger, 2015, 2017, 2018] with a flat prior on all astrophysical parameters within the ranges shown in all of the corner plots (e.g., Figure 2.6). The likelihood was determined by current observations of the EoR history, galaxy luminosity functions and 21cm upper limits (discussed in detail in Section 2.4.1).

We use all of the `Multinest` outputs, including both accepted and rejected samples, resulting in 1.8M parameter samples. Of these, we randomly select 1.28M for training, 183k for validation and 330k for testing. The database is standardized (subtract the mean and divide by the standard deviation of each summary statistic) before being passed into the network for training.

Our datasets are generated with the public 21cmFAST v3 code [Mesinger and Furlanetto, 2007, Mesinger et al., 2011, Murray et al., 2020]. 21cmFAST is a semi-numerical simulation code that operates under the assumption that dark matter halos host galaxies which source inhomogeneous, large-scale cosmic radiation fields. Matter density and velocity fields are generated using second-order Lagrangian perturbation theory (e.g., Scoccimarro 1998). Galaxy properties are assigned to dark matter halo fields using empirical scaling relations, following the parametrization in Park et al. [2019]. The ionizing, X-ray and soft UV cosmic radiation fields sourced by these galaxies are computed with a combination of excursion set and direct integration along the lightcone. The ionization and thermal state of the IGM gas are then tracked with a set of coupled differential equations, allowing us to compute the various observables discussed below. The HERA22

[3] https://github.com/21cmfast/21cmEMU
[4] https://github.com/21cmfast/21cmFAST
[5] https://github.com/21cmfast/21CMMC

runs that form our database assumed a simulation box length of 250 cMpc, with a 128^3 grid. For further details on the simulation code, the interested reader is directed to [Mesinger and Furlanetto, 2007, Mesinger et al., 2011, Murray et al., 2020]. Below we summarize the astrophysical parameters used as input to 21cmEMU, and the summary observables that are the corresponding output.

2.2.1 Galaxy model and astrophysical parameters

The input consists of nine parameters that characterize bulk galaxy properties. Two parameters $(f_{*,10}, \alpha_*)$ describe the stellar-to-halo mass relation (SHMR), which is a power-law for the faint galaxies (hosted by $M_h \lesssim 10^{12} M_\odot$ halos) that dominate the cosmic radiation fields at $z > 5$ (e.g., Kuhlen and Faucher-Giguère 2012, Dayal et al. 2014, Behroozi and Silk 2015, Mitra et al. 2015, Sun and Furlanetto 2016, Yue et al. 2016):

$$\frac{M_*}{M_h}(M_h) = f_{*,10} \left(\frac{M_h}{M_{10}}\right)^{\alpha_*} \left(\frac{\Omega_b}{\Omega_m}\right). \tag{2.1}$$

Here Ω_b is the universal baryon energy density (as a fraction of the critical energy density), Ω_m is the total matter (i.e., cold dark matter and baryon) energy density, and $f_* \equiv f_{*,10} \left(\frac{M_h}{M_{10}}\right)^{\alpha_*} \in [0, 1]$ is the stellar fraction, with $f_{*,10}$ corresponding to the fraction of galactic gas in stars normalized to the amount in a halo of mass $M_{10} \equiv 10^{10} M_\odot$, and α_* the power-law index.

Star formation is assumed to occur on a time-scale that goes with the Hubble time, $H^{-1}(z)$, (or analogously the dynamical time, which also scales with the Hubble time during matter domination):

$$\dot{M}_* = \frac{M_*}{t_* H^{-1}(z)}. \tag{2.2}$$

The characteristic star formation timescale, $t_* \in [0, 1]$, is another free parameter.

The typical ionizing escape fraction, $f_{\rm esc}(M_h) \in [0, 1]$ is similarly described by a power-law (e.g., Paardekooper et al. 2015, Kimm et al. 2017, Lewis et al. 2020):

$$f_{\rm esc}(M_h) = f_{\rm esc,10} \left(\frac{M_h}{M_{10}}\right)^{\alpha_{\rm esc}}, \tag{2.3}$$

with two free parameters: the normalization, $f_{\rm esc,10}$, and the power-law index, $\alpha_{\rm esc}$.

Star formation is suppressed in small mass halos due to inefficient gas cooling and/or feedback (e.g., Hui and Gnedin 1997, Springel and Hernquist 2003, Okamoto et al. 2008, Sobacchi and Mesinger 2013, Xu et al. 2016, Ocvirk et al. 2020, Ma et al. 2020). We account for this suppression by assuming only a fraction $\exp\left(-M_{\rm turn}/M_h\right)$ of halos host active star forming galaxies. The characteristic halo mass scale below which the abundance of galaxies is exponentially suppressed, $M_{\rm turn}$, is another free parameter.

The specific X-ray luminosity escaping the galaxies is also taken to be a power-law in energy (e.g., Das et al. 2017), $L_X \propto E^{-\alpha_X}$, with the index α_X left as a free parameter. The luminosity is normalized via the soft band X-ray luminosity per unit SFR, another free parameter:

$$L_{X<2\rm keV}/\rm SFR = \int_{E_0}^{2\rm keV} \rm dE\ L_X/SFR, \tag{2.4}$$

where E_0, the last input parameter, is the minimum energy of X-ray photons capable of escaping their host galaxy.

In summary, the nine input parameters are:

1. $\boldsymbol{f_{*,10}}$: normalization of the SHMR, defined at $M_h = M_{10}$.
2. $\boldsymbol{\alpha_*}$: power-law index of the SHMR.
3. $\boldsymbol{f_{\rm esc,10}}$: normalization of the ionizing escape fraction to halo mass relation, defined at $M_h = M_{10}$.
4. $\boldsymbol{\alpha_{\rm esc}}$: power-law index of the ionizing UV escape fraction to halo mass relation.
5. $\boldsymbol{t_*}$: characteristic star formation timescale, defined as a fraction of the Hubble time.
6. $\boldsymbol{M_{\rm turn}/\rm M_\odot}$: characteristic mass below which halos become exponentially less likely to host an active star forming galaxy.
7. $\frac{\boldsymbol{L_{\rm X<2keV}/SFR}}{\rm erg\ s^{-1}\ M_\odot^{-1}\ yr}$: soft-band X-ray luminosity per unit SFR escaping the galaxies.
8. $\boldsymbol{E_0/\rm keV}$: minimum X-ray energy that can escape the galaxies.
9. $\boldsymbol{\alpha_{\rm X}}$: power-law index of the X-ray spectral energy distribution (SED).

This simple parametrization is easy to interpret physically and is consistent with observations of the UV LFs as well as the scaling relations found in galaxy simulations and semi-analytic models.

2.2.2 Observational summaries

For a given set of cosmological and astrophysical parameters, 21cmFAST calculates the corresponding 3D lightcones of IGM properties. When performing inference, these lightcones are generally compressed into summary statistics that are compared directly with observations. Here we do not attempt to directly emulate the 3D lightcones of the various cosmological quantities, and instead only emulate the following summary observables (motivated by existing EoR/CD observations discussed in Section 2.4.1):

1. $\boldsymbol{\overline{x}_{\rm HI}(z)}$ — the volume-averaged neutral fraction of hydrogen and helium as a function of redshift (also commonly referred to as the EoR history).
2. $\boldsymbol{\overline{T}_{\rm b}(z)}$ — the volume-averaged (global) 21cm brightness temperature (e.g., Madau et al. 1997, Furlanetto et al. 2006, Pritchard and Loeb 2012):

$$\begin{aligned} T_{\rm b}(\boldsymbol{x}, z) &= \frac{T_{\rm S} - T_{\rm R}}{1+z}(1 - e^{\tau_{21}}) \\ &\approx 27\, x_{\rm HI}(1+\delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1+z}{10}\right)^{1/2}\ {\rm mK} \\ &\times \left(\frac{T_{\rm S} - T_{\rm R}}{T_{\rm S}}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right], \end{aligned} \tag{2.5}$$

where τ_{21} is the 21cm optical depth of the intervening gas, $\delta_b \equiv \rho/\bar{\rho} - 1$ is the baryon overdensity, with ρ being the baryon density, and $T_{\rm S}$ and $T_{\rm R}$ are the spin and background temperatures, respectively. We assume throughout that the radio background is provided by the CMB, $T_{\rm R} = T_{\rm CMB}$ is the temperature of the CMB. We note that 21cmFAST computes the brightness temperature at each cell location, $\mathbf{x}$, using the exact expression in the first line of the equation above; the second line is a Taylor expansion in the limit of $\tau_{21} \ll 1$ that provides physical intuition.

3. $\overline{T}_{\rm S}(z)$ — the mean spin temperature of the neutral IGM as a function of redshift. The IGM spin temperature is only defined for neutral hydrogen that is outside of the cosmic HII regions that surround galaxies. Specifically, the volume average is performed over those cells in the simulation box with $\overline{x}_{\rm HI} \geq 95$ %.

4. $\boldsymbol{\Delta^2_{21}(k, z)}$ — spherically-averaged 21cm power spectrum (PS): $\Delta^2_{21}(k, z)\left[\mathrm{mK}^2\right] \equiv k^3/(2\pi^2)\langle\tilde{T}_b\tilde{T}_b^*\rangle$, where $k = |\boldsymbol{k}|$, and $\tilde{T}_b(\boldsymbol{k}, z)$ is the Fourier dual of the brightness temperature from eq. (3.7).

5. $\boldsymbol{\phi(M_{1500}, z)}$ — the non-ionizing UV luminosity function (UV LF), defined as the number density of galaxies per UV magnitude, M_{1500}, as a function of redshift. The $\sim$1500 Å rest frame luminosity is calculated from the SFR: $\dot{M}_*(M_h, z) = \mathcal{K}_{\rm UV} \times L_{\rm UV}$, where $\mathcal{K}_{\rm UV} = 1.15\cdot10^{-28}\mathrm{M}_\odot\mathrm{yr}^{-1}\ \mathrm{Hz\,s\,erg}^{-1}$ assumes a Salpeter initial mass function (e.g., Madau and Dickinson 2014, Sun and Furlanetto 2016). The UV luminosity is related to the AB magnitude using [Oke and Gunn, 1983]: $\log\left(\frac{L_{\rm UV}}{\mathrm{erg\ s^{-1}\ Hz^{-1}}}\right) = 0.4 \times (51.63 - M_{\rm UV})$.

6. $\boldsymbol{\tau_{\rm e}}$ — the Thompson optical depth to the last scattering surface (LSS): $\tau_e = \sigma_T \int_0^{z_{\rm LSS}} dz \left|\frac{cdt}{dz}\right| n_e$, where σ_T is the Thompson scattering cross section and n_e is the electron number density calculated assuming hydrogen and helium are singly ionized at a fraction $(1 - \overline{x}_{\rm HI})$ and that helium is doubly ionized at $z < 3$.

Although the last two quantities are computed analytically by 21cmFAST and are therefore reasonably fast, we still emulate them for two reasons. The first is to provide users of 21cmEMU with a standalone package. The second is that the analytic calculation is still slower than the emulator prediction time: emulation reduces the runtime from $\sim$ 1 s to $\lesssim$ 50 ms for a single parameter combination ($\lesssim$ 1 ms per parameter set if in a large ($\gtrsim$ 100) batch), with a relatively low emulation error (see Section 3.4).

We use 84 redshift bins in the range $z \sim 5 - 35$ for all summaries except the 21cm PS. For the 21cm PS we exclude high redshift bins that generally have a very weak signal, keeping 60 redshift bins spanning $z \sim 6 - 21$, and 12 k bins spanning $k \sim 0.04 - 1$ Mpc^{-1}. We also floor the PS values to 0.1 mK^2, in order to reduce the dynamic range of the data and improve training. We note that the value of the floor is an order of magnitude smaller than the accuracy of the 21cmFAST simulator itself (e.g., Mesinger et al. 2011, Zahn et al. 2011), and thus has no effective impact on the accuracy of our emulator.

2.3 Emulator architecture and performance

21cmEMU is implemented using `Tensorflow` [Abadi et al., 2015, Developers, 2022], with an architecture consisting of (see diagram in Fig. 2.1):

- one large block (8 layers with 1k nodes each) of fully-connected (dense) layers whose output is fed into all of the branches.
- one branch per summary observable.

Since the 21-cm power spectrum is a smooth function of wavemode and redshift (e.g., Fig. 2.3), it can be interpreted as a 2D image. Therefore we use a convolutional neural network (CNN) in the 21-cm PS branch and fully-connected layers in the other branches. Note that the branches are not connected with one another. The only nodes they have in common are those from the main block which each branch receives as input.

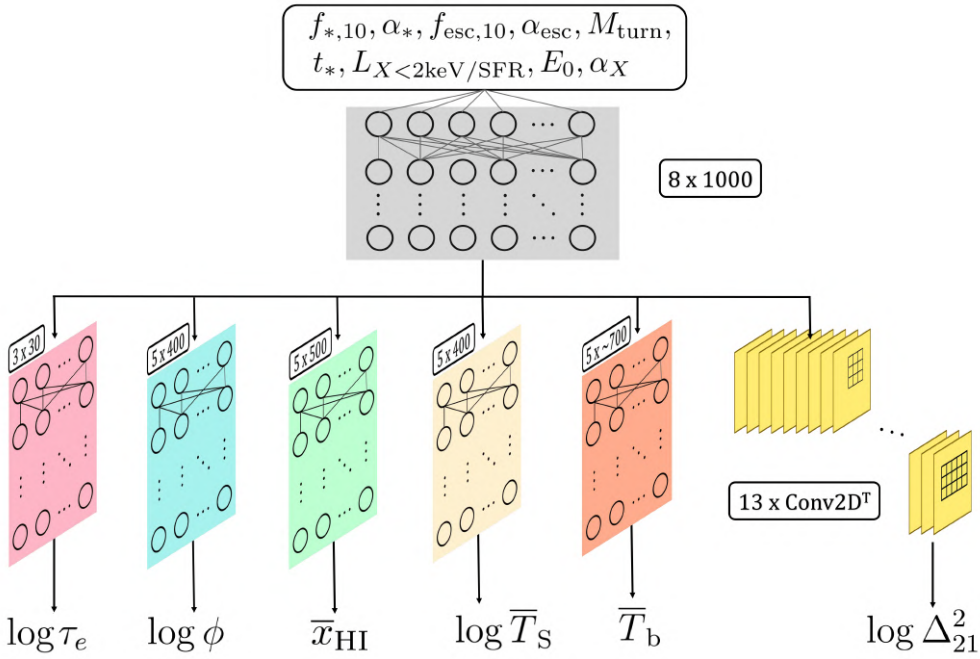

Figure 2.1: Schematic of the 21cmEMU architecture. Astrophysical parameters (*top*; c.f. Section 2.2.1 are inputted through a large block of fully-connected layers. The output from this shared block is then passed on into five blocks (much smaller than the shared block). The pink, blue, green, yellow, and orange fully-connected branches output the Thomson scattering optical depth, UV LFs, mean hydrogen neutral fraction, spin temperature, and global signal, respectively. The output from the shared block is also reshaped into an image and is passed into a 2D convolutional neutral network which outputs the 21cm power spectrum. The convolutions gradually build the PS image. The window size varies among the layers. The number of filters (yellow layers) decreases toward the end of the CNN.

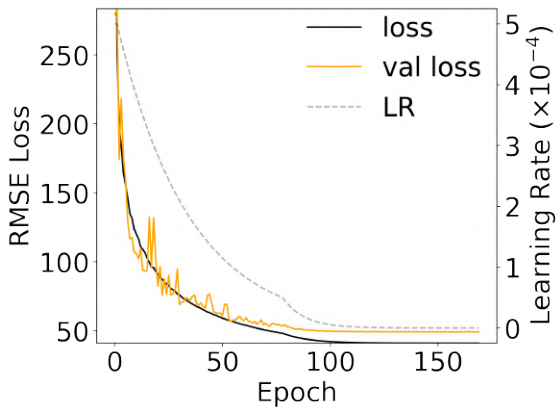

Figure 2.2: Training loss (black line) and validation loss (orange line) as a function of training epoch. The learning rate curve is also shown with the dashed gray line and the corresponding right axis.

The network trains on all of the summaries at once (i.e., multi-task learning), using a weighted sum of root mean squared error (RMSE) losses with one loss term per branch, where each branch loss has a different weight. We assign the largest weight to the 21-cm PS branch as it is higher dimensional with the largest dynamic range, and thus more difficult to learn. The final set of weights chosen is obtained from a trial of about 50 different weight combinations with the goal of choosing the best weights such that the 21-cm power spectrum, brightness temperature, and neutral fraction are learned best. The performance of the other summaries is not significantly affected by the choice of weights. These trials are done ad hoc since the training is computationally expensive.

We perform a few tests to motivate the importance of the block of fully-connected layers. First, we train a network equivalent to the brightness temperature branch alone i.e., whose input is the astrophysical parameters that are straight away passed into the brightness temperature branch of fully-connected layers. We find that the median brightness temperature fractional error over the test set in this network is $\sim 45\%$ larger than the one in the final network. [6] This means that, on average[7], our final architecture performs better than just having individual networks for each summary. The final architecture can contribute to improving the performance in two ways: (i) combining the losses of the summaries allows the network to learn from the correlations between the summaries; and (ii) simply making the network larger and deeper. To test the relative importance of (i) and (ii), we train a network without the shared block but with the rest of the architecture the same. This significantly reduces the number of trainable parameters in the network (by about 50%), but still allows different summaries to influence each other through the shared loss. We do see an increase of up to a few percent in the median and 68% CL of the fractional error for the smaller network as expected. Most notably, for the brightness temperature we see an increase of $\sim 1\%$, $\sim 13\%$, and $\sim 27\%$ for the median, 68% CL, and 95% CL of the fractional error, respectively. We conclude that combining the losses of all the summaries is the main cause of performance improvement, while the large shared block is needed to get the best performance for the most challenging summaries: the brightness temperature and 21-cm power spectrum.

In Figure 2.2, we show the total training and validation losses as a function of epoch in black and orange, respectively. We also show the learning rate schedule used during training with a dashed gray line. We see a smooth decline of the validation loss up to ~ 100 epochs. Our final network is taken at the minimum of the validation loss, at epoch 150. The training takes about eleven GPU hours ($\sim$3.5 min per epoch) with the full database (1.8M samples).

Below, we discuss the branch architecture and performance for each summary observable in turn, summarizing the results in Table 3.2. Throughout, we illustrate the emulator performance using examples from the test set, as well as the distributions of absolute differences (Abs Diff) and fractional errors (FE) over the entire test set. The latter two are defined for each observational summary, y, as:

$$\text{Abs Diff} \equiv |y_{\rm true} - y_{\rm pred}| \tag{2.6}$$

$$\text{FE}(\%) \equiv \frac{\text{Abs Diff}}{\max\left(|y_{\rm true}|, y_{\rm floor}\right)}, \tag{2.7}$$

where $y_{\rm true}$ refers to the 21cmFAST direct simulation output and $y_{\rm pred}$ is the corresponding 21cmEMU prediction. We compute the above averaged over different bins in y and/or different models in

[6] We also test slightly changing the brightness temperature branch itself: adding an additional layer and increasing the number of nodes increases the median fractional error (see below for specific definition) by about 50%, while increasing the number of layer nodes slightly and adding one additional layer increases it by about 20%.

[7] We did not perform this test for all of the other summaries. We did perform it for the 21-cm power spectrum and found that the performance of the final network is a few percent better than that of the CNN branch alone.

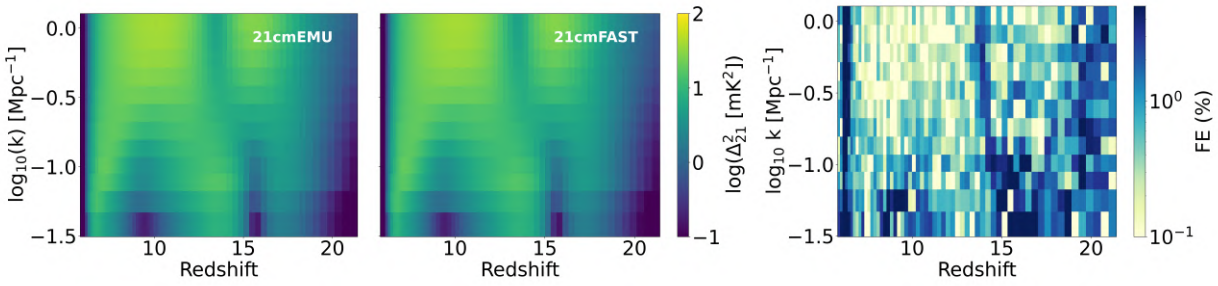

Figure 2.3: The spherically-averaged 21-cm power spectrum as a function of wavemode and redshift for a sample in the test set. The 21cmEMU prediction is shown on the left while the 21cmFAST result is on the right. This sample has a 21-cm PS FE that is roughly comparable to the median value of the whole the test set, and can thus be considered representative of the emulator performance. The rightmost panel shows the fractional error for this single sample.

the test set, as described below. One drawback of the FE metric is that it can diverge to infinity as the denominator goes to zero. To avoid this, we use floors for the values of the denominator: $\log(\Delta^2_{21,\text{floor}}) = 0.1$; $\bar{T}_{b,\text{floor}} = 5$ mK, and $\bar{x}_{\text{HI,floor}} = 10^{-4}$. The specific values of these floors was chosen relatively arbitrarily; however, they are lower than the expected accuracy achievable by any near term experiment.[8] The other summaries, τ_e, $\overline{T}_S$, and UV LFs do not have a floor value.

2.3.1 The 21-cm power spectrum

The power spectrum branch consists of thirteen 2D convolution layers with wide (up to 7 redshift bins $\times$ 3 k bins) kernels and two upsampling layers that gradually build the (k, z) PS image based on the output of the shared block, as seen in Figure 2.1. We use a pixel-based RMSE loss, weighted by the inverse of the estimated thermal noise corresponding to a 1000h SKA1-low observation (taken from Prelogović et al. 2022; for more details see Section 2.2.1 in that work). Weighting by the inverse of the noise forces the CNN to be more accurate in (k, z) bins that are easier to observe: generally corresponding to lower redshifts and larger scales.

In Figure 2.3, we compare the emulator prediction for the 21-cm power spectrum with its corresponding target from 21cmFAST. We show a single sample from the test set, with the 21cmEMU prediction on the left and the 21cmFAST target in the middle panel. This sample was chosen as it has the closest median fractional error to that of the entire test set; thus it can be considered representative of the typical emulator performance. It is difficult to see a difference between the two PS with the naked eye. We the FE of this single sample in the rightmost panel. The FE is generally sub-percent, rising to $\sim$percent in regions of low power.

In these 2D images we clearly see the well-known trend of three peaks in the redshift evolution of the large-scale 21-cm PS and two peaks in the small-scale evolution (e.g., Pritchard and Furlanetto 2007). In general, the features evolve smoothly over (k, z), showcasing why we use a CNN in the 21-cm PS branch of 21cmEMU.

We quantify the 21-cm PS prediction error in the top left panel of Figure 3.3. In the top

[8]For the 21-cm power spectrum for example, the expected mean noise level from thermal noise and sample variance for a 1000hr observation with the SKA1-low instrument is $\gtrsim$0.1 mK^2 (e.g., see Figure 2, bottom left panel in Kaur et al. 2020). Similarly, global signal experiments have measurement noise that is orders of magnitude larger than the floor value we chose (e.g. Singh et al. 2022, Murray et al. 2022), and are instead limited mostly by foregrounds and instrument systematics. For the mean neutral fraction, estimates have typical uncertainties of order 0.1 (see e.g. Greig et al. 2022 and references therein), orders of magnitude larger than the floor value we use.

sub-panel, we plot the redshift evolution of the PS amplitude at $k = 0.1$ Mpc^{-1}, with 21cmEMU predictions shown via dash-dotted lines and the corresponding 21cmFAST targets shown with solid lines. We chose to plot $k = 0.1$ Mpc^{-1} because the strongest constraints by current interferometers are around these scales; smaller scales are dominated by thermal noise and larger scales by foregrounds (e.g., Mertens et al. 2020, Trott et al. 2020, HERA Collaboration et al. 2023, 2022b). The ten models plotted here were chosen at random from the test set. We again see that the differences between the emulator and "truth" are difficult to spot with the naked eye.

In the bottom sub-panel we show the absolute difference between each pair of curves in the top sub-panel, as well as the median absolute difference (dashed black line) and the 68% / 95% confidence limits (CL; dark / light gray) computed over the entire test set. We see that the median (68%) 21cmEMU absolute error at $k \sim 0.1$ Mpc^{-1} is $\left|\log\left(\Delta^2_{21,\rm true}/\rm mK^2\right) - \log\left(\Delta^2_{21,\rm pred}/\rm mK^2\right)\right| \leq$ 0.01 ($\sim$ 0.02). This translates to a median (68%) fractional error of 0.70% (1.0%) [9] at this wavemode and 0.55% (2.4%) when averaged over all wavemodes. This is far below observational uncertainties in the near-term, thus justifying the use of an emulator. The error rises slightly at lower redshifts, owing to the broader distributions of possible PS, including very small values post reionization. In Appendix A, we show the evolution of the 21-cm power spectrum fractional error as a function of the input 9D astrophysical parameters.

2.3.2 The 21-cm global signal

The 21-cm global signal branch consists of seven fully-connected layers with 600 - 1000 nodes each. We quantify the performance of 21cmEMU on the global signal in the top right panel of Figure 3.3. We show the redshift evolution of the global signal (*top*) and absolute differences (*bottom*) for the same ten random samples from the test set.

As for the 21-cm power spectra, the difference between the 21cmFAST calculation and 21cmEMU prediction is difficult to see with the naked eye and is generally $\lesssim 1$ mK. We see from the bottom sub-panel that the 95% CL of the errors in the test set is also $\lesssim 1$ mK. This translates to a median (68%) FE of 0.34% (1.2%).

We see from both the global signal and the PS that our training set spans a wide range of heating and ionization histories. This is due to the fact that we include both accepted and rejected livepoints of the HERA22 inference in the training set, in order to have the largest dataset possible. Extending beyond the ranges of the most likely models allows 21cmEMU to generalize beyond the HERA22 posterior distribution, accurately predicting even unlikely models that, e.g., have not reionized by $z = 5$.

2.3.3 The 21-cm spin temperature in the neutral IGM

The $\overline{T}_{\rm S}$ branch consists of five fully-connected layers with 400 nodes each. We quantify the network performance on the mean 21-cm spin temperature in the right panel of the middle row of Figure 3.3. In the top sub-panel, we show ten examples of the emulated spin temperature curve (dash-dotted line) and the corresponding true curves from the test set (solid line). In the bottom sub-panel of the plot, we show the absolute error for each of the ten examples, the median for the entire test set with the black dashed line, and the 68% / 95% CL regions in shaded in dark/pale grey as a function of redshift. We can see that the absolute difference is

[9] Note that these errors are calculated on the emulator PS output which is in log space. Computing the corresponding error distributions in linear space, we obtain a median (68%) FE of 1.53% (1.94%) at $k \sim 0.1$ Mpc^{-1}, and 1.39% (3.76%) over the entire test set. Note that since we return to linear space, we do not need to apply a floor on the power spectrum in this FE calculation.

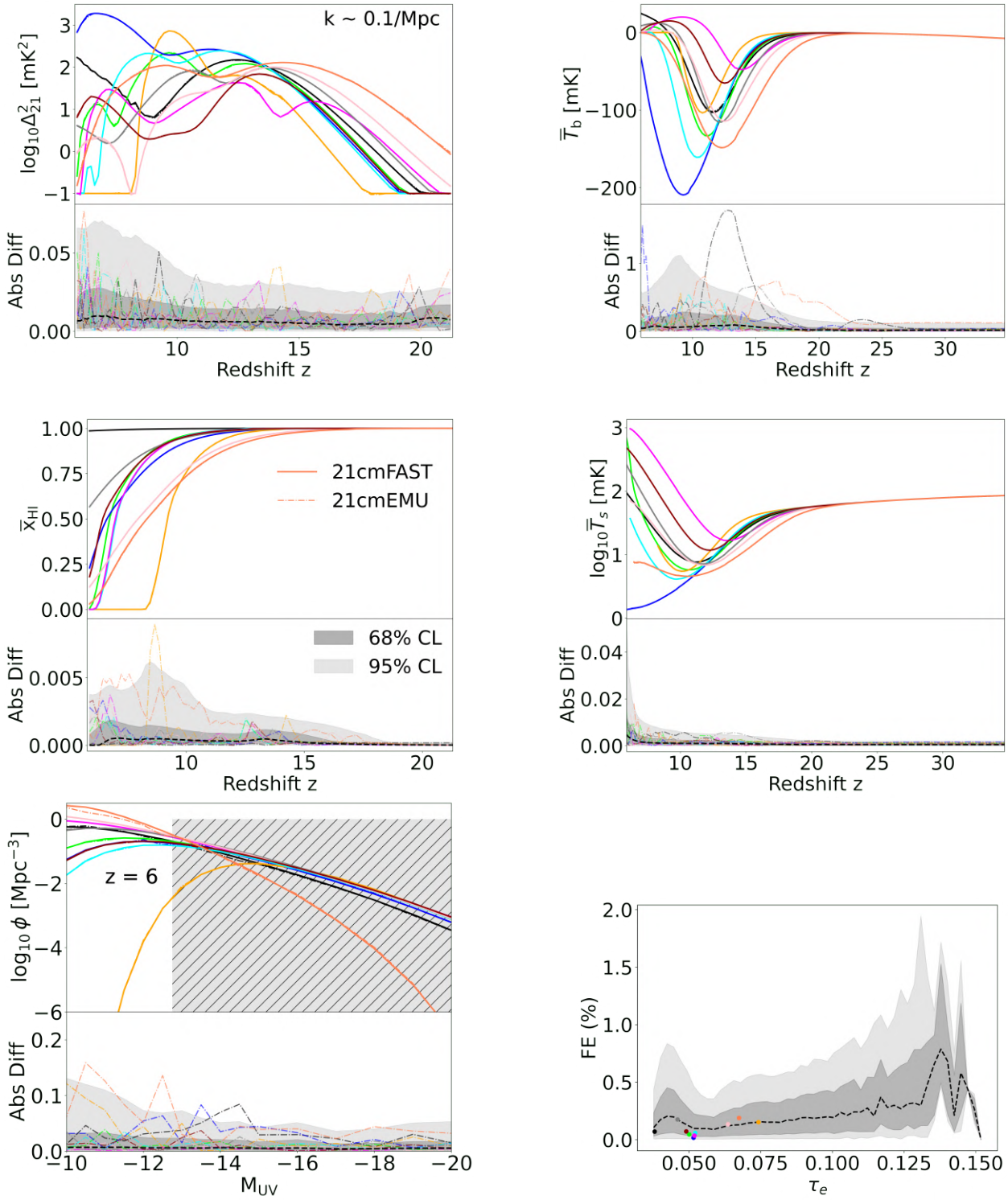

Figure 2.4: A subset of summary outputs from 21cmEMU for ten random samples from the test set. Panels show: redshift evolution of the $k = 0.1$ Mpc^{-1} 21-cm PS amplitude, redshift evolution of the mean 21cm brightness temperature, redshift evolution of the mean spin temperature in the neutral IGM, the CMB optical depth, UV LFs at $z = 6$, the EoR history (*clockwise from upper left*). Colors denote the astrophysical parameter sample with solid (dashed) lines corresponding to outputs from 21cmFAST (21cmEMU). In the bottom sub-panels, we show the absolute differences between the predicted and true quantities shown in the top sub-panels. Absolute differences of the ten random samples are shown with the corresponding colors, while the median absolute differences (FE in the case of τ_e) computed over the entire test set are shown with dashed, black curves. Dark (light) shaded regions enclose 68% (95%) CL.

$\left|\log\left(\mathrm{T_{S,true}/K}\right) - \log\left(\mathrm{T_{S,pred}/K}\right)\right|$ < 0.01 at 95% CL over most of the redshift range. The FE of the log of the mean spin temperature over the entire test set is 0.032 % and the 68% CL is 0.13 %.

We recall that the spin temperature is calculated by taking the global average over all cells in the simulation box that have $\overline{x}_{\rm HI} \geq 95\%$. When there are no cells satisfying this condition, the spin temperature becomes undefined. We account for this by having the emulator predict the redshift at which the spin temperature becomes undefined.[10] The emulator correctly predicts the exact redshift bin below which $\overline{T}_{\rm S}$ becomes undefined for 95.1% of the models in our test set, and is only one bin off ($\Delta z \sim 0.1$) for 4.89% of the models.

2.3.4 The global history of reionization

The EoR history branch, like the spin temperature branch, consists of five fully-connected hidden layers with 500 nodes each. In the left panel of the middle row in Figure 3.3, we show the EoR histories of our ten parameter samples (*top sub-panel*), and the corresponding prediction error (*bottom sub-panel*). We see that the absolute differences are $\lesssim$0.005 for 95% of the models in the test set. The FE is 0.0075% for the median and 0.095% at the 68% CL.

2.3.5 The CMB Thomson scattering optical depth

The Thomson scattering optical depth branch consists of three layers of 30 nodes each as it outputs only one number. We show the FE of the $\tau_{\rm e}$ prediction in the lower right panel of Figure 3.3. The ten parameter samples are denoted with different color dots. Over the entire test set, we see a median fractional error of 0.1% and a 0.25% FE at 68% CL. There is a notable increase in the prediction error as well as its bin-to-bin variance toward higher values of $\tau_{\rm e}$. This is due to a small number of samples in this unlikely corner of parameter space: fewer than 1% of the models in the test set have $\tau_{\rm e} > 0.11$.

2.3.6 Galaxy UV luminosity functions

The LFs branch consists of five layers of 400 nodes each. The network outputs the LFs at four redshifts ($\mathrm{z} = 6, 7, 8, 10$) and magnitude bins ranging from -20 to -10. In the lower left panel of Figure 3.3, we show the emulated and simulated LFs at $z = 6$ (the performance at the other redshifts is comparable). The hatched region denotes the range spanned by LF observations used in the inference in the following section.

We can see that the emulator is very accurate in the flat range spanned by the existing observations, while it is less accurate around the faint-end turnover. At all of the redshift bins, we have that the absolute difference $\left|\log\left(\phi_{\rm true}/\mathrm{Mpc}^{-3}\right) - \log\left(\phi_{\rm pred}/\mathrm{Mpc}^{-3}\right)\right| < 0.1$ over the majority of the magnitude range.

We provide an alternative setting in 21cmEMU that allows the user to skip the emulation and directly calculate the CMB optical depth and UV LFs using 21cmFAST. This improved accuracy however comes at the cost of a slower runtime: $\sim$ 700 ms per call compared with $\lesssim$ 50 ms using emulation.

[10] In principle, one could use the EoR history emulator prediction to find the redshift at which the volume averaged neutral fraction drops below 0.05. However, this is not identical to our definition for $\overline{T}_{\rm S}$, since our simulations account for partially neutral and self-shielded clumps inside the reionized cells. Therefore we include a separate output for the redshift at which there are no cells with $\overline{x}_{\rm HI} \geq 95\%$. We note that 21cmFAST also includes partially-ionized cells, both by UV and X-rays. Partial ionization by UV is assumed to correspond to unresolved HII regions surrounding nascent galaxies (see the discussion in Zahn et al. 2011).

2.3.7 Summary of 21cmEMU performance and context with other emulators

In Table 3.2, we summarize the performance of 21cmEMU for each summary in the first row, using the fiducial training set of 1.3M samples. In general, the median (68%) emulation fractional error is at the level of $\lesssim$ 0.5% (1%). The most accurate prediction is achieved with the EoR history, most likely due to the fact that it is a monotonic and smooth function, making it easier to learn. The least accurate summary is the power spectrum, which is understandable as it is two dimensional with the largest dynamic range.

It is difficult to directly compare the performance of 21cmEMU with other emulators of EoR/CD observables, due to their different astrophysical parametrizations and training set sizes. Nevertheless, at face value 21cmEMU's accuracy is better than achievable with state-of-the-art emulators (e.g., Mondal et al. 2020, Bevins et al. 2021, Bye et al. 2022, Yoshiura et al. 2023). For example, comparing with the recent, bespoke 21-cm global signal emulator 21cmVAE [Bye et al., 2022], we obtain a factor of 2.2 (1.5) lower median (95th percentile) RMS error (see their eq. 1). Our median 21-cm PS FE is a factor of $\sim$ 10–100 lower than that of the bespoke PS emulators in Kern et al. 2017, Ghara et al. 2020, when compared over the same redshift/wavemode ranges.

This improvement in 21cmEMU over previous works could be attributed to several factors. Firstly, we have a training set of unprecedented size: 1.3M samples. This is orders of magnitude larger than used in previous works (generally ranging from thousands to tens of thousands). We quantify how 21cmEMU's accuracy changes with the training set size in the following section.

Secondly, the improvement in power spectrum emulation could be attributed in part to our novel CNN architecture. Previous 21-cm PS emulators used only fully-connected layers which are not as efficient in processing 2D images such as the PS.

Finally, the fact that 21cmEMU emulates many different observables allows the prediction of any one of these to be helped by the others. Indeed, we verified explicitly that the 21-cm PS emulation is improved when the other summary outputs are included in the loss (i.e., when all branches are trained together). In addition to improving performance, including multiple EoR/CD observables is extremely important in the current era where 21-cm observations are not strongly constraining. As we show in Section 2.4.2, complementary galaxy and EoR observations are needed to obtain a likelihood-dominated (as opposed to prior-dominated) posterior (see also HERA22).

2.3.8 Varying the size of the training set

Since 21cmEMU was trained on an uncharacteristically-large training set, it is useful to see how it performs with smaller training sets. To do so, we remove some models at random, retrain 21cmEMU on the reduced training set, and test its performance on the same test set.

In Figure 2.5, we plot the median FE in each summary as a function of the training set size. We normalize the FE so that unity corresponds to the fiducial, 1.3 M training set. We also explicitly list the performance using half of the database (640k samples), and 1% of the database (13k samples) in the middle and bottom rows of Table 3.2.

We see that there is a sharp increase in emulator accuracy with training set size, up to a size of $\sim$ 100k. Doubling the size of the training set roughly doubles the emulator accuracy. This relationship flattens beyond sizes of $\gtrsim$ 100k, such that a ten-fold increase in the training set from $\sim$100k $\rightarrow$ 1.3 M only improves the FE by a factor of $\sim$two.

Training Size	Summary	Median FE (%)	68% CL (%)
1.3M Full	$\log \Delta^2_{21}$	0.55	2.4
	$\overline{T}_{\rm b}$	0.34	1.2
	$\log \overline{T}_{\rm S}$	0.032	0.13
	$\overline{x}_{\rm HI}$	0.0073	0.10
	$\tau_{\rm e}$	0.11	0.26
	$\log \phi$	0.50	2.1
640k Random	$\log \Delta^2_{21}$	0.71	3.0
	$\overline{T}_{\rm b}$	0.43	1.51
	$\log \overline{T}_{\rm S}$	0.047	0.17
	$\overline{x}_{\rm HI}$	0.0086	0.12
	$\tau_{\rm e}$	0.15	0.35
	$\log \phi$	0.57	2.5
13k Random	$\log \Delta^2_{21}$	3.2	13.0
	$\overline{T}_{\rm b}$	4.8	16.6
	$\log \overline{T}_{\rm S}$	0.40	1.2
	$\overline{x}_{\rm HI}$	0.035	0.57
	$\tau_{\rm e}$	0.45	1.0
	$\log \phi$	2.5	10.0

Table 2.1: Performance of the 21cmEMU network when trained on the full database, half of the database and 1% of the database.

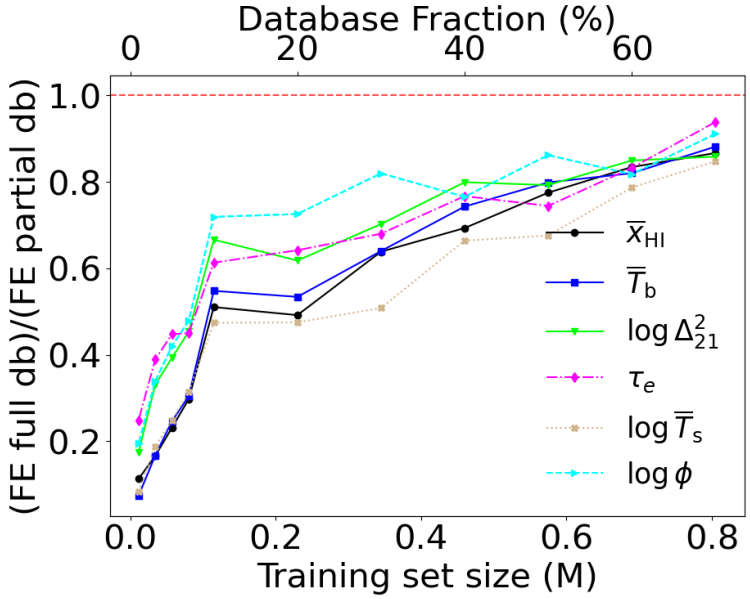

Figure 2.5: Median fractional error of each summary as a function of the training set size. The FE is normalized so that unity corresponds to the fiducial, 1.3 M training set.

2.4 Application to inference

In this section, we apply 21cmEMU to inference problems. We use the 21cmMC driver [Greig and Mesinger, 2015], which now includes the option to use either 21cmFAST or 21cmEMU as the simulator. 21cmMC incorporates three highly parallel samplers: EMCEE [Foreman-Mackey et al., 2013], Multinest [Feroz et al., 2009], and Ultranest [Buchner, 2016, 2019, 2021]; in this work we use the latter two as discussed further below.

First, we run the same inference as was previously run in HERA22 using 21cmFAST in order to see how emulator error affects the posterior. After this validation, we showcase the potential of 21cmEMU by performing several new inferences demonstrating: (i) how different observations are complimentary; (ii) the approximate impact of new, late-ending EoR constraints; (iii) the potential impact of upcoming H6C HERA observations. Each 21cmEMU inference took roughly a day on a GPU, compared with a few weeks on a cluster had we used 21cmFAST directly.

2.4.1 Comparison with direct simulation

We run the same inference as in HERA22 (10k livepoints) with 21cmEMU. Doing this allows us to directly compare the inference results between the two methods.

The likelihood in HERA22 incorporates four data sets:

1. *Thomson scattering CMB optical depth* - this term compares the Thomson scattering CMB optical depth from the proposed model with the one from the analysis of Planck Collaboration et al. [2020] by Qin et al. [2020], whose posterior is characterized by median and 68% credible interval (CI): $\tau_e = 0.0569^{+0.0081}_{-0.0086}$. The likelihood function is a two-sided Gaussian.

2. *The Lyman forest dark fraction* - this term compares the mean neutral fraction at $z = 5.9$ with the upper limit of $\overline{x}_{\rm HI} < 0.06 \pm 0.05$ at 68% CI obtained from QSO dark fraction [McGreer et al., 2015]. The likelihood function is unity at $\overline{x}_{\rm HI}(z = 5.9) < 0.06$, decreasing as a one-sided Gaussian for higher neutral fraction values.

3. *UV luminosity functions* - this term compares the model with $z = 6, 7, 8, 10$ UV luminosity functions observed with *Hubble* [Bouwens et al., 2015, 2016, Oesch et al., 2018] in the magnitude range $M_{\rm UV} \in [-20, -10]$. This likelihood term is also a two-sided Gaussian.

4. *21-cm power spectrum upper limits* - this term accounts for HERA H1C 94 night observations at $z = 8$ and $z = 10$, presented in HERA Collaboration et al. [2022b]. The likelihood is the upper limit likelihood discussed in HERA22.

These individual likelihood terms are multiplied to obtain the total likelihood. When using 21cmEMU for inference, we add the median emulator error in quadrature to the measurement uncertainties for each corresponding likelihood term.

In Figures 2.6 and 2.7, we compare posteriors obtained using 21cmFAST (*cyan*) to that using 21cmEMU (*orange*). Both were run using the `MultiNest` sampler with the same number of livepoints (10k, yielding $\sim$ 60k posterior samples). In the lower left of Fig. 2.6 we plot the 1D and 2D marginal PDFs for our astrophysical parameters, while in the top right we plot 95% CI of some of the summary observations (see caption for details). In Fig. 2.7 we plot the corresponding spin temperature PDFs in the two HERA bands, which was one of the main results of the HERA22 paper. We note that the 21cmFAST and 21cmEMU posteriors are nearly identical, testifying that the emulation error is fairly negligible when performing inference using current data sets. The only notable difference is in the t_* PDF, which is slightly broader when 21cmEMU is used as a

simulator compared with 21cmFAST. We find no notable trends of the emulator error with this parameter, concluding the small difference could be due to stochasticity in sampling and/or a higher dimensional covariance of the emulator error.

In Figure 2.6 we also include a run using 21cmEMU and the same HERA22 likelihood, but with the UltraNest sampler (*purple curves*; 5k livepoints, yielding $\sim$ 70k posterior samples). The resulting posterior is consistent with the previous two. Interestingly, the choice of sampler (purple vs orange) results in a larger difference than the choice of simulator (orange vs cyan) for some marginal PDFs. In particular, the UltraNest posterior is more accurate towards the edges of the prior range, resulting in flatter posteriors at the edges: this behavior is also recovered using the EMCEE sampler as shown in Lazare et al. [2023]. Moreover, UltraNest's vectorization makes it $\sim$ 10x faster when using an efficient simulator like 21cmEMU. Therefore, in subsequent sections we only show posteriors generated with UltraNest.

We remind the reader that the emulator was trained on the HERA22 nested sampling output. This inference took $\sim$400k core hours. Once trained however, the emulator performs amortized posterior estimation in only 225 core hours using Multinest or in 30 core hours using Ultranest.

2.4.2 Impact of different observations on the posterior

Having tested the emulator in the previous section, we now use it to perform multiple inferences that would be too costly with direct simulation.We begin by quantifying how the individual terms from the HERA22 likelihood discussed in the previous section affect the posterior. We do this by removing the terms one by one, and comparing the resulting posteriors in Figure 2.8.

In orange we show the full HERA22 posterior from the previous section, including all likelihood terms. In green, we remove the HERA power spectrum upper limit constraint. We see that the only consequence is that the L_X/SFR parameter becomes unconstrained. In the panel on the right, we can also see the 95% CI of the power spectrum and 21-cm global signal becoming wider around z $\sim 6-10$. As discussed in HERA22, the 21-cm power spectrum limits is the only measurement sensitive to the IGM temperature during the cosmic dawn.

Next, if we remove constraints on the EoR history (here corresponding to the dark fraction and τ_e likelihood terms), using only the UV LFs in the likelihood, we obtain the posterior shown in blue. We see that EoR history measurements allow us to set (lose) constraints on the ionizing escape fraction (here parametrized via $f_{\mathrm{esc},10}$ and α_{esc}), which disappear completely when their corresponding terms are not included in the likelihood. Including only the UV LFs does disfavor very early reionization, $z > 11$, because the redshift evolution of the SFR density implied by UV LF observations is too steep to allow arbitrarily early EoR, even with escape fractions close to unity.

Finally we show the prior distribution in the space of UF LFs, 21-cm PS, 21-cm global signal, and EoR history in gray. We see that all of the posteriors in these spaces are significantly broader than the priors, and are thus likelihood dominated (i.e., are not sensitive to the prior choices). Moreover, each likelihood term adds complimentary information, highlighting the importance of combining observational datasets when interpreting the high-redshift Universe.

2.4.3 Impact of late reionization

Recent observations of the large-scale opacity fluctuations in the Lyman alpha forest (e.g., Becker et al. 2015, Bosman et al. 2018, 2022) imply a late end to reionization $z < 5.6$ [Mondal et al., 2022, Qin et al., 2021]. In this section, we explore how such new EoR history constraints would

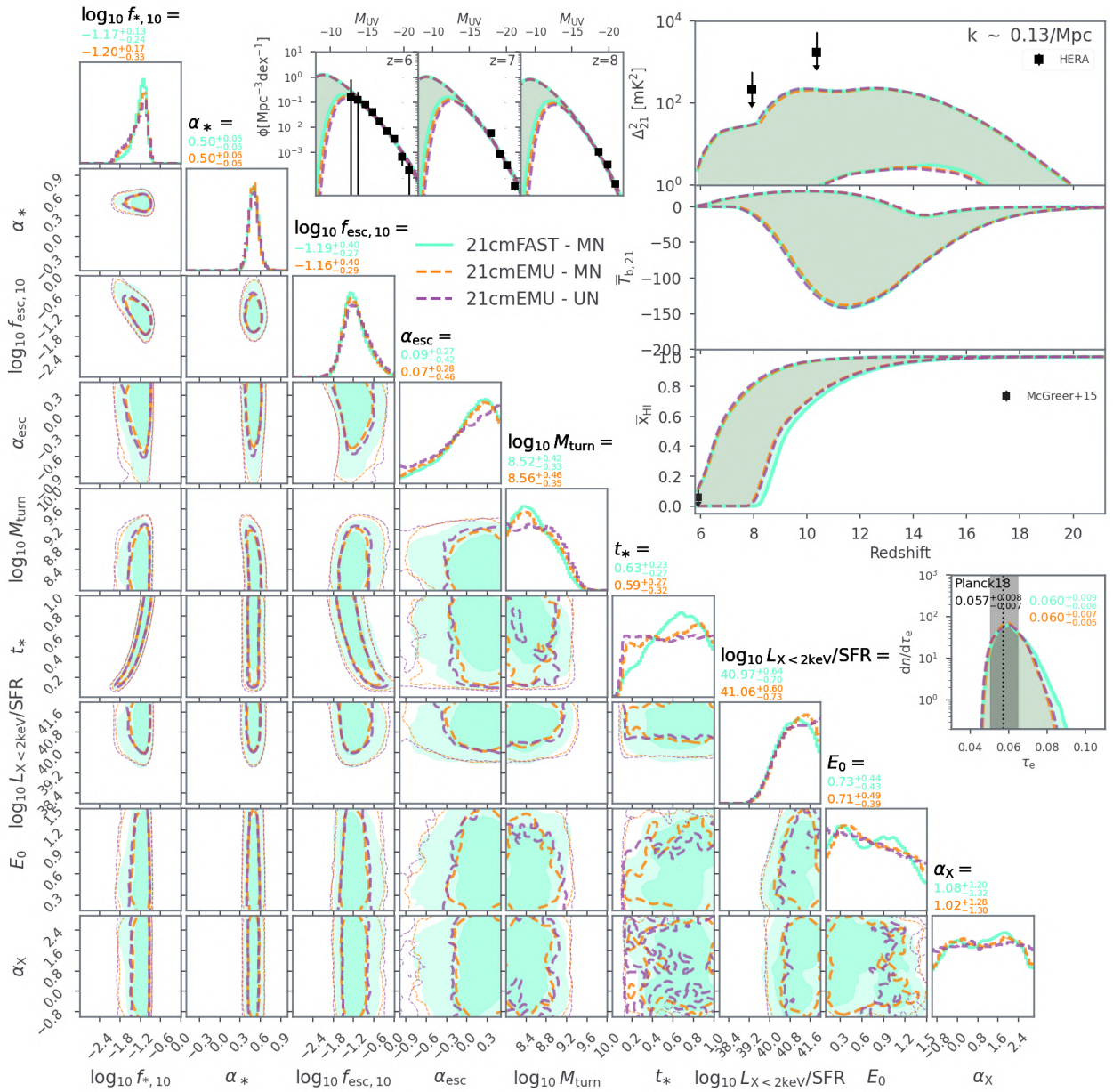

Figure 2.6: Comparison of posteriors obtained using 21cmFAST and 21cmEMU after performing an inference with the same HERA22 likelihood. The darker/dashed regions represent 68% CIs, while pale/thin dashed regions represent 95% CIs. The orange and purple posterior distributions are obtained using the `MultiNest` sampler (10k livepoints, $\sim$ 60k posterior samples), while the cyan posterior distribution is obtained using the `UltraNest` sampler (5k livepoints, $\sim$ 70k posterior samples). The median value and the 68% CIs of the 1D marginal PDFs are written above each column of the corner plot. In the panels on the top right, all highlighted regions correspond to 95% CIs. In the top middle panel, we plot the luminosity functions for redshifts 6,7, and 8. For the luminosity function likelihood, we use the data shown in black squares [Bouwens et al., 2015, 2016, Oesch et al., 2018]. In the top right, we show a panel with three summary statistics, namely the redshift evolution of the 21-cm power spectrum at $k = 0.13$ cMpc^{-1}, the 21-cm global signal and mean neutral fraction, from top to bottom. The black squares in the power spectrum plot correspond to the two deepest limits for each HERA redshift band ($k = 0.13$ cMpc^{-1} at z $\sim$ 8 and $k = 0.17$ cMpc^{-1} at z $\sim$ 10). In the bottom plot, the black square denotes the upper limit on the average neutral hydrogen fraction obtained from the QSO dark fraction [McGreer et al., 2015]. In the bottom right, we show the PDFs of the Thomson optical depth together with the *Planck* result used in the likelihood. The astrophysical parameter ranges shown in the corner plot correspond to the extent of the flat priors assumed for the inferences.

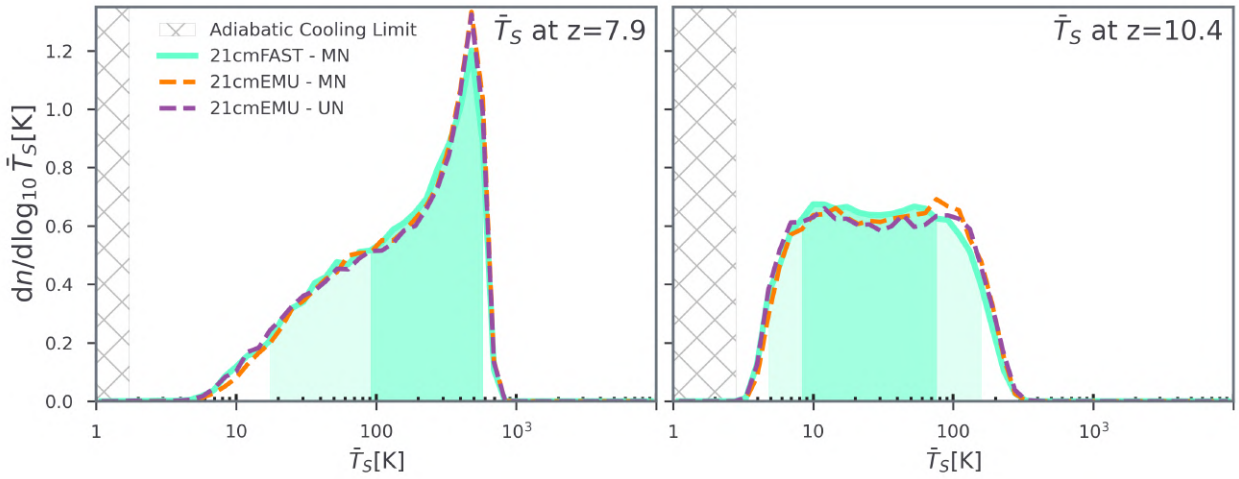

Figure 2.7: Comparison of the mean spin temperature distribution from 21cmFAST and 21cmEMU for each of the two HERA bands after performing an inference with the exact same likelihood. The credible intervals have been calculated using the highest posterior density method. The dark (light) cyan shaded region shows the 68% (95%) CI. The solid cyan line shows the distribution for 21cmFAST with 10k livepoints using MultiNest. The dashed orange line shows the same but for 21cmEMU. The dashed purple line shows the distribution for 21cmEMU but using the UltraNest sampler with 5k livepoints.

impact the previously-shown posterior. Unfortunately, the current version of 21cmEMU does not emulate the Lyman alpha forest, and so we cannot compute a likelihood directly in the observed space of Lyα opacity fluctuations. Instead we take a more approximate approach, computing the likelihood in the space of EoR histories, i.e., $\overline{x}_{\rm HI}(z)$. To construct a likelihood in this space, we use the EoR history posterior from Qin et al. [2021], who did in fact compute a likelihood from forward-modeled Lyα opacities in addition to the dark fraction and $\tau_{\rm e}$ observations. Specifically, we compute a new *Late EoR* posterior by replacing the dark fraction and $\tau_{\rm e}$ likelihood terms with a Gaussian likelihood evaluated at three redshifts, $\overline{x}_{\rm HI}(z = 6) = 0.25 \pm 0.07$, $\overline{x}_{\rm HI}(z = 7) = 0.58 \pm 0.1$, and $\overline{x}_{\rm HI}(z = 8) = 0.79 \pm 0.09$, ignoring any covariance between redshifts. Although this is obviously an approximation to computing the likelihood directly in the space of the observations, it suffices to qualitatively show the impact of new EoR history constraints.

In Figure 2.9, we show the previous (*Fiducial*) posterior in purple (71k samples) together with the new (*Late EoR*) posterior in orange (18k samples). Understandably, the corresponding recovered EoR history in orange is narrowly centered around the three points at z=6, 7, 8 used to define the likelihood. As a consequence, the posterior of the Thomson optical depth also becomes more narrow, shifting toward lower values while still being within the range allowed by Planck observations. The resulting PDF of $f_{\rm esc,10}$ for *Late EoR* is also narrower, and shifted towards smaller values. Even the power law scaling of the escape fraction with halo mass, $\alpha_{\rm esc}$, is constrained to within $\pm$ 0.3 (68% C.I.) for *Late EoR*, whereas the *Fiducial* posterior only sets a lower limit for this parameter. The remaining parameters are unaffected by the change to the Late EoR likelihood.

We also see that the recovered 21-cm large-scale PS for *Late EoR* is narrower at $z < 8$. The large-scale 21-cm PS during the EoR peaks around its midpoint (e.g., Lidz et al. 2007, Pritchard and Furlanetto 2007), which occurs at $z \sim$7–8. The HERA22 upper limits disfavor higher values of the 21-cm PS at $z \sim 8$, but the tail towards small PS values seen in the *Fiducial* posterior (corresponding to small $\overline{x}_{\rm HI}$), shrinks when moving to the *Late EoR* posterior.

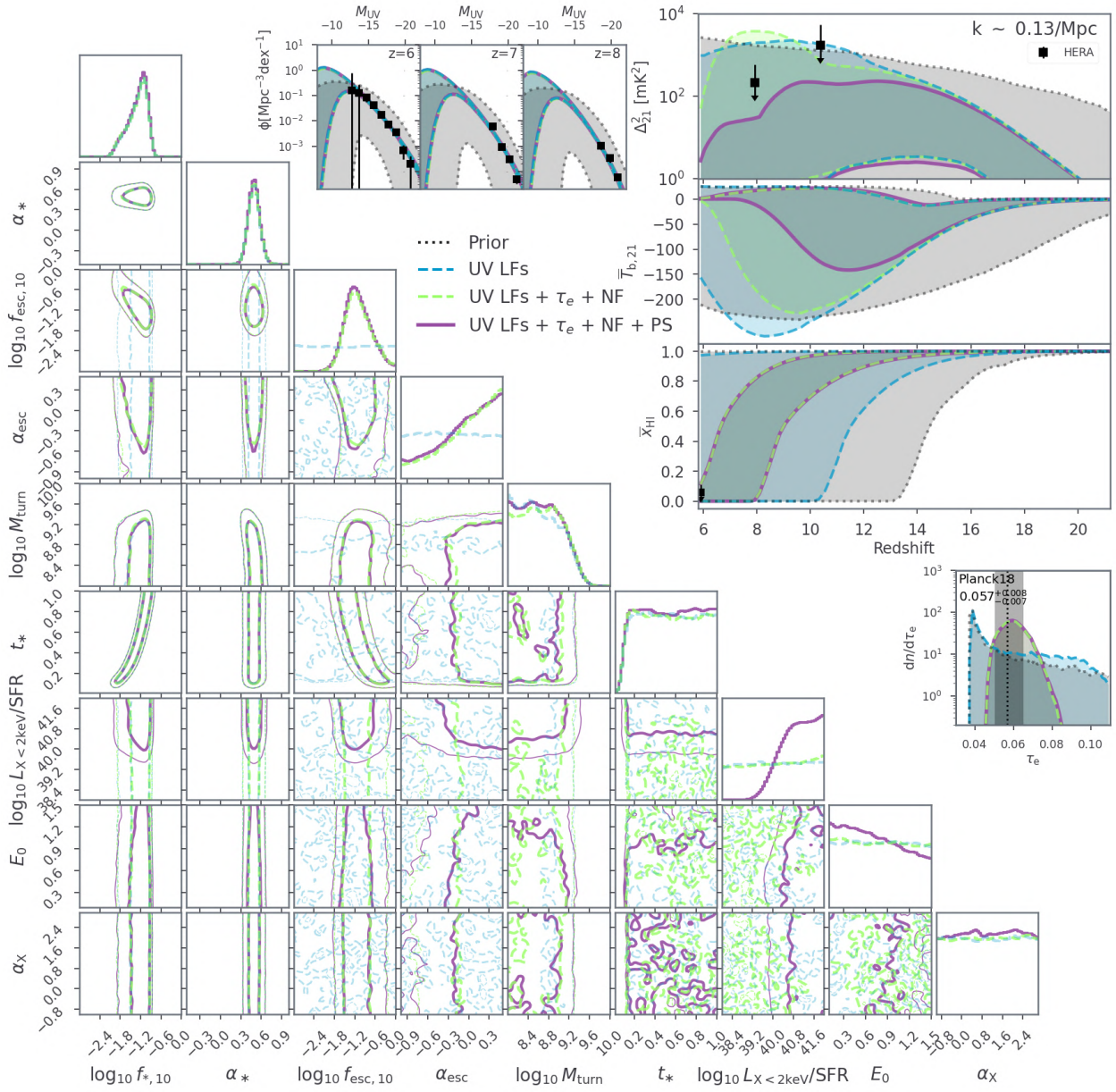

Figure 2.8: Contribution of various likelihood terms to the total posterior. The corner plot on the left shows the 95% CI of three inferences, all run with `21cmEMU` and `UltraNest`. The full posterior with all four probes is plotted in purple (exactly the same as the purple in Figure 2.6). In green, we show the posterior without the HERA power spectrum upper limits term. In blue, we additionally remove the neutral fraction and Thomson optical depth terms, leaving only the UV luminosity functions terms. On the top right half of the plot, we show the 95% CI of the same three posteriors but in the space of summary statistics: first the UV LFs, and then a panel with the 21-cm power spectrum, 21-cm global signal, and EoR history, top to bottom, and finally a panel with the Thomson optical depth. In grey, we plot the summary statistic 95% CI assuming a flat distribution across all nine astrophysical parameters which is what was used for the prior for the `21cmFAST` inference.

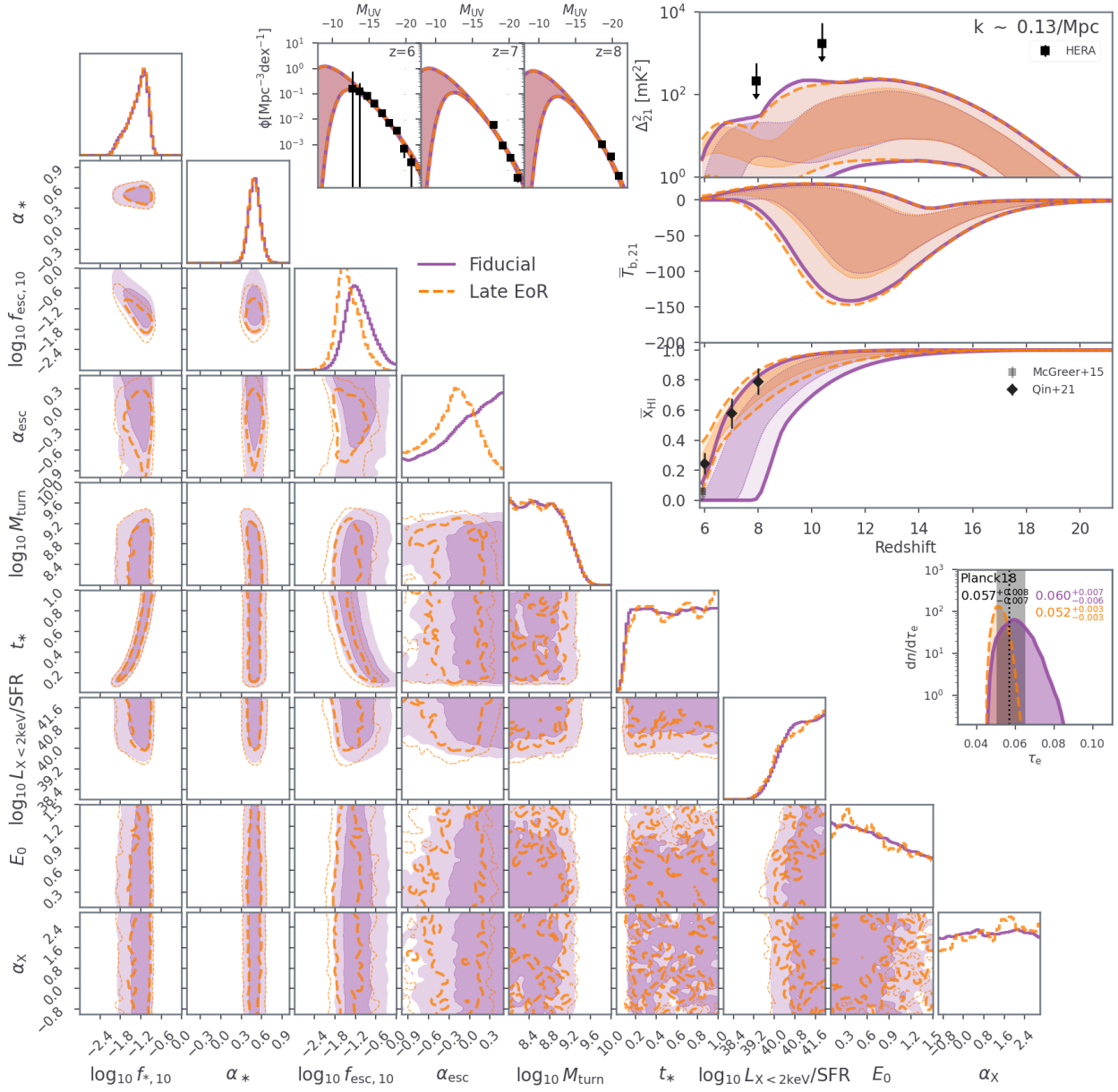

Figure 2.9: Same as Fig. 2.6, but comparing the fiducial posterior (*purple*) with one obtained by replacing the QSO dark fraction and τ_e likelihood terms with a "Late EoR" likelihood denoted by the three points with error bars in the middle right panel (*orange*. The "Late EoR" likelihood is based on the inference results in Qin et al. 2021, which included recent measurements of opacity fluctuations in the Lyman alpha forest. In the top right sub-panels, we show both the 68% (darker) and 95% (paler) C.I.

2.4.4 Forecasts for HERA Phase II sixth-season observations

We now forecast parameter constraints that could be achievable with the sixth season of HERA observations, taken in 2022-2023 (Berkhout *et al.*, *in prep.*). This season of observing used Phase II of the HERA instrument, spanning 50-230 MHz (omitting the FM band, 90-110 MHz), expanding coverage to Cosmic Dawn and late reionization with respect to Phase I (which was used for HERA22). While analysis of this season's data is ongoing, its broad characteristics are known [Dillon and Murray, 2021]: approximately 1300 hours of unflagged data over ~ 150 nights, with an average of ~ 148 un-flagged antennas per night. Although further flagging will certainly occur during processing, this dataset will be HERA's most sensitive data release to date, by a significant factor.

We create a mock observation corresponding to this upcoming dataset. For the "true" cosmic signal, we use the Evolution of Structure (EOS) 2021 release [Muñoz et al., 2022]. EOS2021 is a large simulation (1.5 cGpc per side with 1000^3 cells) made with 21cmFAST, with the goal of being our current "best guess" for the true cosmic signal. Although it used the same parametrization for galaxy scaling relations as is used here (see Section 2.2.1), the physical model of EOS2021 has a few notable differences. Instead of leaving $M_{\rm turn}$ as a free parameter, EOS2021 explicitly calculated a local $M_{\rm turn}(\mathbf{x}, z)$ based on feedback from the local ionizing and photo-disassociating backgrounds, as well as the relative velocities of baryons and dark matter. Furthermore, EOS2021 explicitly accounted for putative PopIII star formation in the first, H_2-cooled galaxies (MCGs, e.g., Tegmark et al. 1997, Abel et al. 2002, Bromm and Larson 2004, Haiman and Bryan 2006), which dominated the background radiation fields at $z > 16$, for their fiducial parameter choices. As a result of the models being different, 21cmEMU could result in a biased recovery of the EOS2021 signal; we quantify this below.

We use 21cmSense[11] [Pober et al., 2013, 2014] to obtain thermal and sample variance estimates of the HERA 6th season data, and describe our methodology and assumptions in App. A.1. We consider our sensitivity estimate to be realistic, with a few important caveats, for example the potential over-estimation of sensitivity when treating 'similar' baselines as identical [Zhang et al., 2018]. The largest unpredictable caveat is of course the presence of instrumental systematics, for which we describe our approach in more detail below.

Radio telescopes, including HERA, impose their own signature on observations – dependent on the primary beam attenuation, antenna layout, channelization and other instrumental characteristics. The effect of this instrumental signature on the observed power spectrum is such that neighbouring Fourier modes are linearly 'mixed' via a 'window function' matrix (e.g., Liu and Tegmark 2011, Gorce et al. 2023). We calculated this window function using the hera_pspec[12] package. We did not use the exact HERA beam as in Gorce et al. 2023. Instead, we used the Gaussian beam approximation which we deemed sufficient for this forecast (see Figure 7 in Gorce et al. 2023 for a comparison). Once we obtain the HERA window function, we matrix multiply it with the emulated model to properly compare with the forecast.

We perform inference using the EOS2021 cosmological signal with the sensitivity estimates from 21cmSense as the mock observation (see Fig. A.3 in Appendix A.1). This inference takes about 30 GPU h to run to convergence with UltraNest. In Fig. 2.10 we show the resulting posterior (*HERA 6th season* in orange) together with the previous result (*Fiducial* in purple). In the top right panel we show the mock PS at $k \sim 0.16 \text{Mpc}^{-1}$ as orange points with associated error bars. We see that based purely on the available S/N, the HERA sixth season data have the potential to detect the cosmic PS during the EoR ($6 < z < 8$). The 95% CI of the inferred

[11] https://github.com/rasg-affiliates/21cmSense

[12] https://github.com/HERA-Team/hera_pspec

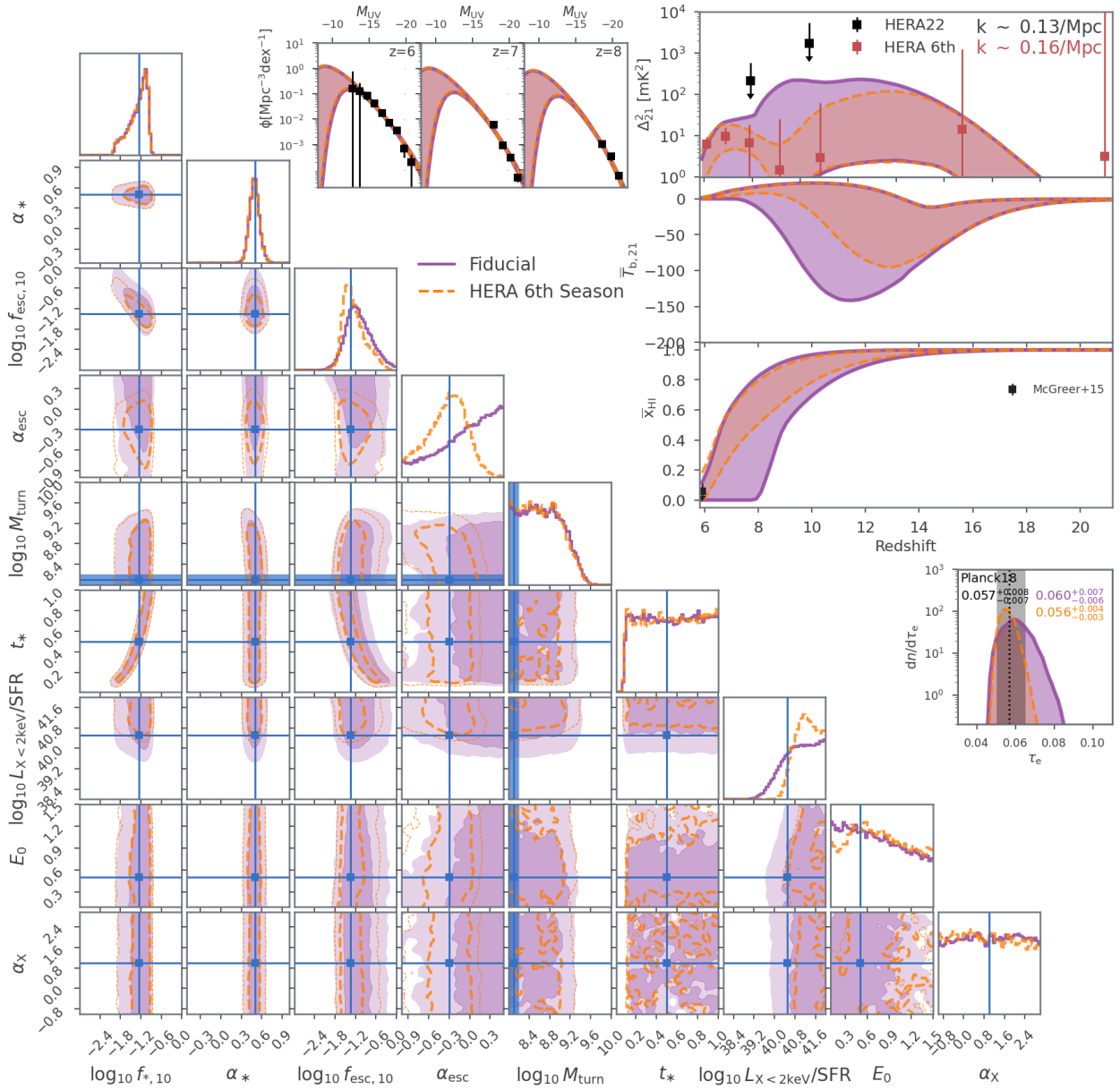

Figure 2.10: Forecasted constraints from mock HERA Phase II season six observations (see text for details) are shown in orange. The mock PS amplitudes at $k \sim 0.16\,\mathrm{Mpc}^{-1}$ are shown as orange points with error bars in the top right panel, together with current upper limits from HERA22. The parameters of the cosmological simulation used for the mock observation, EOS2021, are denoted with blue lines in the corner plot. We caution that the theoretical model used in EOS2021 and that used in 21cmEMU are somewhat different, as discussed in the text. As $M_{\rm turn}$ in EOS2021 evolves with redshift (see Fig. 5 in Muñoz et al. 2022), here we demarcate its range during the EoR (i.e., $6 < z < 8$ where the mock observations imply a detection). For more details about the panels, see the legend in Fig. 2.6.

PS (*orange*) go tightly around the data points. This unbiased recovery is reassuring, given the above-mentioned differences in the theoretical models used for the mock and forward-modeled data. Indeed, most of the "true" astrophysical parameters from EOS2021 (denoted with blue lines in the corner plot) are consistent with the recovered orange posteriors. Parameters governing X-ray heating, $L_{\rm X<2keV}/{\rm SFR}$ and E_0, are recovered with the lowest accuracy, with the true values residing outside of the 68% CI of their 2D PDF. This is understandable, because the 21cmEMU forward models do not include the additional radiation from H_2-cooling galaxies, which dominate the X-ray heating at $z > 16$.

Comparing to current constraints (*Fiducial* posterior in purple), we see that that HERA 6th season data have the potential to drastically improve our knowledge of the EoR. The *HERA 6th season* EoR history $\overline{x}_{\rm HI}(z)$ is constrained to within $\pm$0.1 (95% C.I.): a factor of $\gtrsim 3$ improvement over current limits. As a result, we can place much stronger constraints on the characteristic ionizing escape fraction, $f_{\rm esc,10}$, and its dependence on galaxy mass, $\alpha_{\rm esc}$, which are almost completely unknown currently.

It is important to note that these two posteriors use a different form for the likelihood. For the *HERA 6th season* forecast, we assume that there are *no residual systematics* after processing of the HERA data. This is in contrast to the previous likelihoods, which assume that each k-mode has a positive systematic whose prior amplitude is uniform and unbounded [cf. HERA Collaboration et al., 2022b]. In practice, assuming no residual systematics results in a two-sided Gaussian likelihood, corresponding to a 'detection', whereas the traditional likelihood has been a one-sided error-function corresponding to an 'upper-limit'. We make this choice as it is not straightforward to sample from the unbounded uniform prior for systematics when creating the mock data for the forecast. The resulting tighter parameter posteriors for the new data are therefore the result of an admixture of the new more sensitive data *and* the (effectively) more constrained priors on systematics.

2.5 Conclusion

Here we introduced 21cmEMU: a publicly-available emulator of several summary observables from 21cmFAST. We trained the emulator on 1.3M pseudo-posterior samples from the inference in HERA22. The input consists of a nine parameter model characterizing the UV and X-ray outputs of high redshift galaxies. The output consists of: (i) the 21-cm power spectrum as a function of redshift and wavemode; (ii) the IGM mean neutral fraction as a function of redshift; (iii) the UV luminosity function at four redshifts 6, 7, 8, and 10; (iv) the Thompson scattering optical depth to the CMB; (v) the mean spin temperature as a function of redshift; and (vi) the 21-cm global signal as a function of redshift. The emulator predicts all of these quantities with under $\sim 2.4\%$ error at 68% CL, and a computational cost that is lower by a factor of $\sim$10000 compared to 21cmFAST.

We varied the size of the training set, finding only a modest decrease in performance (a factor of $\sim$2 decrease in the FE) as the number of samples was reduced from 1.3M to $\sim$100k. Below $\sim$100k samples, we saw a sharp drop in performance, with the fractional error increasing roughly as the inverse of the size of the training set.

We validated the emulator's performance in inference by comparing the posteriors obtained with 21cmEMU vs 21cmFAST using the same likelihood (taken from HERA22). We found a very modest difference between these two posteriors, further illustrating that the emulator error is negligible when performing inference using current data.

Next, we profited from the speed of our trained emulator to perform multiple inferences that would otherwise be very costly using direct simulation. First, we studied the constraining power of each term in our fiducial likelihood. We found that current observations are very complemen-

tary, with UV LFs constraining the stellar-to-halo mass relations, EoR history probes constraining the ionizing escape fraction, and the addition of 21-cm PS upper limits constraining the X-ray luminosity to SFR relation.

We also explored the impact of new EoR history constraints, driven by opacity fluctuations in the Lymanα forest. These recent observations imply much tighter constraints on the EoR history, finishing at $z < 5.6$ (e.g., Qin et al. 2021, Mondal et al. 2022). The inclusion of these new limits tightened the recovered constraints on the ionizing escape fraction and its scaling with halo mass. The impact on other parameters was modest.

Finally, we presented forecasts of parameter constraints achievable with ongoing 6th season phase II observations with the HERA telescope. Optimistically, we could expect a detection of the 21-cm PS at $z \sim$6–7. This would result in a large improvement in the recovered timing of the EoR, allowing us to infer $\overline{x}_{\rm HI}(z)$ to within $\pm$ 0.1 (95% C.I.): a factor of $\gtrsim 3$ improvement over current limits. As a result, we could place stronger constraints on the characteristic ionizing escape fraction and its dependence on galaxy mass, which are almost completely unknown currently. We cautioned however that this forecast is optimistic, mainly because it assumed there are no residual systematics in the processed data (see Appendix A.1 for more details).

21cmEMU was trained on a database of summary observables where only one seed i.e., random set of initial conditions is available per combination of astrophysical parameters. In the future, we hope train the emulator on a database that samples many different seeds in order to emulate a full likelihood function rather than only approximate the mean as we do right now. This is important since Prelogović and Mesinger [2023] showed that using a single random seed when forward modeling can bias the inference results.

We make 21cmEMU publicly-available at https://github.com/21cmfast/21cmEMU, and include it as an alternative simulator in the public 21cmMC[13] sampler. We will periodically release updated versions, trained on the latest galaxy models and expanding the choice of summary outputs.

Acknowledgements

We thank A. Liu and J. Dillon for useful comments on a draft version of this paper. We also thank A. Gorce for helpful discussion concerning the HERA window function calculation. We gratefully acknowledge computational resources of the Center for High Performance Computing (CHPC) at SNS. A.M and R.T. acknowledge support from the Italian Ministry of Universities and Research (MUR) through the PRO3 project "Data Science methods for Multi-Messenger Astrophysics and Cosmology", as well as partial support from the Fondazione ICSC, Spoke 3 “Astrophysics and Cosmos Observations”, Piano Nazionale di Ripresa e Resilienza Project ID CNoooooo13 “Italian Research Center on High-Performance Computing, Big Data and Quantum Computing” funded by MUR Missione 4 Componente 2 Investimento 1.4: Potenziamento strutture di ricerca e creazione di “campioni nazionali di R&S (M4C2-19)” - Next Generation EU (NGEU). S.M. has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101067043. Y.Q. is supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project # CE170100013. R.T. acknowledges co-funding from Next Generation EU, in the context of the National Recovery and Resilience Plan, Investment PE1 – Project FAIR “Future Artificial Intelligence Research”. This resource was co-financed by the Next Generation EU [DM 1555 del 11.10.22].

[13] https://github.com/21cmfast/21CMMC

Data Availability

The trained emulator is on a publicly accessible github repository, as well as available for install as a Python package using `pip`.

Appendix A

Parameter space dependence of the 21-cm PS emulation error

In this appendix, we look at how the emulation error is distributed over the 9D input parameter space. In Figure A.1, we show the 21-cm power spectrum test set fractional error as a 2D histogram as a function of each pair of input astrophysical parameters. On the diagonal, we show the histogram (probability density) of each astrophysical parameter in the test set.

As expected, the emulation error peaks at the edges of parameter space where the density of samples is the lowest (see also Fig. 9 in Kern et al. 2017 and top plot in Fig. 18 in HERA Collaboration et al. 2022a). However, the inclusion of the rejected livepoints in the training allowed our emulator to generalize well beyond the peak of the posterior (c.f. Fig. 2.6). Importantly, the mean FE remains modest ($\lesssim 2\%$) throughout the prior volume.

A.1 21cmSense Sensitivity Estimates for HERA's 6th-Season

To obtain mock error estimates for the forecasted sixth season of HERA observations, we used the updated open-source `21cmSense`[1] tool. The general algorithm of `21cmSense` can be found in Pober et al. [2014] and in the extensive documentation and tutorials of the updated codebase[2] (see also Liu and Shaw [2020] for a review including a similar argument). A brief outline of the calculations is as follows: `21cmSense` estimates thermal noise on any 3D $\vec{k}$-mode as

$$P_N(\vec{k}_\perp, k_{||}) \propto \frac{T_{\rm sys}^2}{N_{k_\perp} \Delta\nu \tau_{\rm int}} \xi(\vec{k}_\perp, k_{||}), \tag{A.1}$$

where $T_{\rm sys}$ is frequency-dependent system temperature

$$T_{\rm sys} = T_{\rm sky}(\nu) + T_{\rm rcv}(\nu), \tag{A.2}$$

$\Delta\nu = 122.07\,$kHz is the channel width of the observation and $\tau_{\rm int} = 300\,$sec is the coherently-averaged LST-bin size used in the analysis[3]. Furthermore, ξ is a 'flag' function that takes the value 0 or 1 depending on the location of the 3D mode with respect to the foreground wedge (see below).

[1] https://github.com/rasg-affiliates/21cmSense.

[2] e.g., https://21cmsense.readthedocs.io/en/latest/tutorials/understanding_21cmsense.html

[3] In general, `21cmSense` uses the more fundamental snapshot integration time of the instrument, and re-phases observations over a longer 'coherent observation duration', however HERA is a drift-scan telescope that performs no re-phasing, and all observations within an LST bin are considered coherent without re-phasing.

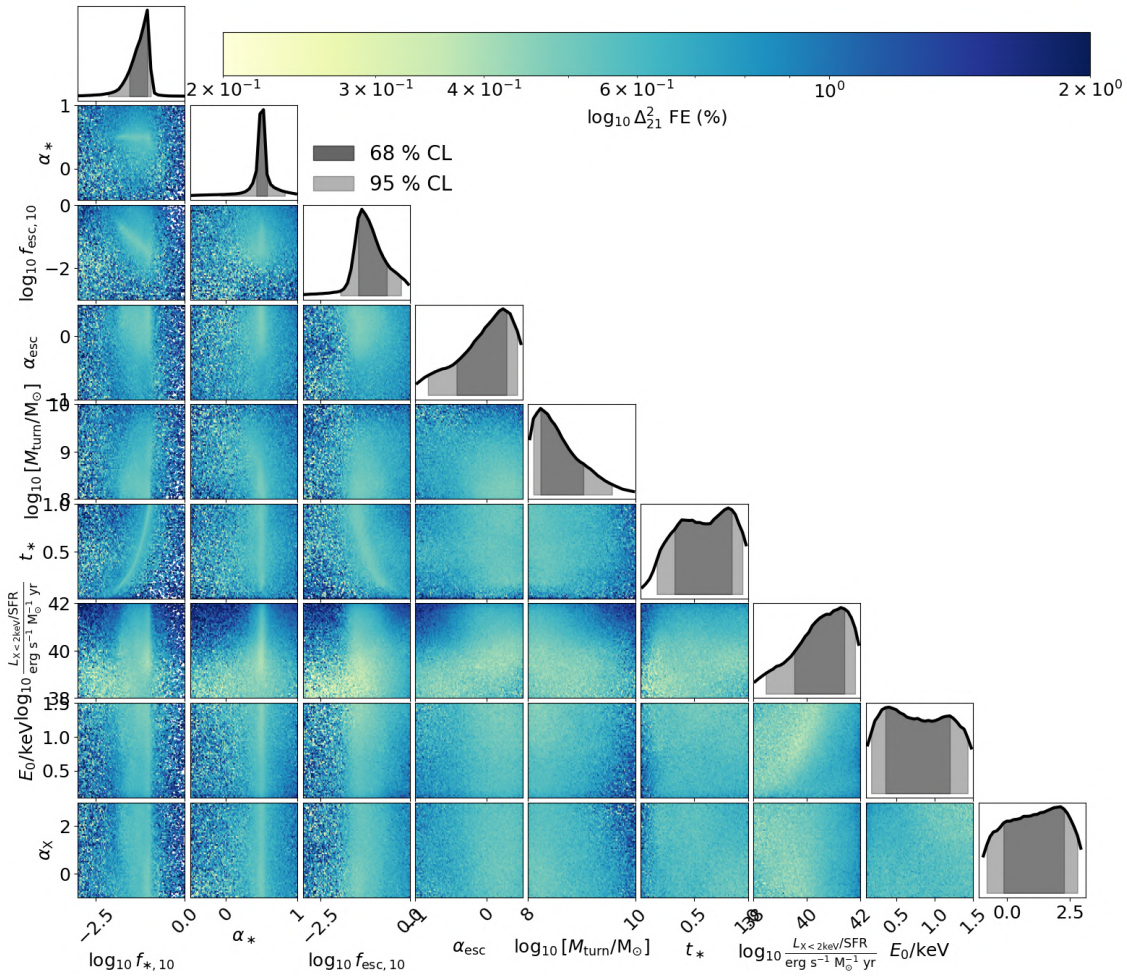

Figure A.1: Distribution of the mean fractional error of the emulated $\log \Delta^2_{21}$. The colour of each bin in the 2D histogram is a weighted mean of the fractional error of the samples therein. On the diagonal, we show the 1D marginal density distribution of each astrophysical parameter in the test set. Note that the range of astrophysical parameters in the corner plot corresponds to the ranges taken for the flat prior of the inference used to generate the database.

In this equation, $N_{k_\perp}$ represents the number of samples of this angular scale observed *coherently* throughout the observing season (i.e., observations that are averaged together as visibilities). In `21cmSense`, this is estimated by creating a grid on the $\vec{k}_\perp$ plane, whose cells are the size of the primary beam of the instrument in Fourier-space (for HERA, this is 7λ at 150 MHz), and binning the the baseline coordinates into this grid[4]. In addition to the number of samples in a bin coming from different (redundant) baselines, we also have samples from the same baseline at different *times*. Here, samples at the same LST on different nights are averaged coherently, but samples at different LSTs are averaged *incoherently* (i.e., after forming power spectra). Currently, `21cmSense` only has support for specifying the number of nights observed and the number of hours observed each night (thereby specifying the number of LST bins in conjunction with the LST bin duration). However, in realistic observational programs, the same LST bins are not observed each night (whether due to the evolution of the sky throughout the season, or through flagging / data quality concerns). To partially account for this, we define a function $n_{\rm obs}(\rm LST)$ which counts the number of unflagged days observed over the season for any given 300-second-long LST bin (note that this accounts for flags of the entire observation, due to things like poor weather or correlator malfunctions, but not antenna- or channel-specific flags). To map this non-constant function of LST bin onto the schema available in `21cmSense`, which assumes the same LST bins are observed each night, we set

$$n_{\rm days,eff} = \sqrt{\frac{\sum_{\rm LST}^{n_{\rm LST}} n_{\rm obs}^2(\rm LST)}{n_{\rm LST}}} \tag{A.3}$$

and $t_{\rm day} = n_{\rm LST} \times 300\,\rm sec$. This achieves the same resulting thermal noise level, under the assumption that the sky temperature is constant over the LST bins. We use actual sixth-season HERA measurements for $n_{\rm obs}$, as shown in Fig. A.2. We calculate $n_{\rm days,eff} = 67.4$ for coherent averaging and $t_{\rm day} = 21\,hr$ for incoherent averaging (i.e., the thermal noise from our observing pattern is equivalent to observing 253 300-second LST bins each for 67.4 days). Finally, we apply a further factor of $\epsilon = 0.9$ to $n_{\rm days,eff}$ to broadly account for finer-scale flags applied during analysis that are unaccounted in the LST-bin observing pattern of Fig. A.2. In summary, we have

$$N_{k_\perp} = \epsilon N_{{\rm bl},k_\perp} N_{\rm days,eff} \sqrt{n_{\rm LST}}. \tag{A.4}$$

The line-of-sight modes observed depend on the channel width, as already defined, and also the bandwidth of the observation. While HERA Phase II observes 200 MHz of bandwidth from 50 - 250 MHz, power spectra are estimated in smaller 'spectral windows' whose size is determined by a number of factors. Chiefly, the windows are as wide as possible, so as to include the largest scales where the signal is strongest, but are constrained by lightcone evolution [Trott, 2016, Datta et al., 2012, Greig and Mesinger, 2018] to be effectively smaller than $\sim 10\,$MHz. In practice, spectral windows are chosen to lie between strongly flagged channels (eg. FM band and Orbcomm), which means their width is not constant. Here, we use constant 20 MHz spectral windows, where we assume a Blackman tapering function is applied to each window to reduce the effective bandwidth to $\sim 10\,$MHz (and an appropriate scaling factor of 1.737 is applied to the final noise level). We calculate noise estimates for all spectral windows between 50-250 MHz, excluding the FM band between 90-110 MHz.

We use a model for $T_{\rm sky}$ that is a power-law in frequency, with amplitude and spectral-index obtained from simulated auto-correlations of the diffuse sky, using the GSM [de Oliveira-Costa

[4]This is probably the greatest departure from the actual HERA analysis, which coherently averages only *redundant* baselines, i.e., those that are equivalent to within several centimeters.

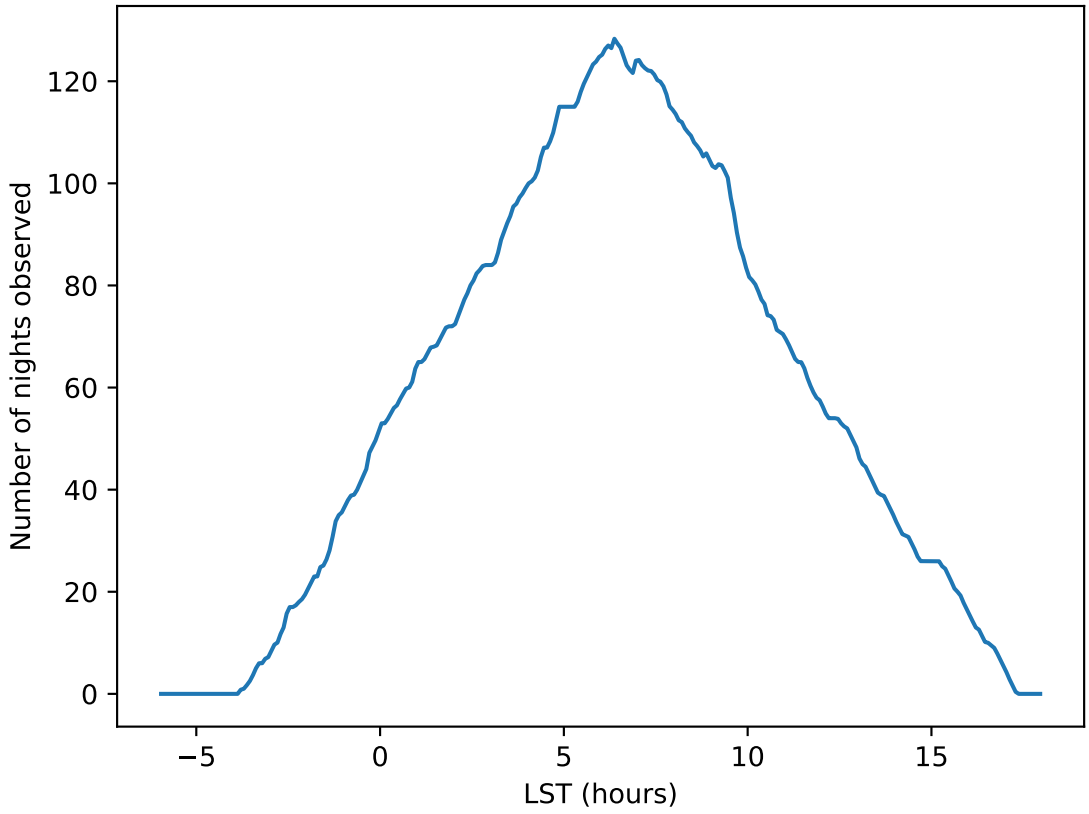

Figure A.2: The number of times each 300-second LST bin was observed and un-flagged in HERA's sixth season, used for sensitivity estimates. Note that this accounts only for flags arising from strong effects that affect large swathes of the observed antennas and/or channels (eg. lightning storms, correlator outages), and further flags are applied in the downstream analysis.

et al., 2008] and the HERA Phase II primary beam [Fagnoni et al., 2021] at LST=7 hours:

$$T_{\rm sky} = 150\,{\rm K} \times \left(\frac{\nu}{150\,{\rm MHz}}\right)^{-2.5}. \tag{A.5}$$

Currently, `21cmSense` is not able to use different sky models for different LST-bins, so this choice represents the temperature for the most-observed LST bin. For $T_{\rm rcv}$, we use a frequency-dependent model based on electromagnetic simulations performed in Fagnoni et al. [2021], interpolated by a cubic spline. This model is close to a power-law at low frequencies, with an amplitude of $\sim 600\,{\rm K}$ at 50 MHz and asymptoting to a const $\sim 60\,{\rm K}$ by 200 MHz.

We construct several estimates of the noise based on different effective observing arrays. The sixth season of HERA data observed with a maximum of 209 antennas simultaneously in any given night (of the total 350 antennas available). The bulk of these antennas observed consistently throughout the season, though a fraction of them were swapped in and out. In our estimates here, we assume that the same antennas observe consistently throughout the season, which is a reasonable approximation. Nevertheless, in practice, even though 209 antennas are being correlated at any given moment, some fraction of them are flagged over all channels (e.g., due to swapped polarizations, non-redundancies from physical effects such as feed displacement, or X-engine failures that affect a subset of antennas, etc.). The average number of antennas actually observing per-night throughout the season is as-yet unknown, though initial estimates place it at $\sim$ 150 antennas [Dillon and Murray, 2021]. Here we use $N_{\rm ants} = 148$, where the antennas are drawn randomly from the set of 209 antennas that actually observed throughout the season. In all cases, we use only baselines whose East-West length is greater than 15 m (i.e., we exclude North-South baselines, as their systematics are more difficult to filter out), and also only include baselines shorter than 150 m, similar to analyses of previous HERA seasons.

After obtaining the 3D sensitivity grid, we incoherently average into 1D spherical $|k|$-shells with bins of width $\Delta k_{||}$. In this process, we flag out $(|k_\perp|, k_{||})$-bins within the foreground 'wedge' [Liu et al., 2014a,b], defined by

$$k_{||}^{\rm wedge} = 0.15 h{\rm Mpc}^{-1} + \frac{{\rm d}k_{||}}{{\rm d}\eta}(\nu)\frac{|b|}{c}, \tag{A.6}$$

with $|b|$ the baseline length (in meters) corresponding to a given $k_\perp$, and $dk_{||}/d\eta$ a redshift-dependent cosmological factor converting bandwidth into cosmic distance. This corresponds to the 'horizon' limit of foregrounds in delay-space, plus a conservative buffer of 0.1 h/Mpc (corresponding to the buffer used in first-season HERA analyses).

In addition to the thermal variance, cosmic- (or sample-) variance is added, proportional to a fiducial cosmological power spectrum, $P^2_{\rm theory}$ divided by the number of LST bins and $k_\perp$-modes in a spherical shell. We note that using the number of LST-bins is inspired by the idea that LST-bins should capture the entire duration of 'coherence', equal to roughly the beam-crossing time for an antenna. However, HERA is conservative in using shorter coherence times, resulting in many more LST-bins. This reduces the thermal sensitivity, but artificially reduces the cosmic variance estimated by `21cmSense`. Nevertheless, since cosmic variance is generally a sub-dominant contribution to the total variance, this should not have a large effect on the results presented here. For the fiducial theoretical model, we here use the model from Muñoz et al. [2022].

We summarize the parameters used in Table A.1 and show the full HERA phase II 6th season sensitivity forecast in Figure A.3.

There are a few caveats to these estimates. Most importantly, baselines found within 7λ UV-bins together are considered redundant, while in the HERA analysis only baselines within 10 cm are considered redundant. This will artificially increase thermal sensitivity estimates. Secondly,

Parameter	Description	Value
$N_{\rm ants}$	Number of antennas within the 209 available actually observed	75, 100, 120, 148, 209
$T_{\rm sky}$	Sky temperature model	$150{\rm K}(\nu/150{\rm MHz})^{-2.5}$
$T_{\rm rcv}$	Receiver Temperature	Empirical, 600 K at 50 MHz, 60 K above 200 MHz
$\Delta\nu$	Channel width	122.07 kHz
$\tau_{\rm int}$	Coherent integration time (LST bin width)	300 sec
Δ u	UV-grid size for coherent baseline averaging	7 λ
$N_{\rm days,eff}$	Effective number of days observed coherently	$67.4^{\dagger}$
$t_{\rm day}$	Effective observed hours per day	21 hours (253 LST bins)
ϵ	Efficiency factor for frequency-dependent flags	0.9
B	Spectral window bandwidth	20 MHz
$B_{\rm eff}$	Effective spectral window bandwidth after Blackman taper	11.51 MHz
FG Wedge Level	Line-of-sight scale below which modes are filtered	0.15 h/Mpc + horizon
Theory Model	Cosmological PS from which to calculate cosmic variance	Muñoz et al. 2022

Table A.1: Parameters used within `21cmSense` to obtain sensitivity estimates for the sixth season of HERA observations. See App. A.1 for details on the algorithm. † note that $N_{\rm days,eff}$ and $t_{\rm int}$ are effectively equivalent to the actual LST footprint of the season in terms of thermal noise, under the assumption that the sky temperature is constant with LST.

the sky temperature is considered constant over the LST bins. To minimize the effect of this limitation, we use a sky model that is based at the most-observed LST (7 hours). Thirdly, cosmic variance is reduced as the square root of the number of LST bins, instead of the number of independent 'fields' observed. This artificially increases the sensitivity from cosmic variance, though this should not have a large effect since this is the sub-dominant contribution on most scales and redshifts. Finally, in this forecast we did not decimate the k-bins as was done in previous analyses. This results in some unaccounted covariance between k-bins that would tend to over-estimate the sensitivity. We do not expect this to significantly affect the qualitative conclusions derived from the forecast.

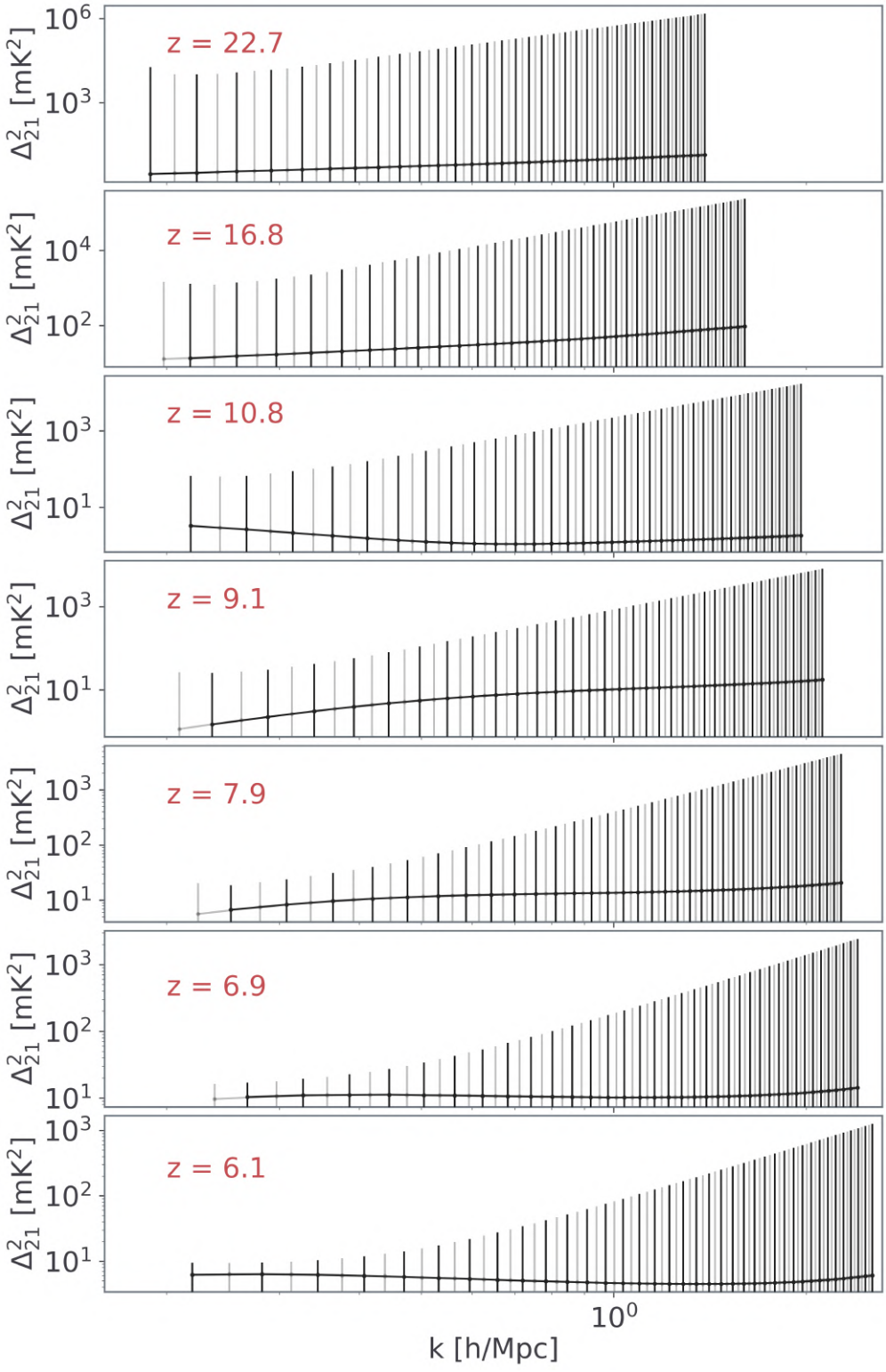

Figure A.3: HERA phase II 6th season sensitivity forecast obtained using 21cmSense with the parameters specified in Table A.1. Note that in practice, HERA decimates the k-bins to avoid requiring non-diagonal covariance (e.g., HERA Collaboration et al. 2022a). Here we have approximated this practice by using only half of the above k-bins (those highlighted in black) when computing the likelihood for our inference.

Chapter 3

21cmEMUv3: a score-based diffusion emulator of 21cmFAST summary observables including Pop II and Pop III stars

Daniela Breitman[1], Andrei Mesinger[1], Steven G. Murray[1], Ivan Nikolić[1], Roberto Trotta[234]
Draft to be submitted to Astronomy and Astrophysics shortly.

3.1 Introduction

The cosmic dawn (CD) of the first luminous objects and subsequent reionisation of the intergalactic medium (IGM) remain two of the greatest mysteries in modern cosmology. In recent years, observations of the CD and the epoch of reionisation (EoR) have increased dramatically, particularly through efforts to detect the 21-cm line of neutral hydrogen. Dedicated experiments such as the Murchison Widefield Array (MWA; Tingay et al. 2013), the Hydrogen Epoch of Reionisation Array (HERA; DeBoer et al. 2017), and the LOw Frequency ARray (LOFAR; van Haarlem et al. 2013) are focused on detecting the 21-cm power spectrum. In the coming decade, the Square Kilometre Array (SKA, e.g. Mellema et al. 2013, Koopmans et al. 2015, Mesinger 2019) is expected to deliver a tomographic map of the 21-cm signal from over half of the observable Universe.

Robust interpretation of these observations requires Bayesian inference. Such inferences require hundreds of thousands of forward model evaluations, and can thus be computationally demanding with even the most efficient semi-analytical simulators (e.g. HERA Collaboration et al. 2022a, 2023) that typically require $\gtrsim 1$ core hour per forward model. Moreover, with direct simulation, a new inference must be performed from scratch to interpret every new observation. To keep up with the fast pace of observational data releases, there has recently been an increased interest in methods that reduce the computational cost of Bayesian inferences such as emulation

[1]Scuola Normale Superiore (SNS), Piazza dei Cavalieri 7, Pisa, PI, 56125, Italy
[2]Centro Nazionale “High Performance Computer, Big Data and Quantum Computing"
[3]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136 Trieste, Italy
[4]Imperial Centre for Inference and Cosmology (ICIC), Imperial College, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, U.K.

(e.g. Kern et al. 2017, Shimabukuro and Semelin 2017, Schmit and Pritchard 2018, Jennings et al. 2019, Ghara et al. 2020, Mondal et al. 2022).

Emulation is a form of so-called *amortized* inference, where we pay an upfront cost when building the database to train the neural network (NN), typically computationally cheaper than one inference. After the emulator is trained, inferences are fast and can be preformed in minutes [2] every time a new observational data set is released [3].

In Breitman et al. [2024], we presented 21cmEMUv1, the very first emulator of *six* 21cmFAST summary observables that produces inferences in a matter of a few GPU hours [4], over 10^5 times faster than previous state-of-the-art inferences. We used 21cmEMUv1 to demonstrate how the synergy of current multi-tracer EoR/CD observations, such as UV luminosity functions [Bouwens et al., 2015, 2016, Oesch et al., 2018], Thomson scattering CMB optical depth [Planck Collaboration et al., 2020, Qin et al., 2020], and Lyman forest dark fraction [McGreer et al., 2015] allows us to learn about the first galaxies even with preliminary observations such as upper limits on the 21-cm PS. We show that current preliminary observations of 21-cm PS, in fact, *rely* on synergies with other probes in order to improve constraints, and are otherwise powerless on their own. In Cang et al. 2024, we introduced 21cmEMUv2 where we provide a different astrophysical model which includes a population of molecularly-cooling galaxies (MCGs) to model the very first luminous sources (pop III stars) in addition to the population of atomically-cooling galaxies (ACGs, aka pop II stars) already present in 21cmEMUv1. Moreover, 21cmEMUv2 features an exotic heating model, in which these MCGs source a radio background that is in excess of the CMB. We perform 21(!?) inferences with different foreground and systematics models, and, using the Bayesian evidence, find that current observations of the global 21-cm signal from EDGES [Bowman et al., 2018], in fact, strongly *disfavour* exotic models with a strong radio background.

In this work, we present 21cmEMUv3, a significant expansion to 21cmEMUv1and v2, both in terms of the physical model as well as the network architecture: (i) 21cmEMUv3 is trained on a 21cmFAST model including both populations of atomic and molecularly-cooling galaxies where we vary the escape fractions, stellar-to-halo mass relation (SHMR), and X-ray luminosity for the MCG population, for a total of ten astrophysical parameters with the new addition of the matter power spectrum normalisation parameter σ_8, for a grand total of eleven input parameters; (ii) 21cmEMUv3 is the first emulator of the cylindrical (2D) 21-cm PS, which is required in order to compare to interferometric observations in a "like-to-like" fashion (see Breitman et al. 2025a for an in-depth discussion). Since the 21-cm PS is anisotropic due to RSDs (e.g. Mao et al. 2012, Jensen et al. 2013, Pober et al. 2014, Ross et al. 2021) and lightcone evolution (e.g. see Mao et al. 2012, Datta et al. 2014, Greig and Mesinger 2018), averaging the 21-cm PS over different regions of k-space, as often happens since instruments probe much larger scales than simulations, can

[2] The time it takes to run an inference depends on the complexity of the posterior being sampled, as well as the sampler itself. Samplers such as MultiNest [Feroz et al., 2009] prioritise speed over correctness and often require much fewer likelihood evaluations than, for example, UltraNest [Buchner, 2016, 2019, 2021] that prioritises correctness over speed. In our application, the inference can be performed in a few GPU minutes with MultiNest and in a few GPU hours with UltraNest.

[3] One drawback of emulation is the need to explicitly specify a likelihood, which can be intractable for many summary statistics. Simulation based inference (SBI; e.g. Cranmer et al. 2020) gets around this problem by learning the likelihood or posterior through repeated samples from a stochastic simulator and is therefore another form of amortized inference. However, SBI methods can be limited by the accuracy of the density estimation of the likelihood or posterior (e.g. Meriot et al. 2024). Current and upcoming EoR/CD observations are of low S/N making the corresponding likelihoods close to Gaussian (e.g. Prelogović and Mesinger 2023). Furthermore, SBI relies on forward modelling the observational uncertainties and as such would need to be retrained every time there is a new observation. Therefore, emulation is currently by far the most practical form of amortized inference for EoR/CD observations, and is likely to remain so in the near future before datasets mature.

[4] See footnote 1.

lead to biases in the recovered posterior. Emulating the 2D 21-cm PS enables choosing a k-space footprint that is more similar to that of the observation, thus allowing for an unbiased inference; (iii) It is also the first score-based diffusion [Hyvärinen, 2005, Sohl-Dickstein et al., 2015, Song and Ermon, 2020, Song and Ermon, 2019, Song et al., 2020] emulator of the 21-cm PS, which learns to generate *realisations* of the 21-cm PS. Moreover, its performance is vastly superior to a convolutional NN (CNN) as in v1 and v2; (iv) In 21cmEMUv3 the rest of the summary observables are emulated with long-short term memory (LSTM) layers [Hochreiter and Schmidhuber, 1997], whose purpose is to better capture the temporal structure in timeseries data (e.g. Prelogović et al. 2022, Dorigo Jones et al. 2024). In v1 and v2, these are instead emulated with fully-connected layers.

We begin by introducing the physical model being emulated in Section 3.2. Next, we describe the machine learning architecture of the emulator in Section 3.3. In Section 3.4, we quantify the performance of 21cmEMUv3 on a test set. In Section 3.5, we apply 21cmEMUv3 in inference problems. We conclude in Section 4.7.

3.2 Simulated database

21cmEMUv1 (Breitman et al. 2024) was trained on a 21cmFASTv3 (Mesinger et al. 2011, Murray et al. 2020) model where all galaxies are characterised by a single galaxy population. Theory suggests that the first generation of stars, formed of primordial gas, cooled inside $10^6 - 10^8$ $M_\odot$ halos through the molecular hydrogen cooling channel (e.g. see Barkana and Loeb 2001). Such molecularly-cooled galaxies (MCGs) can have vastly different physical properties from their larger descendents that cool through atomic cooling (atomically-cooled galaxies; ACGs)

There have been many studies exploring the role MCGs play in cosmic history (e.g see Barkana and Loeb 2001, Holzbauer and Furlanetto 2011, Fialkov et al. 2013, Wolcott-Green et al. 2017, Schauer et al. 2021, Muñoz et al. 2022). Here we build an emulator for such a model that has two galaxy populations (ACGs and MCGs; see also Lazare et al. 2023) allowing for each one to have its own set of scaling relations (for a detailed description of the physical model, see Qin et al. 2020 and munoz), which we introduce briefly below.

3.2.1 Pop II and Pop III hosting galaxies

Galaxies form when gas from the IGM accretes onto its host dark matter halo and cools until it collapses to form stars. Galaxies can be distinguished via the dominant cooling channel which enabled their formation: (i) atomic-cooling galaxies (ACGs, aka population II stars) that obtained the majority of their gas through cooling via HI and He line transitions (efficient at virial temperatures $T_{\rm vir} \gtrsim 10^4$ K); and (ii) molecularly-cooling galaxies (MCGs, aka population III stars) that cooled through H_2 rotational and vibrational transitions efficient around virial temperatures $10^3{\rm K} \lesssim {\rm T_{vir}} \lesssim 10^4{\rm K}$ and $T_{vir} \lesssim 10^3{\rm K}$, respectively. However, H_2 can be dissociated by the build up of a Lyman-Werner (10.-13.6eV) background, and we therefore expect that eventually ACGs dominate cosmic radiation fields. Here, we allow ACGs and MCGs to have different properties, assuming that their stellar populations are dominated by PopII and PopIII stars respectively.

The stellar mass $M_{*,\rm ACG\ (MCG)}$ of an ACG (MCG) galaxy hosted by a halo of mass M_h is power-law for the faint galaxies (hosted by $M_h \lesssim 10^{12} M_\odot$ halos) that dominate the cosmic radiation fields at $z > 5$ (e.g., Kuhlen and Faucher-Giguère 2012, Das et al. 2014, Behroozi and Silk 2015, Mutch et al. 2016, Sun and Furlanetto 2016, Yue et al. 2016):

$$M_{*,\rm ACG\ (MCG)} = \min\left[1, f_{*,10(7)}\left(\frac{M_h}{M_{10(7)}}\right)^{\alpha_{*,\rm ACG\ (MCG)}}\right] \times \left(\frac{\Omega_b}{\Omega_m}\right) M_h, \tag{3.1}$$

where Ω_b is the universal baryon energy density (as a fraction of the critical energy density), and Ω_m is the total matter (i.e., cold dark matter and baryon) energy density. We may write the stellar fraction in ACGs (MCGs) as $f_{*,\text{ACG (MCG)}} \equiv f_{*,10(7)} \left(\frac{M_h}{M_{10(7)}}\right)^{\alpha_{*,\text{ACG (MCG)}}} \in [0,1]$, with $f_{*,10(7)}$ corresponding to the fraction of galactic gas in stars normalized to the amount in a halo of mass $M_{10(7)} \equiv 10^{10(7)} M_\odot$. These stellar-to-halo mass relations (SHMRs) introduce a total of four free parameters: the normalisation and power-law index of each of the two relations. We keep $f_{*,10(7)}$ and $\alpha_{*,\text{ACG}}$ as three free parameters, and fix $\alpha_{*,\text{MCG}} = 0.5$ (because...too high dimensional, can't put a prior on it anyway, will not be constrained by observations etc.).

We assume that the star formation rate (SFR) is proportional to the Hubble time, $H^{-1}(z)$, (or analogously the dynamical time, which also scales with the Hubble time during matter domination):

$$\dot{M}_{*,\text{ACG (MCG)}} = \frac{M_{*,\text{ACG (MCG)}}}{t_* H^{-1}(z)}, \tag{3.2}$$

where the characteristic star formation timescale, $t_* \in [0,1]$, is another free parameter.

Star formation is suppressed in low mass halos due to inefficient gas cooling and/or feedback (e.g., Hui and Gnedin 1997, Springel and Hernquist 2003, Okamoto et al. 2008, Sobacchi and Mesinger 2013, Xu et al. 2016, Ocvirk et al. 2020, Ma et al. 2020). We account for this suppression by including an exponential duty cycle to the mass function of halos (HMF) that host star-forming ACGs (MCGs):

$$\phi^{\text{ACG (MCG)}} = \frac{\mathrm{d}n}{\mathrm{d}M_h} \times \begin{cases} \exp\left(-\frac{M^{\text{ACG}}_{\text{crit}}}{M_h}\right) \\ \exp\left(-\frac{M^{\text{MCG}}_{\text{crit}}}{M_h}\right) \exp\left(-\frac{M_h}{M^{\text{cool}}_{\text{crit}}}\right) \end{cases} . \tag{3.3}$$

$M^{\text{ACG (MCG)}}_{\text{crit}}$ is the turnover mass below which star formation is suppressed, which we write as:

$$M^{\text{ACG (MCG)}}_{\text{crit}} = \max\left[M^{\text{cool (diss)}}_{\text{crit}}, M^{\text{ion}}_{\text{crit}}, M^{\text{SN}}_{\text{crit}}\right], \tag{3.4}$$

where we assume it is dominated by three physical processes: (i) inefficient cooling $M^{\text{cool}}_{\text{crit}}$ below the atomic cooling threshold for ACGs and Lyman-Werner feedback that photodissociates molecular hydrogen via the two-step Solomon process, $M^{\text{diss}}_{\text{crit}}$, for MCGs; (ii) photoheating feedback, $M^{\text{ion}}_{\text{crit}}$; (iii) supernova feedback, $M^{\text{SN}}_{\text{crit}}$. Moreover, for MCGs, in addition to the lower mass threshold, $M^{\text{MCG}}_{\text{crit}}$, we also include a second exponential term with an upper mass threshold, $M^{\text{cool}}_{\text{crit}}$, to transition between MCGs and ACGs around the atomic cooling threshold $T_{\text{vir}} \sim 10^4$K.

The typical ionizing escape fraction, $f_{\text{esc,ACG (MCG)}} \in [0,1]$ is described by a power-law (e.g., Paardekooper et al. 2015, Kimm et al. 2017, Lewis et al. 2020), similar to the SHMR:

$$f_{\text{esc,ACG (MCG)}} = \min\left[1, f_{\text{esc},10(7)} \left(\frac{M_h}{M_{10(7)}}\right)^{\alpha_{\text{esc}}}\right], \tag{3.5}$$

where we have a total of three free parameters for both relations: the normalizations $f_{\text{esc},10(7)}$, and the power-law index, α_{esc}.

The specific X-ray luminosity escaping the galaxies is also taken to be a power-law in energy (e.g., Das et al. 2017), $L_{X,\text{ACG (MCG)}} \propto E^{-\alpha_X}$. We normalise it via the *soft-band* (i.e. < 2 keV) X-ray luminosity per unit SFR for each galaxy population:

$$L^{\text{ACG (MCG)}}_{X<2\text{keV}/}/\dot{M}_{*,\text{ACG (MCG)}} = \int_{E_0}^{2\text{keV}} \mathrm{dE}\; L_{\text{X}}/\dot{M}_{*,\text{ACG (MCG)}}, \tag{3.6}$$

where E_0 is the minimum energy of X-ray photons capable of escaping their host galaxy, for a total of three additional free parameters. We only consider the soft-band luminosity because

Parameter	Limits	Description
$\log_{10} f_{*,10}$	$\in [-2, -0.5]$	SHMR normalisation for ACGs
$\alpha_{*,\mathrm{ACG}}$	$\in [0, 1]$	power-law index for SHMR for ACGs
t_*	$\in [0.01, 1]$	characteristic SF timescale
$\log_{10} f_{\mathrm{esc},10}$	$\in [-3, 0]$	normalisation of the ionizing escape fraction to halo mass relation for ACGs
α_{esc}	$\in [-1, 1]$	power-law index of the ionizing escape fraction to halo mass relation for ACGs and MCGs
$\log_{10} f_{*,7}$	$\in [-4, -1]$	SHMR normalisation for MCGs
$\log_{10} f_{\mathrm{esc},7}$	$\in [-3, -1]$	normalisation of the ionizing escape fraction to halo mass relation for MCGs
$\log_{10} L^{\mathrm{ACG}}_{\mathrm{X<2keV}}/\mathrm{SFR}$	$\in [38, 43]$ $\frac{\mathrm{erg\,s^{-1}}}{M_\odot\ \mathrm{yr}^{-1}}$	soft-band X-ray luminosity per unit SFR for ACGs
$\log_{10} L^{\mathrm{MCG}}_{\mathrm{X<2keV}}/\mathrm{SFR}$	$\in [39, 44]$ $\frac{\mathrm{erg\,s^{-1}}}{M_\odot\ \mathrm{yr}^{-1}}$	soft-band X-ray luminosity per unit SFR for MCGs
E_0	$\in [100, 1500]$ keV	minimum energy of X-ray photons capable of escaping their host galaxy
σ_8	$\in [0.75, 0.85]$	matter power spectrum normalisation

Table 3.1: In 21cmEMUv3, we model ACGs and MCGs by varying a total of eleven parameters: seven ACG astrophysical parameters, three MCG astrophysical parameters, and one cosmological parameter σ_8.

harder photons have a mean free path longer than the Hubble length and thus do not interact with the IGM.

In 21cmEMUv3, we model ACGs and MCGs by varying a total of eleven parameters: seven ACG astrophysical parameters, three MCG astrophysical parameters, and one cosmological parameter σ_8. In Table 3.1, we list the 21cmEMUv3 input parameters and their limiting values in the database used to train the emulator [5].

3.2.2 Summary observables

For a set of cosmological and astrophysical parameters, 21cmFAST produces 3D lightcones of IGM properties. When performing inference, these lightcones are generally compressed into summary statistics that are compared directly with observations. As such, rather than emulating full 3D lightcones, we also emulate only lower dimensional summary statistics:

(i) $\overline{\boldsymbol{x}}_{\mathrm{HI}}(\boldsymbol{z})$ — volume-averaged (global) neutral fraction of hydrogen and helium as a function of redshift (aka EoR history).

(ii) $\overline{\boldsymbol{T}}_{\mathrm{s}}(\boldsymbol{z})$ — the mean IGM spin temperature as a function of redshift.

(iii) $\overline{\boldsymbol{\delta T}}_{\mathrm{b}}(\boldsymbol{z})$ — global 21-cm brightness temperature (e.g., Madau et al. 1997, Furlanetto 2006,

[5]Note that these limits do not correspond to the prior of the database as we did not use a flat prior for most of the parameters.

Pritchard and Loeb 2012):

$$\begin{aligned}\delta T_{\rm b}(\boldsymbol{x}, z) &= \frac{T_{\rm S} - T_{\rm R}}{1+z}(1 - e^{\tau_{21}}) \\ &\approx 27\, x_{\rm HI}(1+\delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1+z}{10}\right)^{1/2}\ {\rm mK} \\ &\times \left(\frac{T_{\rm S} - T_{\rm R}}{T_{\rm S}}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right],\end{aligned} \tag{3.7}$$

where τ_{21} is the 21-cm optical depth of the intervening gas, $\delta_b \equiv \rho/\bar{\rho} - 1$ is the baryon overdensity, with ρ being the baryon density, and $T_{\rm S}$ and $T_{\rm R}$ are the spin and background temperatures, respectively. We assume throughout that the radio background is provided by the CMB, $T_{\rm R} = T_{\rm CMB}$ is the temperature of the CMB. We note that 21cmFAST computes the brightness temperature at each cell location, $\mathbf{x}$, using the exact expression in the first line of the equation above; the second line is a Taylor expansion in the limit of $\tau_{21} \ll 1$ that provides physical intuition.

(iv) $\boldsymbol{\Delta^2_{21}(k_\perp, k_\parallel, z)}$ — cylindrically-averaged 21-cm power spectrum (PS) as a function of sky-plane mode $k_\perp$ and line-of-sight (LOS) mode $k_\parallel$: $\Delta^2_{21}(k_\perp, k_\parallel, z)\ [{\rm mK}^2] \equiv k^3/(2\pi^2)\langle \tilde{T}_b \tilde{T}_b^* \rangle$, where $k = \sqrt{k_\perp^2 + k_\parallel^2}$, and $\tilde{T}_b(\boldsymbol{k}, z)$ is the Fourier dual of the brightness temperature from eq. (3.7).

(v) $\boldsymbol{\phi(M_{1500}, z)}$ — the non-ionizing UV luminosity function (UV LF), defined as the number density of galaxies per UV magnitude, M_{1500}, as a function of redshift. The ∼1500 Å rest frame luminosity is calculated from the SFR: $\dot{M}_*(M_h, z) = \mathcal{K}_{\rm UV} \times L_{\rm UV}$, where $\mathcal{K}_{\rm UV} = 1.15 \cdot 10^{-28} {\rm M}_\odot {\rm yr}^{-1}$ Hz s erg^{-1} assumes a Salpeter initial mass function (e.g., Madau and Dickinson 2014, Sun and Furlanetto 2016). The UV luminosity is related to the AB magnitude using [Oke and Gunn, 1983]: $\log\left(\frac{L_{\rm UV}}{\rm erg\ s^{-1}\ Hz^{-1}}\right) = 0.4 \times (51.63 - M_{\rm UV})$.

(vi) $\boldsymbol{\tau_e}$ — the Thompson optical depth to the last scattering surface (LSS): $\tau_e = \sigma_T \int_0^{z_{\rm LSS}} dz \left|\frac{cdt}{dz}\right| n_e$, where σ_T is the Thompson scattering cross section and n_e is the electron number density calculated assuming hydrogen and helium are singly ionized at a fraction $(1 - \overline{x}_{\rm HI})$ and that helium is doubly ionized at $z < 3$.

We create a database of about 47k samples by varying the 11 parameters listed in Table 3.1 using a Bayesian inference framework whose prior corresponds to the limits shown in the table. The inference is informed by three datasets: (i) Thomson scattering optical depth to the CMB [Planck Collaboration et al., 2020, Qin et al., 2020]: Gaussian likelihood around $\tau_e = 0.0569^{+0.0081}_{-0.0086}$; (ii) UV luminosity functions [Bouwens et al., 2015, 2016, Oesch et al., 2018]: Gaussian likelihood at redshifts $z = 6, 7, 8, 10.$; (iii) Lyα forest [Bosman et al., 2018, D'Odorico et al., 2023]: Gaussian likelihood comparison to measured CDF of Lyα forest effective optical depth from Bosman et al. 2018.

3.3 21cmEMUv3 architecture

As shown in Figure 3.1, the 21cmEMUv3 emulator is composed of two independent parts: the generative score-based diffusion model for the cylindrical 21-cm PS (rightmost branch in Figure 3.1), and the LSTM and fully-connected NN that emulates the remaining five summary observables.

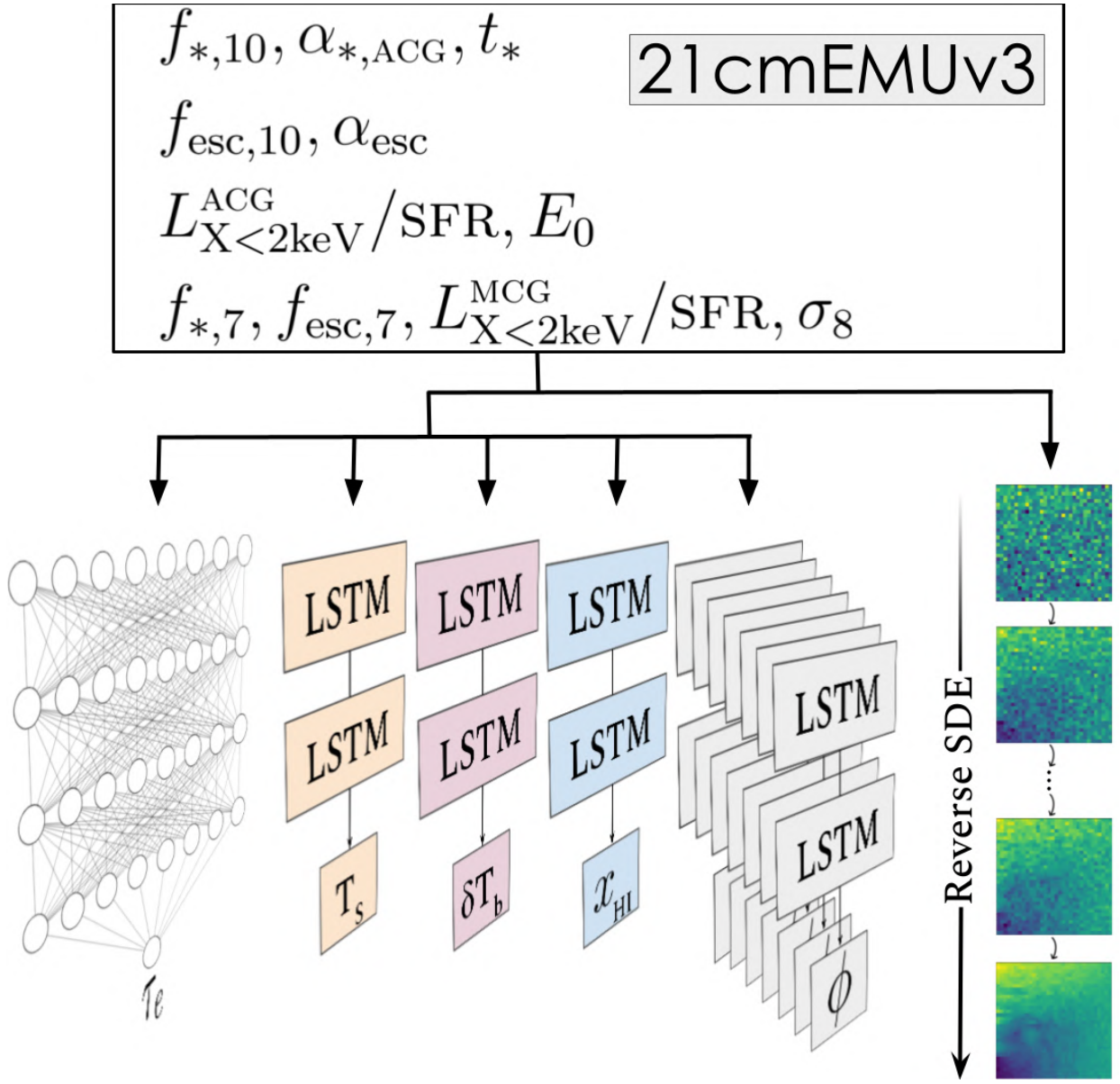

Figure 3.1: Architecture of the 21cmEMUv3 emulator. On the right: a network composed of five branches predicts the timeseries statics with LSTM layers and the Thomson scattering optical depth with a fully-connected NN. On the left: the cylindrical 2D 21-cm PS is emulated with a score-based diffusion network.

3.3.1 Generative score-based diffusion emulator of the cylindrical 21-cm PS

Emulating 2D PS is a conditional image generation task for which diffusion models exhibit state-of-the-art performance.

We adopt a state-of-the-art NN architecture for conditional image generation: a score-based diffusion model (e.g. Kawar et al. 2022). In this work, we use the same architecture as in Breitman et al. 2025a. In the rightmost branch of Figure 3.1, we illustrate the general idea behind diffusion models. Bottom to top: the forward diffusion process can be interpreted as a continuous noise-adding stochastic process where we corrupt a data sample with Gaussian noise with increasing variance until the data sample is transformed into a sample from a known Gaussian prior distribution. This stochastic process can be written as a solution to a stochastic differential equation (SDE):

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}, \tag{3.8}$$

where $\mathbf{x} = \big(x(0), \ldots, x(T)\big)$ is the diffused data (i.e. a 2D PS) at a given time in the diffusion process, $t \in [0, T]$, with $x(0)$ denoting a sample from the data distribution and $x(T)$ a sample from the Gaussian prior. $\mathbf{w}$ is a Wiener process (aka Brownian motion). The drift function $\mathbf{f}(\mathbf{x}, t)$ and the diffusion coefficient $g(t)$ are hyperparameters of our model. In this work, we choose a drift function and diffusion coefficient that correspond to the commonly-used variance preserving (VP) SDE, the continuous time-limit of the DDPM in Ho et al. 2020. Top to bottom in the rightmost branch of Figure 3.1: in order to generate new data samples, we need to reverse the forward process. This forward process can be reversed with another SDE (Anderson 1982):

$$d\mathbf{x} = \big[\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log P_t(\mathbf{x}|\tilde{\mathbf{x}})\big]dt + g(t)d\overline{\mathbf{w}}, \tag{3.9}$$

where the only unknown is the *score function* $\nabla_{\mathbf{x}} \log P_t(\mathbf{x}|\tilde{\mathbf{x}})$ of the probability density function (PDF) P of the data x (in this case, the 2D PS means) explicitly conditioned on the 2D PS realisation $\tilde{\mathbf{x}}$ in addition to the continuous time index t.

In this work, we use a U-Net autoencoder architecture similar to the one in the diffusion denoising probabilistic model (DDPM, Ho et al. 2020). U-Nets have been developed for image segmentation as they are efficient in recognizing local information in images over a range of scales [Ronneberger et al., 2015]. We use as many samples as possible to train the most accurate emulator possible, since we find that the performance does change when we vary the number of training samples. We use 40k power spectra for training, and 4k for validation. Each PS has 32 redshift bins logarithmically spaced between 5.5 and 29. Once we trained our NN to accurately predict the score for varying noise levels, we can generate new mean 2D PS samples given one 2D PS realization by solving this reverse-time SDE. Since our goal is to use the score-based diffusion model as part of an inference pipeline, we solve the reverse SDE via the probability-flow ODE method (Song et al. 2020, for more details see Appendix ??). This algorithm modestly sacrifices accuracy for a significant speed increase, yielding a mean 2D PS estimated from 200 pulls in $\sim$2 s. Our entire code is publicly available[6].

3.3.2 LSTM emulator for timeseries

Previous works (e.g. Prelogović et al. 2022, Dorigo Jones et al. 2024) have found that LSTM networks perform well on the 21-cm 1D PS and the 21-cm global signal. As such, we use LSTMs to emulate all of the timeseries summaries i.e. the global 21-cm signal, the spin temperature, the

[6] https://github.com/21cmfast/21cmEMU

neutral fraction and the UV LFs. The Thomson scattering optical depth to the CMB is emulated with a classic fully-connected network as it is an integrated summary statistic. We find that the LSTM outperforms other architectures such as the fully-connected architecture from 21cmEMUv1 as well as 1D generative score-based diffusion.

Similar to Dorigo Jones et al. 2024, all the LSTM branches have only two layers, as more layers do not enhance the performance. The database for training this NN has about 39k samples with about 4k samples each for validation and testing.

3.4 21cmEMUv3 performance

We assess the performance of 21cmEMUv3 by evaluating it on a test set and calculating the error using the following metrics:

(i) absolute difference (Abs Diff):

$$\text{Abs Diff} \equiv |y_{\text{true}} - y_{\text{pred}}| \tag{3.10}$$

(ii) fractional error (FE):

$$\text{FE}(\%) = \frac{\text{Abs Diff}}{\max(|y_{\text{true}}|, y_{\text{floor}})} \times 100, \tag{3.11}$$

where the floor is set to $y_{\text{floor}} = 10^{-2}$.

3.4.1 Cylindrical 21-cm PS

We assess the performance of 21cmEMUv3 to predict the *mean* cylindrical 21-cm PS by directly comparing it against 21-cm cylindrical PS mean estimates. We obtain these mean estimates by building a small test database of 100 parameters with about 50 realisations each. We then average over these 50 realisations to produce an estimate of the mean 21-cm PS. We find that the median (1σ) error over the entire test set and ten redshift bins is 2.40% (8.5%). We also confirm that the error on the test set means is smaller than the error when comparing the emulated PS to individual realisations from the test set.

In Figure 3.2, we show on the first row an average example of the average of 200 generated samples (right plot), and the corresponding PS realisation (left plot). On the bottom row, we show the fractional error when comparing the first row with the test set mean.

3.4.2 Other summaries

In Figure 3.3, we plot test set samples (solid line) and emulated samples (dashed line) for 10 examples for each summary. We typically cannot distinguish the two by eye. We summarise the emulator performance over all summaries in Table 3.2, where the reported statistics are evaluated over 4k test samples for all summaries except the PS where we only have 100 mean estimates. Looking at the table, we see that the LSTM performs at sub-percent median accuracy for all summaries.

3.5 HERA23 inference with MCGs

In this section, we repeat the inference from HERA Collaboration et al. [2023] but with the new 21cmEMUv3 that features a more complex astrophysical model, including both Pop II and

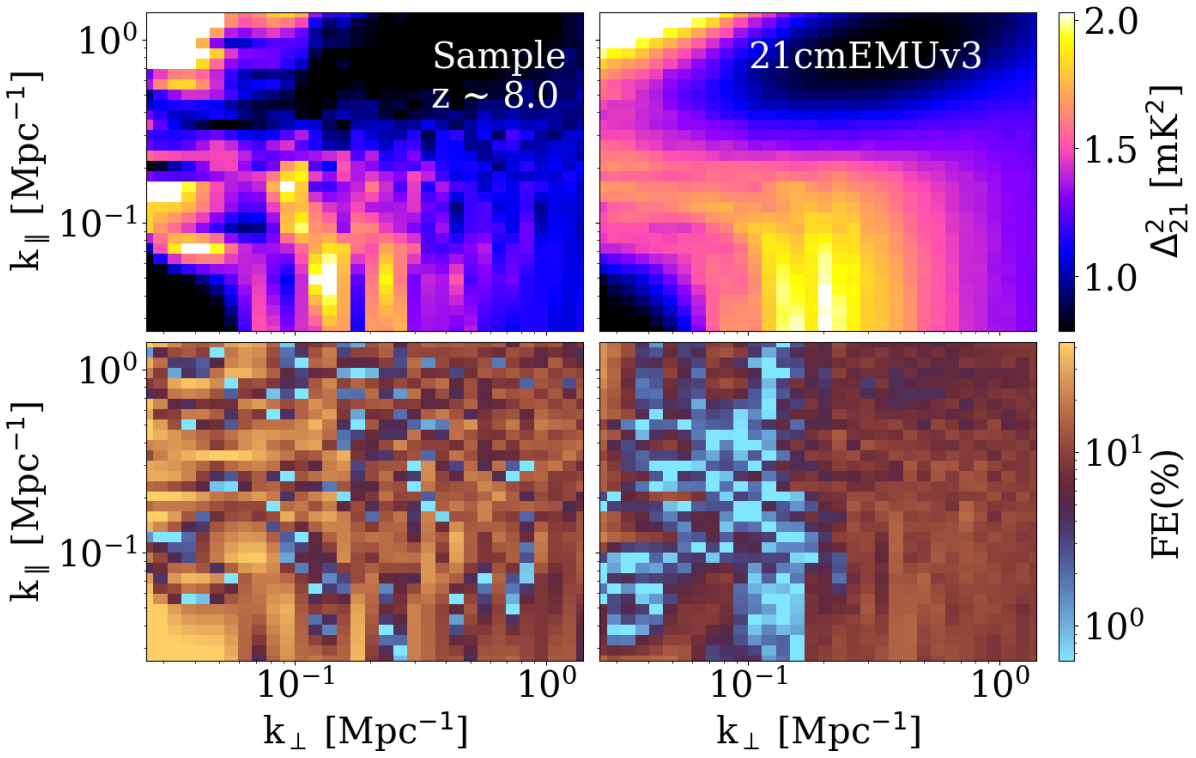

Figure 3.2: On the top row, we plot an average cylindrical 21-cm PS realisation from the test set (left) and the corresponding emulated PS from 21cmEMUv3 (right). On the bottom row, we use the fractional error to compare an estimate of the mean obtained from averaging 22 2D PS realisations with the corresponding 2D PS from the top row.

Summary	Median FE (%)	68% CL (%)
Δ^2_{21}	2.4	8.5
$\overline{T}_\mathrm{b}$	0.92	2.6
$\log \overline{T}_\mathrm{S}$	0.87	2.8
$\overline{x}_\mathrm{HI}$	0.06	0.57
τ_e	0.27	0.55
$\log \phi$	0.74	6.0

Table 3.2: Performance of the emulator over the test set for all summaries. Note that the test set for the PS consists of 100 mean PS, while for all other summaries, the test set contains 4k samples.

Parameter	Lazare+24	Lazare+24-like	Informed
$\log_{10} f_{*,7}$	$\in [-3.5, -1]$	$\in [-3.5, -1]$	$\mathcal{N}(-3, 0.6)$
$\alpha_{*,\mathrm{MCG}}$	$\in [-0.5, 0.5]$	$\alpha_{*,\mathrm{ACG}}$	1
$\log_{10} f_{*,10}$	$\in [-3, 0]$	$\in [-2, 0.5]$	$\mathcal{N}(-1.5, 0.6)$
$\alpha_{*,\mathrm{ACG}}$	$\in [-0.5, 1]$	$\in [0, 1]$	1
$\log_{10} f_{\mathrm{esc},7}$	$\in [-3, 0]$	$\in [-3, -1]$	$\in [-3, -1]$
α_X	$\in [-1, 3]$	1	1

Table 3.3: We choose more informed MCG priors on the SHMR and consequently adjust the ACG priors to ensure continuity around 10^8 $\mathrm{M}_\odot$.

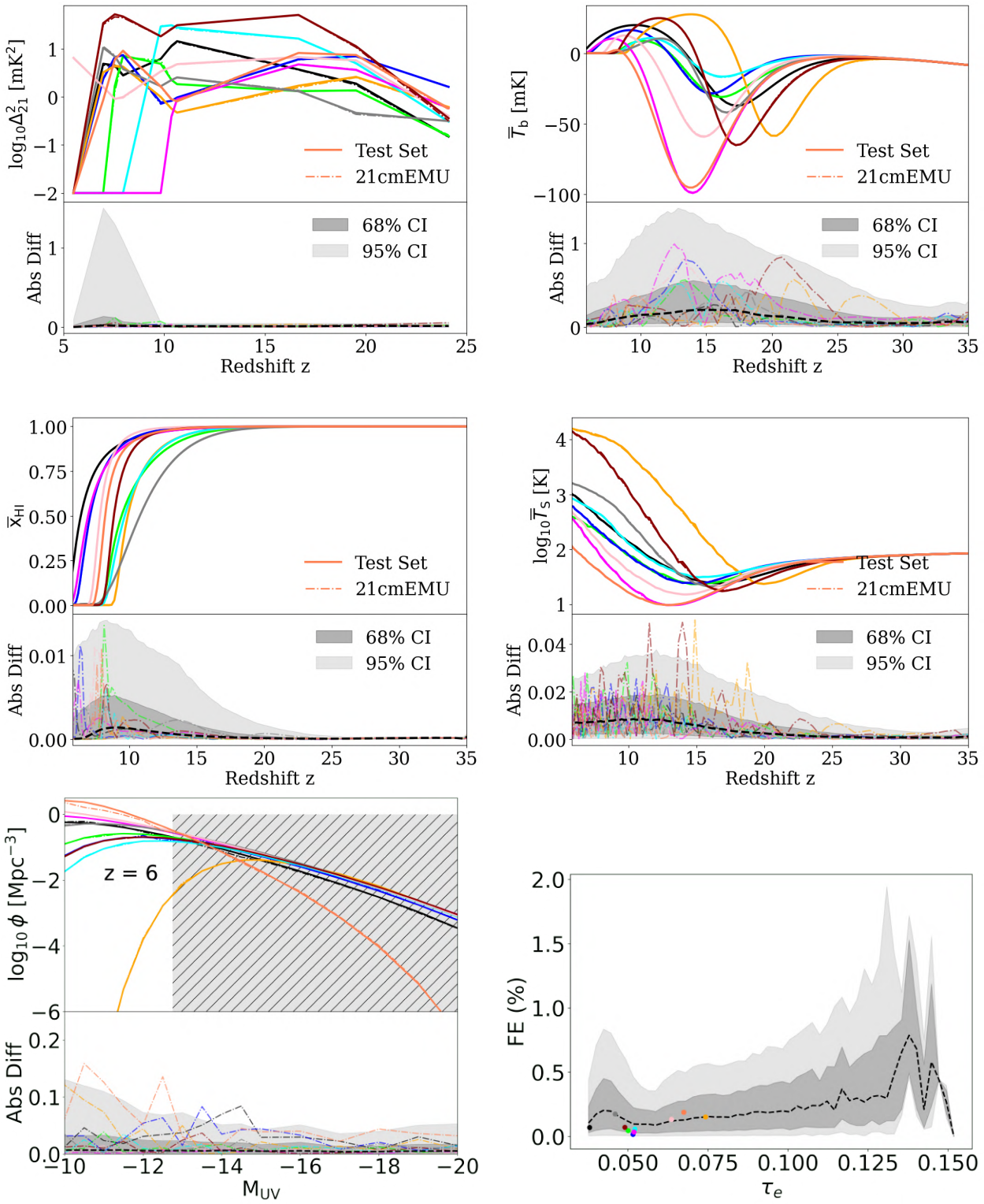

Figure 3.3: A subset of summary outputs from 21cmEMUv3 for ten random samples from the test set. Panels show: redshift evolution of the $k = 0.1$ Mpc^{-1} 21-cm PS amplitude, redshift evolution of the mean 21cm brightness temperature, redshift evolution of the mean spin temperature in the neutral IGM, the CMB optical depth, UV LFs at $z = 6$, the EoR history (*clockwise from upper left*). Colors denote the astrophysical parameter sample with solid (dashed) lines corresponding to outputs from 21cmFAST (21cmEMUv3). In the bottom sub-panels, we show the absolute differences between the predicted and true quantities shown in the top sub-panels. Absolute differences of the ten random samples are shown with the corresponding colors, while the median absolute differences (FE in the case of τ_e) computed over the entire test set are shown with dashed, black curves. Dark (light) shaded regions enclose 68% (95%) CL.

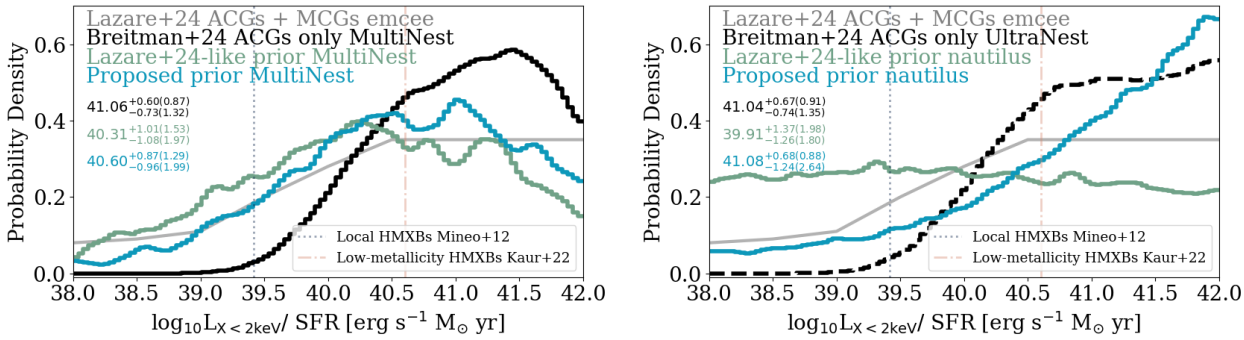

Figure 3.4: Left plot: The posterior on L_X from HERA23 as obtained with `21cmEMUv1` in Breitman et al. 2024 is shown in black. In green, we show the posterior obtained with `21cmEMUv3` with a prior similar to Lazare et al. 2023. In blue, we show the posterior obtained with `21cmEMUv3`with a prior informed by the Renaissance hydrodynamic simulations. All these posteriors are obtained with the `MultiNest` sampler. Right plot: same as the left plot, but with the `nautilus` sampler for the blue and green posteriors and `UltraNest` sampler for the black. These are expected to be more correct (e.g. see Lange 2023, Albert et al. 2025).

Pop III stars. We compare it to Lazare et al. 2023 who were the first to perform such an inference with the inclusion of Pop III stars. We also compare the resulting posterior with two different samplers: `MultiNest` [Feroz et al., 2009] and `nautilus` [Lange, 2023]. We also propose a new prior informed by hydrodynamical simulations [Xu et al., 2016], improving on the flat prior used in Lazare et al. 2023, and discuss its implications.

In HERA Collaboration et al. [2023], we have a likelihood with four CD/EoR observables: (i) UV luminosity functions [Bouwens et al., 2015, 2016, Oesch et al., 2018] with a Gaussian likelihood; (ii) Thomson scattering optical depth to the CMB [Planck Collaboration et al., 2020, Qin et al., 2020] which uses a Gaussian likelihood around $\tau_e = 0.0569^{+0.0081}_{-0.0086}$; (iii) upper limit $\overline{x}_{\rm HI} < 0.06 \pm 0.05$ on the EoR history at $z = 5.9$ obtained with the Lyα dark fraction method [McGreer et al., 2015]. The likelihood function is unity at $\overline{x}_{\rm HI}(z = 5.9) < 0.06$, decreasing as a one-sided Gaussian for higher neutral fraction values; (iv) HERA upper limits on 21-cm PS with a positive and infinite prior on systematics [HERA Collaboration et al., 2023].

We begin by reproducing the result from Lazare et al. 2023 who were the first to perform an inference on HERA23 upper limits with a model including both PopII and PopIII stars. The two main differences between our analysis and theirs are: (i) We do not have the exact same database and training data priors. As such, we reproduce their result with the closest possible prior to theirs as shown in the first two columns of Table 3.3; and (ii) we do not use the `emcee` [Foreman-Mackey et al., 2013] sampler.

In the left plot of Figure 3.4, we show in grey a rough outline of the posterior on the X-ray luminosity per SFR parameter obtained by Lazare et al. 2023, which we compare to our result obtained with `MultiNest` in green. In the right plot of Figure 3.4, we instead plot in green the posterior obtained from the `nautilus` sampler. The posterior obtained with `MultiNest` roughly reproduces the result from Lazare et al. 2023, except at the edge of the posterior, at the highest luminosities, where `MultiNest` is known to be inaccurate (e.g. see Breitman et al. 2024). We see, however, that the posterior obtained with `nautilus`, which is expected to be more correct than the one obtained from `MultiNest` (e.g. see Lange 2023, Albert et al. 2025), is completely flat. This suggests that when performing an inference in such a high-dimensional parameter space with a weakly-constraining likelihood, the posterior is heavily dominated by the prior and sampler choice.

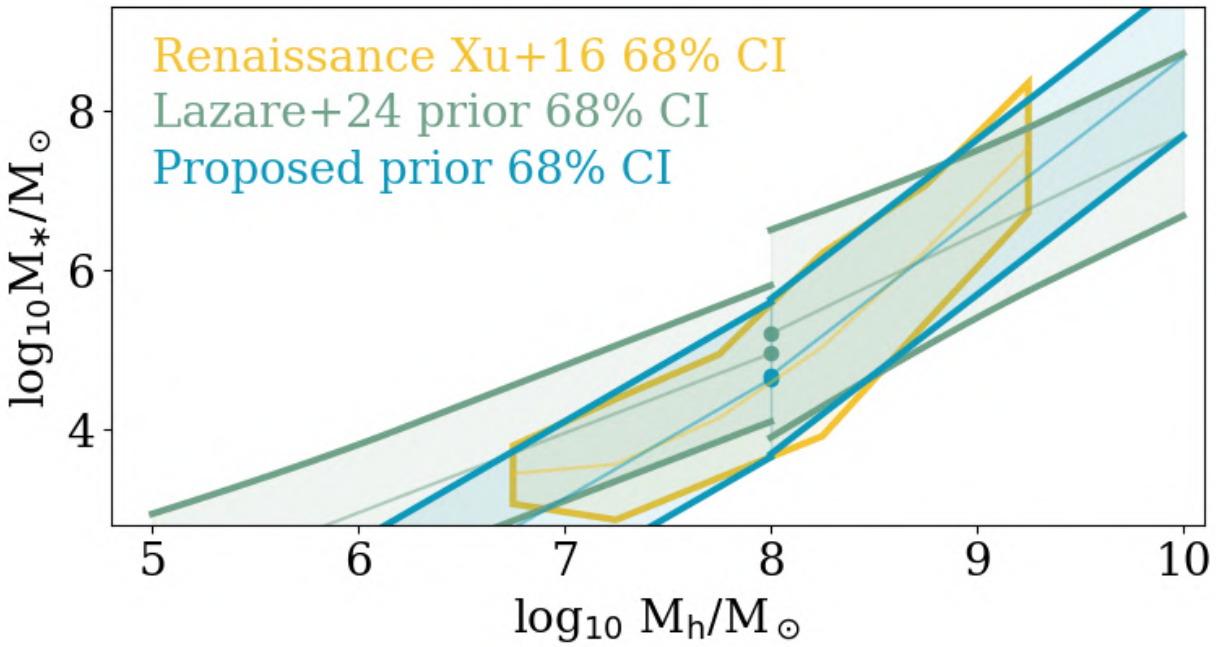

Figure 3.5: In yellow, we show the SHMR for the Renaissance simulations [Xu et al., 2016] as a reference. Previous works such as Lazare et al. 2023 had a broad prior favouring higher stellar masses. We choose a prior informed by the Renaissance simulations that favours lower stellar masses in comparison to Lazare et al. 2023 as shown in blue.

Since the result is heavily prior-dependent, we are interested in seeing how choosing a more informative prior would affect the result. We expect that a prior that favours higher stellar masses would influence the posterior to favour lower X-ray luminosities per unit SFR. We therefore propose a new prior informed by the Renaissance hydrodynamical simulations [Xu et al., 2016]. In Figure 3.5, we show in yellow the SHMR from Xu et al. 2016 and in green the flat prior from Lazare et al. 2023. We can see that the Lazare et al. 2023 favours much higher stellar masses than those observed in Xu et al. 2016. As such, we propose a prior based on Xu et al. 2016 where lower stellar masses are favoured as shown in blue in Figure 3.5.

In Figure 3.4, we show the 1D marginal distribution of the X-ray luminosity per unit SFR for the proposed prior in blue. Indeed, we find that the inference with the prior informed by the Renaissance simulations which favours lower stellar messes yields an L_X posterior that consequently favours higher X-ray luminosities per unit SFR. The result obtained with the new prior supports the original claims made in HERA23.

3.6 Conclusion

In this work, we present a score-based diffusion emulator of the cylindrical 21-cm PS and an LSTM emulator for five other summary statistics. We find that all summaries perform at an accuracy of order percent or less. We showcase an application of the emulator to reproduce results from Lazare et al. 2023. We then highlight that this result is not only prior-dependent, but also sampler dependent. We show this by varying both the sampler and the prior. We find that choosing a prior informed by the Renaissance hydrodynamical simulations which favour lower stellar masses produces a posterior on the X-ray luminosity per unit SFR that favours higher values, in support with the original result published by HERA Collaboration et al. 2022a. We also

caution that in cases where the likelihood is weakly constraining such as in this case, both the prior and sampler choice can significantly affect the resulting conclusions.

Acknowledgements

We gratefully acknowledge computational resources of the Center for High Performance Computing (CHPC) at SNS. A.M acknowledges support from the Italian Ministry of Universities and Research (MUR) through the PRO3 project "Data Science methods for Multi-Messenger Astrophysics and Cosmology", as well as partial support from the Fondazione ICSC, Spoke 3 “Astrophysics and Cosmos Observations”, Piano Nazionale di Ripresa e Resilienza Project ID CN00000013 “Italian Research Center on High-Performance Computing, Big Data and Quantum Computing” funded by MUR Missione 4 Componente 2 Investimento 1.4: Potenziamento strutture di ricerca e creazione di “campioni nazionali di R&S (M4C2-19)” - Next Generation EU (NGEU). S.M. has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101067043.

Data Availability

The emulator is on a publicly accessible github repository, as well as available for install as a Python package using `pip`.

Chapter 4

Sample Variance Denoising in Cylindrical 21-cm Power Spectra

Daniela Breitman[1], Andrei Mesinger[1], Steven G. Murray[1], Anshuman Acharya[2]

[1]Scuola Normale Superiore (SNS), Piazza dei Cavalieri 7, Pisa, PI, 56125, Italy
[2]Max-Planck-Institut für Astrophysik, Garching 85748, Germany

Abstract

State-of-the-art simulations of reionisation-era 21-cm signal have limited volumes, generally orders of magnitude smaller than observations. Consequently, the Fourier modes in common between simulation and observation have limited overlap, especially in cylindrical (2D) k-space that is natural for 21-cm interferometry. This makes sample variance (i.e. the deviation of the simulated sample from the population mean due to finite box size) a potential issue when interpreting upcoming 21-cm observations. Here, we introduce `21cmPSDenoiser`, a score-based diffusion model that can be applied to a single, forward-modelled realisation of the 21-cm 2D power spectrum (PS), predicting the corresponding population mean on-the-fly during Bayesian inference. Individual samples of 2D Fourier amplitudes of wave modes relevant to current 21-cm observations can deviate from the mean by over 50% for 300 cMpc simulations, even when only considering stochasticity due to sampling of Gaussian initial conditions. `21cmPSDenoiser` reduces this deviation by an order of magnitude, outperforming current state-of-the-art sample variance mitigation techniques like Fixing & Pairing by a factor of few at almost no additional computational cost ($\sim$ 2s per PS). Unlike emulators, `21cmPSDenoiser` is not tied to a particular model or simulator since its input is a (model-agnostic) realisation of the 2D 21-cm PS. Indeed, we confirm that `21cmPSDenoiser` generalises to power spectra produced with a different 21-cm simulator than those on which it was trained. To quantify the improvement in parameter recovery, we simulate a 21-cm PS detection by the Hydrogen Epoch of Reionization Arrays (HERA) and run different inference pipelines corresponding to commonly-used approximations. We find that using `21cmPSDenoiser` in the inference pipeline outperforms other approaches, yielding an unbiased posterior that is 50% narrower in most inferred parameters.

4.1 Introduction

The cosmic dawn (CD) of the first luminous objects and eventual reionisation of the intergalactic medium (IGM) remain among the greatest mysteries in modern cosmology. One of the most informative probes of the CD and epoch of reionisation (EoR) is the 21-cm line of the hyperfine structure of the neutral hydrogen atom. The 21-cm line has unmatched potential, ultimately able to provide us with a 3D map of more than half of our observable Universe, as expected with the upcoming Square Kilometre Array (SKA[2], e.g. Mellema et al. 2013, Koopmans et al. 2015, Mesinger 2019).

Precursors to the SKA telescope, such as the the Murchison Widefield Array (MWA[3], Tingay et al. 2013), the Hydrogen Epoch of Reionisation Array (HERA[4], e.g. DeBoer et al. 2017), LOw Frequency ARray (LOFAR[5], e.g. van Haarlem et al. 2013), and New Extension in Nançay Upgrading loFAR (NENUFAR[6], e.g. Zarka et al. 2012) are instead focused on a first detection of the 21-cm power spectrum (PS), since as a well-motivated summary statistic of interferometric observations, it has an enhanced signal-to-noise ratio (S/N) compared to 3D maps. Robustly interpreting such measurements is only possible with Bayesian inference. The current approach to Bayesian inference of 21-cm power spectra, however, relies on several approximations, whose validity is poorly understood[7].

Typically, one samples astrophysical and cosmological parameters $\tilde{\theta}$ from priors $p(\theta)$, forward-models the 3D non-Gaussian 21-cm signal with a simulator, compresses the simulated signal into a summary statistic (e.g. the spherically-averaged 1D PS), and compares the forward model to the observed summary, which in the case of the 1D PS is a function of redshift z and Fourier scale k: $\Delta^2_{21,\,\mathrm{obs}}(k,z)$. Comparison of the forward model to observations is quantified by a likelihood that is approximated to be a Gaussian at each scale and redshift bin:

$$\ln \mathcal{L}\left(\Delta^2_{21,\,\mathrm{obs}}\middle|\tilde{\theta}\right) \propto -[\Delta^2_{21,\,\mathrm{obs}} - \mu(\tilde{\theta})]^T \Sigma^{-1}(\tilde{\theta})[\Delta^2_{21,\,\mathrm{obs}} - \mu(\tilde{\theta})]\ , \tag{4.1}$$

where $\mu(k,z|\tilde{\theta})$ and $\Sigma(k,z|\tilde{\theta})$ are the expectation values and the covariance matrix of the 21-cm PS, averaged not just over modes in given (k,z) bins of a single simulation, but also averaged over many different realisations, i, of the initial conditions (and any other important source of scatter): e.g. $\mu(k,z|\tilde{\theta}) = \langle\Delta^2_{21,i}(k,z|\tilde{\theta})\rangle_i$. However, for computational reasons, sample variance of the initial conditions is ignored i.e. $\mu(\tilde{\theta})$ is generally computed from a single realisation, $\mu(k,z|\tilde{\theta}) \approx \Delta^2_{21,i}(k,z|\tilde{\theta})$, while the covariance is assumed to be diagonal at some fiducial parameter set, $\Sigma(\tilde{\theta}) \approx \sigma^2(\theta_{\mathrm{fid}})$ (for an in-depth analysis of the accuracy of these approximations, see Prelogović and Mesinger 2023). The above-mentioned steps are then repeated many times in order to map out the parameter posterior via Bayes' theorem.

One problem with this approach is that the volume of the forward model does not correspond to that probed by observations. Due to computational restrictions, simulations which resolve relevant scales have much smaller volumes than observed 21-cm fields. For example, the HERA interferometer observes a volume of over 4 cGpc^3, over 100 times larger than typical forward model volumes. This essentially means that we cannot compare like to like when interpreting observations with theory, resulting in two limitations:

[2] https://www.skao.int/en
[3] https://www.mwatelescope.org/
[4] https://reionization.org/
[5] http://www.lofar.org/
[6] https://nenufar.obs-nancay.fr/en/homepage-en/
[7] It is worth noting that these approximations persist in a frequentist framework and have the same detrimental effects as in a Bayesian context, but are more difficult to interpret.

(i) sample variance — forward models may not have enough independent modes to obtain an accurate estimate of the mean 21-cm PS, $\mu(\tilde{\theta})$, on large scales (i.e. small k).

(ii) different PS footprint — forward models and observations could probe very different wave modes in cylindrical (2D) Fourier space (e.g. Pober 2015).

Issue (i) can be problematic since initial 21-cm PS detections will likely be limited to a handful of large-scale wave modes; therefore, an "unlucky" realisation of the forward model could result in large biases (e.g. Zhao et al. 2022b; see the detailed analysis in Prelogović and Mesinger 2023). Naively, sample variance can be mitigated by taking an ensemble average over many forward models while varying the initial conditions (ICs; e.g. Giri et al. 2023, Acharya et al. 2024), or performing an initial exploration to pick a set of ICs which result in forward model that are close to the mean at some fiducial parameter set $\theta_{\rm fid}$ (e.g. Prelogović and Mesinger 2023). Both of these approaches, however, could require hundreds of additional simulations. The number of simulations required to accurately estimate the power spectrum mean can be reduced by choosing correlated ICs (see Rácz et al. 2023 for an overview). For example, fixing and pairing (F&P; e.g., Angulo and Pontzen 2016, Pontzen et al. 2016, Giri et al. 2023, Acharya et al. 2024) requires only one additional evaluation of the forward model. Nevertheless, this is non-negligible computational overhead, considering that typical inferences require over 200k evaluations of the likelihood (e.g., HERA Collaboration et al. 2022a). Moreover, the residual uncertainty on the 1D PS mean can still be of order tens of percent even when using paired 21-cm simulations (e.g., Giri et al. 2023, Acharya et al. 2024).

Issue (ii) can be problematic since observation and theory cannot be compared in the same region of 2D Fourier space (i.e. $k_\perp, k_\parallel$; see Figure 4.2). This might not seem like an issue, since the 21-cm likelihood has always been evaluated using the spherically-averaged (1D) PS, comparing theory to the observation at the same magnitude $k = \sqrt{k_\perp^2 + k_\parallel^2}$. However, because the $(k_\perp, k_\parallel)$ modes contributing to a given k-bin are generally very different for the model and the observation (see Figure 4.2), such a comparison would only be unbiased if the signal were isotropic. Indeed, the cosmic 21-cm PS is not isotropic due to redshift-space distortions (RSDs) (e.g. Bharadwaj and Ali 2004, Barkana and Loeb 2006, Mao et al. 2012, Jensen et al. 2013, Pober 2015, Ross et al. 2021), as well as the redshift evolution of the signal along the line of sight direction (e.g. Mao et al. 2012, Datta et al. 2014, Greig and Mesinger 2018). RSDs can boost the 21-cm PS in k-modes relevant to observations by up to a factor of $\sim$5 in comparison to the spherically-averaged 21-cm PS at moderate to high neutral fractions (e.g., see Figure 7 in Jensen et al. 2013). Moreover, ignoring redshift evolution in the 21-cm lightcone could bias inferred constraints by $\sim$ few $-\,10\sigma$ (e.g. Greig and Mesinger 2018). Therefore averaging the observation and the forward model over different $(k_\perp, k_\parallel)$ bins in order to compare them at the same k magnitude might result in sizeable biases.

In this work, we introduce `21cmPSDenoiser`[8]: a model-independent machine-learning-based tool for sample variance mitigation that provides an estimate of the IC-averaged *mean* 2D 21-cm power spectrum given a *single* realisation. We propose an improved inference pipeline where we mitigate sample variance (issue (i)) by applying `21cmPSDenoiser` on the simulated 2D PS on-the-fly. Unlike emulators, `21cmPSDenoiser` is not tied to a particular model or simulator since its input is a (model-agnostic) realisation of the 2D 21-cm PS. To mitigate issue (ii), our pipeline applies a cut in cylindrical k-space, averaging only over the modes closest to the ones available in the observation after removing the region dominated by foregrounds and systematics (e.g. Pober 2015). Cutting out the foreground- and systematic-dominated "wedge" from the forward-

[8] https://github.com/DanielaBreitman/21cmPSDenoiser

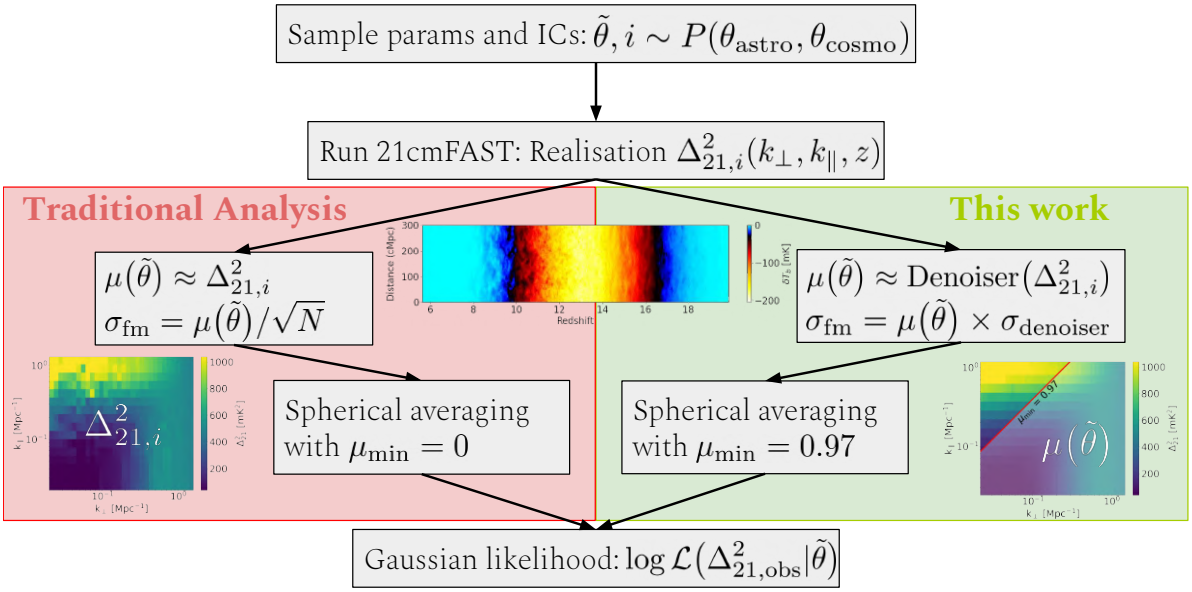

Figure 4.1: Flowchart comparing the current state-of-the-art inference pipeline (left side) with this work (right side). In current state-of-the-art pipelines, we use a single realisation of the initial conditions to estimate the mean 1D 21-cm PS. Moreover, the simulated 1D PS is computed by spherically averaging over different wave modes than those used to compute the observed 1D PS. In this work, we account for sample variance by applying `21cmPSDenoiser` a score-based diffusion model trained to estimate the mean 21-cm PS from a single realisation. We also account for 21-cm PS anisotropy by averaging the 2D PS only over modes above $\mu_{\text{min}} = 0.97$, the region of 2D PS that is closest to where current 21-cm PS instruments observe. Applying a cut in μ_{min} significantly exacerbates the problem of sample variance, and would not be practical without the use of `21cmPSDenoiser`.

modelled 2D PS significantly exacerbates sample variance since it reduces the number of Fourier modes available when spherically averaging. Indeed, although simulation box sizes larger than $\gtrsim$300 cMpc were found to have negligible sample variance in the spherically-averaged PS (e.g., Iliev et al. 2006, Kaur et al. 2020), here we show that is no longer the case after first applying a "wedge" cut in 2D k-space.

This paper is organised as follows. We begin by reviewing the traditional inference pipeline in Section 4.2. Then in Section 4.3, we introduce `21cmPSDenoiser`, demonstrating and testing its performance. In Section 4.4, we compare `21cmPSDenoiser` to Fixing & Pairing, a state-of-the-art sample variance mitigation method. In Section 4.5, we test `21cmPSDenoiser` on "out-of-distribution" (OOD) data, using power spectra from a different simulator than that used for training. In Section 4.6, we apply `21cmPSDenoiser` in a realistic inference and compare it to the traditional pipeline. We conclude with Section 4.7, where we summarise the main achievements of this paper. Throughout this work, we assume a flat ΛCDM cosmology with $(\Omega_\Lambda, \Omega_m, \Omega_b, h, \sigma_8, n_s) = (0.69, 0.31, 0.049, 0.68, 0.82, 0.97)$ consistent with Planck Collaboration et al. 2020. All distances are in comoving units unless explicitly stated otherwise.

4.2 Explicit likelihood inference from 21-cm power spectra

In this section, we describe current, state-of-the-art Bayesian inference pipelines (see left side of Figure 4.1), analogous to those applied to first generation 21-cm interferometers (e.g., HERA

Collaboration et al. 2023, Munshi et al. 2024, Mertens et al. 2025, Nunhokee et al. 2025). We then introduce the changes proposed in this work (right side of Figure 4.1).

4.2.1 Forward modelling the 21-cm signal

A single forward model of the cosmic 21-cm signal consists of the following steps:

1. Sample astrophysical and cosmological parameters. For a given set of cosmological parameters, generate initial conditions (e.g. a Gaussian random field sampled from a power spectrum) creating a realisation of the linear matter density and velocity fields.

2. Evolve densities and velocities to lower redshifts (e.g. via second-order Lagrangian perturbation theory; 2LPT Scoccimarro 1998).

3. Assign galaxy properties to the dark matter halos according to the sampled astrophysical parameters (e.g. via semi-empirical relations; Park et al. 2019).

4. Compute corresponding inhomogeneous radiation fields and their role in heating and ionising the IGM.

5. Calculate the corresponding 21-cm brightness temperature at each cell and redshift, $(\mathbf{x}, z)$ (e.g, Madau et al. 1997, Furlanetto 2006, Pritchard and Loeb 2012):

$$\begin{aligned} T_{\rm b}(\mathbf{x}, z) &= \frac{T_{\rm S} - T_{\rm R}}{1+z}(1 - e^{\tau_{21}}) \\ &\approx 27\, x_{\rm HI}(1+\delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1+z}{10}\right)^{1/2}\ {\rm mK} \\ &\times \left(\frac{T_{\rm S} - T_{\rm R}}{T_{\rm S}}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right], \end{aligned} \tag{4.2}$$

   where τ_{21} is the 21-cm optical depth of the intervening gas, and $T_{\rm S}$ and $T_{\rm R}$ are the spin and background temperatures, respectively[9]. $x_{\rm HI}$ is the fraction of neutral hydrogen, $\delta_b \equiv \rho/\bar{\rho} - 1$ is the baryon overdensity, $\partial_r v_r$ is the baryon peculiar velocity gradient along the line of sight, $H(z)$ is the Hubble parameter at redshift z, and Ω_m and Ω_b are the mass densities of cold dark matter (CDM) and baryons, respectively.

6. Compute a summary statistic from the 21-cm brightness temperature light cone, in order to compare it to the same summary of the observation. Currently all[10] analyses use the spherically-averaged 1D 21-cm PS as a summary:

$$\left\langle \tilde{T}_b(k,z)\tilde{T}_b^*(k',z)\right\rangle \equiv (2\pi)^3\delta_D(k-k')\frac{2\pi^2}{k^3}\Delta^2_{21}(k,z) \tag{4.3}$$

[9]Motivated by observations of local, radio-loud galaxies (e.g. Furlanetto et al. 2006) as well as the global 21-cm signal (e.g. Cang et al. 2024, Singh et al. 2022), we assume that the radio background is determined by the cosmic microwave background (CMB), therefore $T_{\rm R} = T_{\rm CMB}$.

[10]While there have been studies using the 21-cm 2D PS (e.g. Greig et al. 2024) and many other summaries (for example the bispectrum e.g., Mondal et al. 2021, Watkinson et al. 2022, Tiwari et al. 2022, wavelet-based methods e.g., Greig et al. 2022, Zhao et al. 2024), and "optimal" learned summaries e.g. Prelogović and Mesinger 2024, Schosser et al. 2025), none of them have yet been applied to real observational data. Using higher-dimensional summaries would decrease the S/N available for instruments seeking a preliminary detection, as well as making it more important to account for covariances/non-Gaussianity in the likelihood.

where $\tilde{T}_b^*(k', z)$ is the conjugate of the Fourier transform of the 21-cm brightness temperature at redshift z and wave mode $k'[\mathrm{Mpc}^{-1}] = \sqrt{\mathrm{k_x^2 + k_y^2 + k_z^2}} = \sqrt{\mathrm{k_\perp^2 + k_\parallel^2}}$ for a spherical average and cylindrical average at sky-plane mode $k_\perp$ and line-of-sight mode $k_\parallel$, respectively. δ_D is the Dirac delta function, and $\Delta^2_{21}(k, z)$ in mK^2 is the dimensionless 21-cm PS per logarithmic wave mode interval.

Once the forward model is compressed into a summary statistic such as the 21-cm power spectrum, it is compared to the observation by evaluating the likelihood.

4.2.2 Evaluating the likelihood on the 21-cm PS

For complex physical models such as the 21-cm signal, the likelihood function is analytically intractable. As such, all current inferences assume a Gaussian functional form[11] as defined in Eq. 4.1. Assuming a diagonal variance and approximating the mean 21-cm PS with a single realisation $\Delta^2_{21,\,i}$ can produce a large bias in the resulting posterior for high S/N data (e.g., over 5σ as seen from Figure 5 in Prelogović and Mesinger 2023).

Another common approximation of the traditional analysis is that the 1D PS is computed by averaging the amplitudes of all wave modes in a given k-bin. However, as shown in Figure 4.2, the footprint of 21-cm interferometers in cylindrical (2D) k-space can be very different from the corresponding footprint of the forward model.

The cylindrical k-space available to interferometers is limited by a combination of foregrounds, instrument layout, and dish size. The distribution of baselines limits the accessible angular scales. Foregrounds dominate in the regime of low $k_\parallel$, since they are spectrally smooth. The chromatic instrument response, however, causes the foregrounds to leak out into a 'wedge'-like region in 2D k-space (e.g. [Parsons et al., 2012, Liu et al., 2014a,b]). As a result, "clean" EoR measurements are performed outside (or near the boundary) of the wedge (c.f. blue shaded region in Figure 4.2). The cylindrical k-space available to forward models is limited by the need to resolve relevant sources and sinks and the corresponding physical processes. This means that cell sizes of physics-rich simulations cannot be much larger than $\sim$ 1 Mpc. On the other hand, needing to compute thousands of forward-models in a reasonable time limits the box sizes to $\sim$ few $\times$ 100 Mpc. Therefore, forward models typically span wave modes of $1 \lesssim \mathrm{k/Mpc^{-1}} \lesssim$ few $\times$ 0.01.

This discrepancy between the k-space footprint of observations and simulations is problematic because the 21-cm PS is anisotropic due to RSDs and the light cone evolution along the line-of-sight axis. This issue can be avoided by cropping the forward-modelled 2D PS to the k-space region closest to that of the observation (e.g. the region above the red line in Figure 4.2). However, cropping out modes exacerbates the sample variance problem at large scales as it significantly reduces the number of modes available to perform the averaging.

In this work, we introduce `21cmPSDenoiser` (right branch of Figure 4.1) to mitigate sample variance: after a forward model is computed, the resulting 2D PS realisation is passed through our neural network that produces an accurate estimate of the mean 21-cm PS. Mitigating sample variance with `21cmPSDenoiser` allows us to also take into account the 21-cm PS anisotropy: we introduce a cut in cylindrical PS space where we exclude all Fourier modes below $\mu_{\min} = 0.97$, marked by a red line in Figure 4.2, where $\mu_{\min} = \cos\theta$, and $\tan\theta = \frac{k_\perp}{k_\parallel}$. Introducing this cut

[11] Simulation-based inference (SBI) can be used to avoid having to define an explicit functional form for the likelihood. However, a Gaussian likelihood for the 1D PS is a decent approximation for preliminary, low S/N data obtained by averaging over many modes and upper limits (e.g., Prelogović and Mesinger 2023, Meriot et al. 2024) such as the data currently available. SBI will be more relevant for high S/N power spectra, as well as for more complicated summaries.

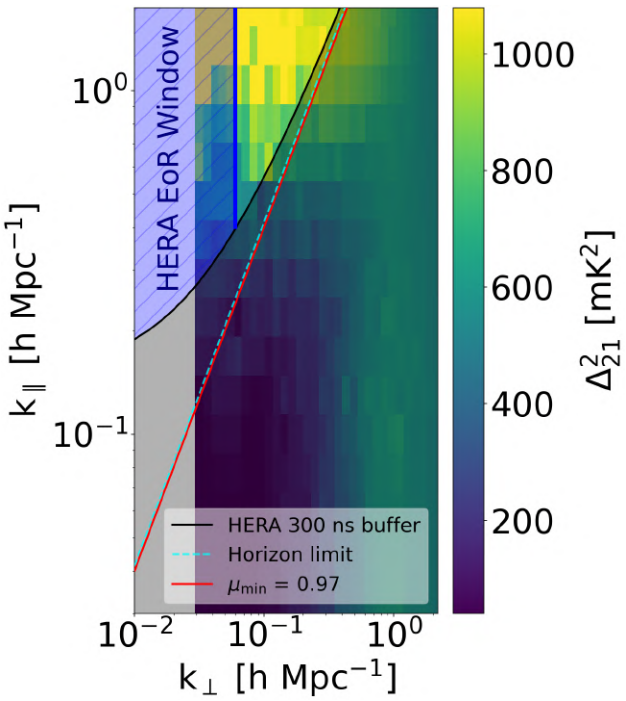

Figure 4.2: Cylindrically-averaged (2D) 21-cm power spectrum as a function of line-of-sight modes $k_\parallel$ and sky-plane modes $k_\perp$. The color map shows the PS amplitude calculated from a slice through a single simulated light cone centred at $z = 9$. The simulation box has a side length of 300 cMpc and was generated with 21cmFASTv3. The blue hashed area is the HERA EoR window. The dashed cyan line is the horizon limit and the black solid line is the horizon limit with a 300 ns buffer added to it to account for additional foreground leakage (see HERA Collaboration et al. 2023). The red solid line is drawn at a value of $\mu_{\min} = 0.97$ where $\mu_{\min} = \cos\theta$, and $\tan\theta = \frac{k_\perp}{k_\parallel}$. In this paper, we use the red line as a rough approximation for the solid black line.

significantly reduces the bias due to PS anisotropy (e.g. Pober 2015). Restricting the k-space footprint comes at the cost of further increasing sample variance and thus further increases the benefits of using `21cmPSDenoiser`.

4.3 Mitigating sample variance with score-based diffusion

In this section, we first introduce the training database and the neural network architecture of `21cmPSDenoiser`. We then evaluate the neural network on a separate database of test samples and comment on its performance.

4.3.1 Simulated Dataset

We build a training database with over one hundred light cone realisations (i.e. IC samples) for each set of astrophysical parameters. We simulate a light cone with `21cmFASTv3`[12] [Mesinger and Furlanetto, 2007, Mesinger et al., 2011, Murray et al., 2020] and use "The Ultimate EoR Simulation Data AnalYser" (`tuesday`)[13], a wrapper around `powerbox`[14][Murray, 2018], to calculate the 2D PS. Our simulation boxes are 300 cMpc on a side, with a cell size of 1.5 cMpc. This is characteristic of the cosmological simulations used to interpret 21-cm observations (e.g. Muñoz et al. 2022, Nunhokee et al. 2025, Ghara et al. 2025).

We vary five astrophysical parameters $\theta_{\rm astro} = (f_{\rm esc,10}, f_{*,10}, M_{\rm turn}, L_{\rm X,<2keV}/{\rm SFR}, {\rm E}_0)$, introduced in Park et al. 2019:

- $f_{\rm esc,10}$: the amplitude of the power-law describing the ionising escape fraction $f_{\rm esc}(M_h) \in [0,1]$ to halo mass relation $f_{\rm esc}(M_h) = f_{\rm esc,10}(M_h/M_{10})^{\alpha_{\rm esc}}$, where $M_{10} = 10^{10} M_\odot$ and the index $\alpha_{\rm esc} = 0.5$ is fixed (e.g., Paardekooper et al. 2015, Kimm et al. 2017, Lewis et al. 2020);
- $f_{*,10}$: the amplitude of the stellar-to-halo mass relation (SHMR) normalised at M_{10}. Similarly to the ionising escape fraction, the faint-end SHMR is described by a power-law whose index we fix to $\alpha_* = 0.5$ (e.g. Mirocha et al. 2017, Munshi et al. 2017, 2021);
- $M_{\rm turn}[M_\odot]$: the characteristic halo mass scale below which the abundance of galaxies becomes exponentially suppressed to account for inefficient star formation in low-mass halos (e.g., Hui and Gnedin 1997, Springel and Hernquist 2003, Okamoto et al. 2008, Sobacchi and Mesinger 2013, Xu et al. 2016, Ocvirk et al. 2020, Ma et al. 2020);
- $\log_{10} \frac{L_{\rm X,<2keV}}{\rm SFR} \left[\frac{\rm erg\ s^{-1}}{\rm M_\odot yr^{-1}}\right]$: the X-ray luminosity escaping the galaxies is modelled as a power-law in energy, which we normalise with the soft-band (i.e. < 2 keV) X-ray luminosity per unit star formation rate (SFR). We fix the power law index of the X-ray SED to $\alpha_X = 1.0$ (e.g. Fragos et al. 2013, Pacucci et al. 2014, Das et al. 2017);
- $E_0[{\rm eV}]$: minimum energy of X-ray photons capable of escaping their host galaxy;

Well-established observations already provide some constraints for $f_{\rm esc}$, f_*, and $M_{\rm turn}$. To build our training set, we sample these three parameters from a posterior informed by UV luminosity functions from Hubble [Bouwens et al., 2015, 2016, Oesch et al., 2018], the Thomson scattering CMB optical depth from Planck [Planck Collaboration et al., 2020, Qin et al., 2020], and the Lyman forest dark fraction [McGreer et al., 2015]. The resulting posterior is shown in green in

[12] https://github.com/21cmfast/21cmFAST/
[13] https://github.com/21cmfast/tuesday
[14] https://github.com/steven-murray/powerbox

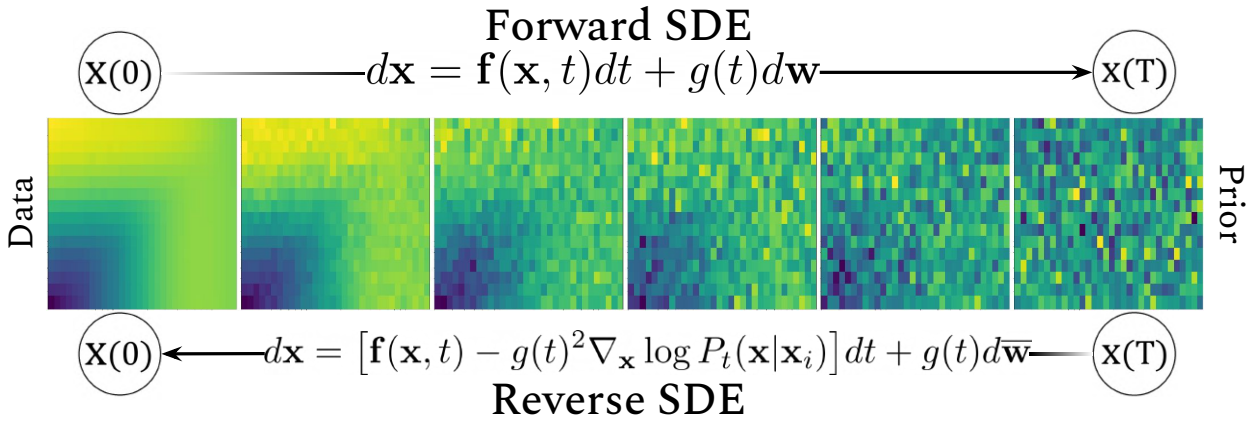

Figure 4.3: An illustration of the forward and backward diffusion processes used in `21cmPSDenoiser`. The forward process adds noise to a mean 21-cm 2D PS sampled from the data distribution (leftmost panel), transforming it into a pre-defined Gaussian prior distribution (rightmost panel). We can then write the reverse process that allows us to sample the Gaussian prior and generate a mean 2D PS, conditioned on the input realisation of the 2D PS.

Figure 8 of Breitman et al. 2024). For the remaining two parameters we sample a flat prior over $\log_{10} L_{\mathrm{X},<2\mathrm{keV}}/\mathrm{SFR} \in [38, 42]$ and $E_0 \in [200, 1500]$ (e.g. Furlanetto 2006, HERA Collaboration et al. 2023).

The database used for training and validation consists of roughly 900 parameter combinations with an average of roughly 100 realisations per parameter set. This amounts to ∼90k total light cones. For each light cone realisation, we calculate the 2D PS on cubic chunks over 40 redshift bins $z \in [5.5, 35]$. We re-bin the 2D PS to be linearly-spaced in log scale in both sky-plane and line-of-sight modes. To minimise the number of empty bins while keeping the 2D PS dimensions in powers of two that are more convenient for the NN, we choose 32 $k_\perp$ bins and 16 $k_\parallel$ bins. Due to the re-binning of $k_\perp$ into log-spaced bins, we end up with three empty bins: the second, fourth and fifth. We fill the power in those bins by interpolating with `SciPy` [Virtanen et al., 2020] which can produce artefacts in individual realisations (e.g. see horizontal bright yellow stripes in the top left plot of Figure 4.4). These artefacts, however, do not affect the mean 21-cm PS as they are just another effect of sample variance and get averaged out. Since the NN is redshift-agnostic, the final database consists of about 3.6M 2D PS. This database is then split into 90% training set with 10% left for the validation. The test set is made separately from the training and validation sets and is described in Section 4.3.3.

We pre-process the data by applying min-max normalisation on the log of the 21-cm PS and then rescaling it to $[-1, 1]$:

$$\left(\log_{10} \Delta^2_{21,}\right)_{\mathrm{norm}} = 2 \times \frac{\log_{10} \Delta^2_{21} - \min(\log_{10} \Delta^2_{21})}{\max(\log_{10} \Delta^2_{21}) - \min(\log_{10} \Delta^2_{21})} - 1. \tag{4.4}$$

The minimum and maximum are each a scalar obtained over the entire training and validation databases.

4.3.2 Denoiser architecture and training

We interpret sample variance as a form of non-uniform, (mildly) non-Gaussian (e.g. Mondal et al. 2015, Shaw et al. 2019) "noise" added to the target mean power spectrum. Obtaining the mean cylindrical PS from a single ("noisy") realisation is thus akin to denoising a 2D image.

This is a common task in image processing, for which machine learning is known to outperform traditional methods (e.g. Kawar et al. 2022) such as Gaussian filter smoothing or principal component analysis (PCA). The idea is to train a neural network (NN) to find a smooth function that interpolates through fluctuating data points, with minimal loss of intrinsic scatter. In our usage case, the small scales with negligible sample variance (c.f. top right corner of Figure 2) can serve as anchor to the NN prediction of the mean for the noisy larger scales (c.f. bottom left corner of Figure 2).

Here, we adopt a state-of-the-art NN architecture for image generation that has also been shown to excel at image denoising (e.g. Kawar et al. 2022): a score-based diffusion generative model (e.g. Sohl-Dickstein et al. 2015, Song and Ermon 2019, Song and Ermon 2020, Song and Ermon 2019, Ho et al. 2020, Song et al. 2020). In Figure 4.3, we illustrate the general idea behind diffusion models:

- Left to right: the forward diffusion process can be interpreted as a continuous noise-adding stochastic process where we corrupt a mean 21-cm PS from the training set with Gaussian noise with increasing variance according to a chosen variance schedule until it is transformed into a sample from a standard normal prior distribution. This stochastic process can be written as a solution to a stochastic differential equation (SDE):

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}, \tag{4.5}$$

  where $\mathbf{x} = \big(x(0), \ldots, x(T)\big)$ is the diffused data (i.e. a 2D PS) at a given time in the diffusion process, $t \in [0, T]$, with $x(0)$ denoting a sample from the data distribution of mean 2D PS and $x(T)$ a sample from the Gaussian prior. $\mathbf{w}$ is a Wiener process (aka Brownian motion). The drift function $\mathbf{f}(\mathbf{x}, t)$ and the diffusion coefficient $g(t)$ are hyperparameters of our model. We choose the most standard drift function and diffusion coefficient leading to the variance preserving (VP) SDE [Song et al., 2020], which is the continuous-time limit of the variance schedule in the Denoising Diffusion Probabilistic Model (DDPM, Ho et al. 2020).

- Right to left: in order to generate new data samples from Gaussian prior samples, we need to reverse the forward process, which can be done by solving the following SDE (Anderson 1982):

$$d\mathbf{x} = \big[\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log P_t(\mathbf{x}|\mathbf{x}_i)\big]dt + g(t)d\overline{\mathbf{w}}, \tag{4.6}$$

  where the only unknown is the score function $\nabla_{\mathbf{x}} \log P_t(\mathbf{x}|\mathbf{x}_i)$ of the probability density function (PDF) P of the data x (in this case, the 2D PS means) explicitly conditioned on a 2D PS realisation $\mathbf{x}_i$ in addition to the continuous diffusion time index t. At $t = 0$, the entire procedure boils down to mapping any input 2D PS realisation to its corresponding mean 2D PS.

In order to solve the reverse SDE written above and generate a mean 2D PS from a given 2D PS realisation, we train a neural network to learn the score function. Here, we use a U-Net autoencoder architecture[15] similar to Ho et al. 2020 implemented with `PyTorch` (e.g. Paszke et al. 2019. U-Nets have been developed for image segmentation as they are efficient in recognising local information in images over a range of scales [Ronneberger et al., 2015]. We train the NN with the continuous-time generalisation of the standard DDPM loss function (see Eq. 7 in Song et al. 2020) with the Adam optimiser [Kingma and Ba, 2017].

[15]The model architecture is based on the `PyTorch` implementation available here: https://github.com/lucidrains/denoising-diffusion-pytorch that is in turn based on the original implementation from Ho et al. 2020 here: https://github.com/hojonathanho/diffusion

During training, the NN learns to accurately predict the score for various levels of noise. It is precisely due to this training method that score-based diffusion models, and diffusion models in general, perform very well on tasks such as image denoising: the denoising task itself is part of the training process. Once the NN is trained to accurately predict the score for varying noise levels indexed by t, we can generate new mean 2D PS samples given one 2D PS realisation by solving this reverse-time SDE. Since our goal is to use the score-based diffusion model as part of an inference pipeline, we solve the reverse SDE via the probability-flow ODE method [Song et al., 2020]. This algorithm modestly sacrifices accuracy for a significant speed increase. In order to average over the network error, our final estimate of the mean 2D PS corresponds to the median obtained over 200 samples (i.e. draws from the prior) from `21cmPSDenoiser` for a given input. For these choices, we obtain a mean 2D PS estimate from a single realisation in ∼6s on a V100 GPU, which reduces to $\sim$ 2s when taking advantage of GPU vectorization by denoising multiple realisations at once.

4.3.3 Performance on the test set

The test set is composed of 50 parameter combinations distinct from the parameters in the training and validation sets with about 200 realisations per parameter and 40 redshift bins for a total of 400k 2D PS. Since the test set is significantly smaller than the training set, we can afford to have a larger number of realisations for each parameter, allowing a more accurate estimate of the true, target mean PS. We assess the performance of `21cmPSDenoiser` using the fractional error (FE) evaluated on a random 40k batch of the test set:

$$\mathrm{FE}(\%) = \left| \frac{\Delta^2_{21,\mathrm{test}} - \Delta^2_{21,\mu}}{\max\left(0.01, \Delta^2_{21,\mathrm{test}}\right)} \right| \times 100 \,, \tag{4.7}$$

where $\Delta^2_{21,\mathrm{test}}$ is the target mean power spectrum, obtained by averaging over 200 realisations per parameter combination, and $\Delta^2_{21,\mu}$ is a mean 2D PS estimate e.g. from `21cmPSDenoiser`. Note that to avoid the fractional error exploding at small power, we floor the denominator to 0.01 mK^2, which is an order of magnitude smaller than the accuracy of the `21cmFAST` simulator itself (e.g., Mesinger et al. 2011, Zahn et al. 2011). We calculate the FE for each realisation from each parameter combination and redshift.

In Figure 4.4, we illustrate the impact of sample variance using a single PS realisation (left column) and after applying `21cmPSDenoiser` (right column). This realisation was chosen from the θ_{mock} parameter vector in the test set. This parameter combination is consistent with the most recent constraints from observations such as the Lyman-α forest (e.g.Qin et al. 2025 and UV luminosity functions [Bouwens et al., 2015, 2016, Oesch et al., 2018]; see Section 4.6.1 for more details. The top row shows the cylindrical PS realisation at $z \sim 6.1$ and the resulting mean estimate obtained from `21cmPSDenoiser` using the realisation shown on the left. In the middle row, we compare them to the test set mean PS via the fractional error where $\Delta^2_{21,\mu} = \Delta^2_{21,\mathrm{NN}}$ on the right and $\Delta^2_{21,\mu} = \Delta^2_{21,i}$ on the left. The plotted PS realisation has a median fractional error very close to that of the entire test set for `21cmPSDenoiser` (see top half of Table 4.1) and can therefore be considered representative. We see that the fractional sample variance error on large scales can be reduced by over an order of magnitude by calling `21cmPSDenoiser`.

To put these errors into better perspective, the bottom row shows the square root of the squared deviation from the test set mean normalised by realistic HERA sensitivities (see Section 4.6 for more details). The left plot on the bottom row shows that the deviation from the mean due to sample variance is comparable to or larger than HERA sensitivity at large scales $k \sim 0.1$ Mpc^{-1} and remains significant ($\sim 10\%$) up to much smaller scales $k \sim 0.5$ Mpc^{-1}. The right plot shows

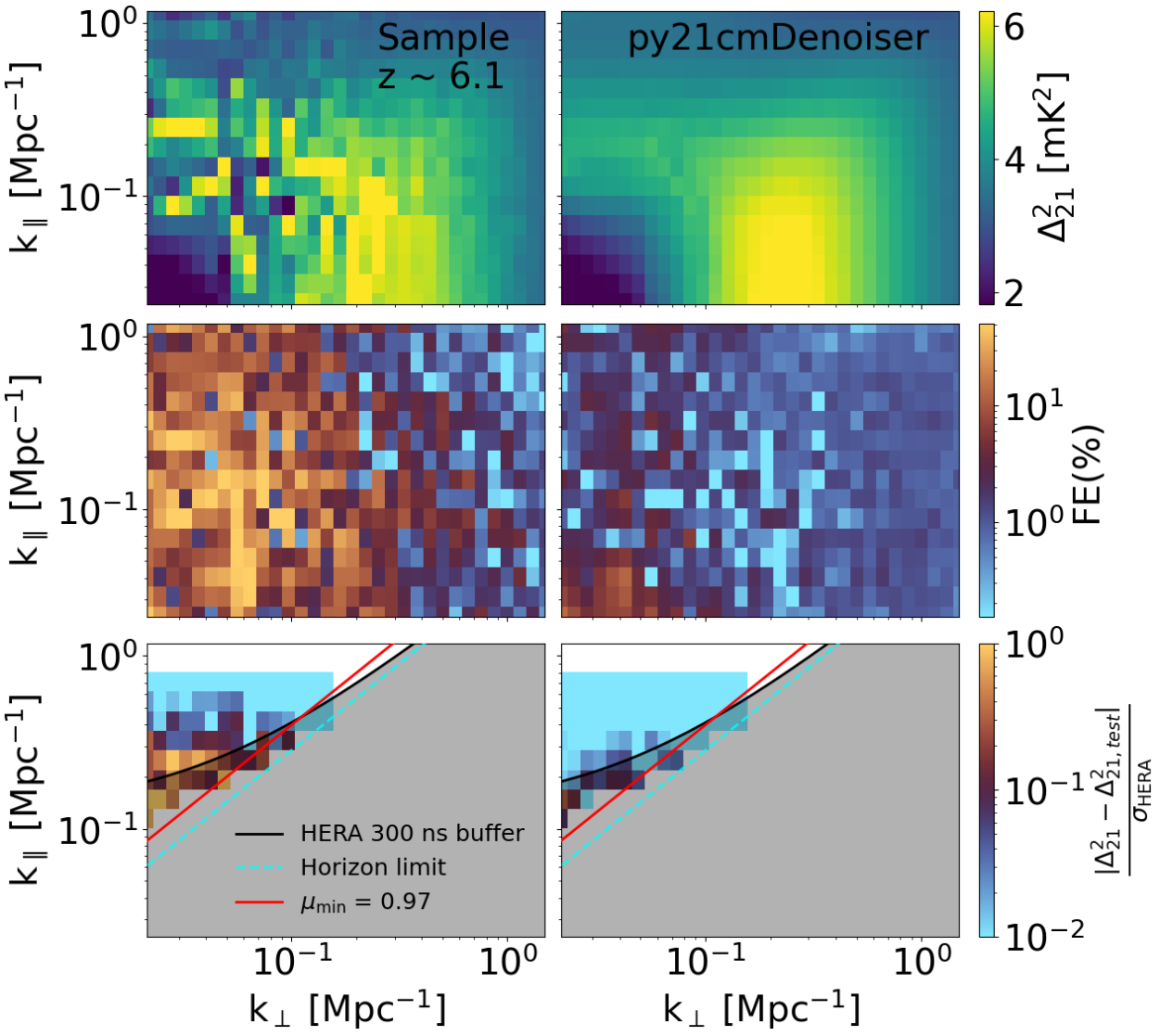

Figure 4.4: Top row: PS sample on the left and NN mean estimate from this same PS sample as input on the right. Middle row: fractional error with respect to the mean PS obtained from an ensemble average of about 200 PS realisations for the sample (left) and `21cmPSDenoiser` (right). Bottom row: error as a fraction of the HERA noise level at the same redshift for the sample (left) and for the `21cmPSDenoiser` (right).

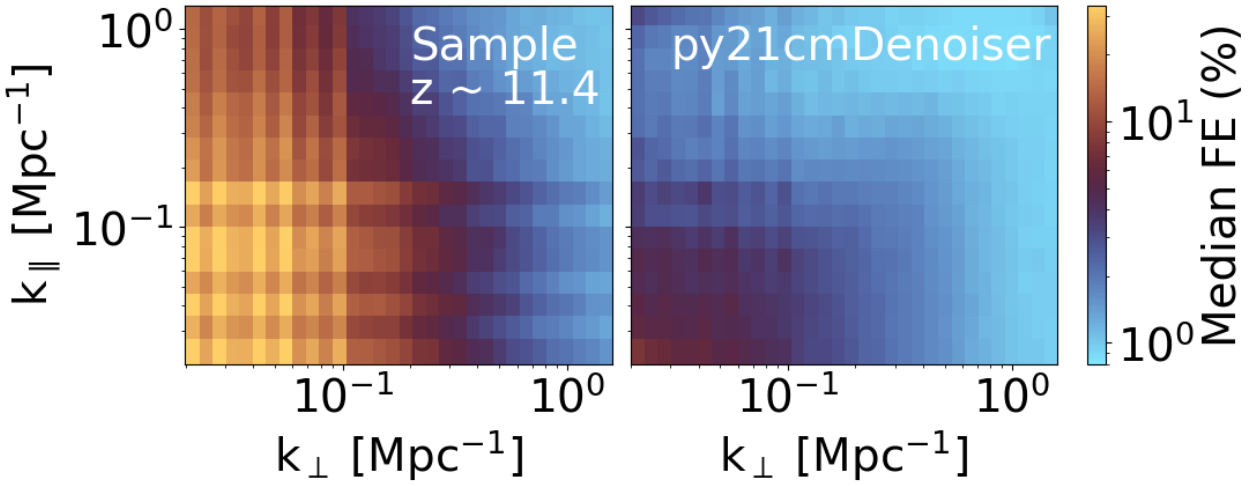

Figure 4.5: Median fractional error on $\sim$ 2.5k test samples at redshift $z \sim 11.4$. The left plot evaluates the FE directly on the PS realisations, while the right plot evaluates it on the output from `21cmPSDenoiser`. The striped pattern in the right plot occurs due to the binning scheme, where certain bins have more samples (and thus less sample variance) than others.

that applying `21cmPSDenoiser` reduces sample variance enough that the residual deviation from the mean is very small ($\lesssim 1\%$) in comparison to the instrument sensitivity.

In Figure 4.5, we show the median of the FE computed from $\sim$ 2.5k test samples at redshift $z \sim 11.4$. In the left plot, we evaluate the FE over the PS realisations while in the right plot we evaluate it over the `21cmPSDenoiser` mean estimate. We can see that when taking the median over $\sim$ 2.5k test samples, individual bins can deviate from the mean by over 30% on large scales and that `21cmPSDenoiser` reduces this deviation down to $\sim$ 3%. Individual bins from single PS realisations as shown in Figure 4.4, on the other hand, can deviate from the mean by over 50% on large scales, which gets reduced down to $\sim$ 5% by `21cmPSDenoiser`. The striped pattern in the right plot occurs due to the binning scheme, where certain bins have more samples (and thus less sample variance) than others.

4.4 Comparing `21cmPSDenoiser` to Fixing & Pairing

In this section we compare our results against fixing and pairing (F&P; e.g. Angulo and Pontzen 2016, Pontzen et al. 2016, Acharya et al. 2024), the benchmark technique for mitigating sample variance. F&P involves pairing a simulation to a given fixed simulation by reversing the sign of the initial matter overdensity field , $\delta \equiv \rho/\bar{\rho} - 1$, such that $\delta_{\rm paired}(\mathbf{k}) = -\delta_{\rm fixed}(\mathbf{k})$, for every wave mode k. Each mode whose amplitude is above the mean initial matter PS in one simulation has a counterpart whose amplitude is equally below the mean in the other simulation. Averaging F&P simulations by construction would yield the mean initial matter PS for the chosen cosmology, but it has also been shown to give a good estimate of the mean PS of evolved fields, including galaxy and line intensity maps (e.g. Angulo and Pontzen 2016, Pontzen et al. 2016, Villaescusa-Navarro et al. 2018).

Nevertheless, F&P has two main shortcomings: (i) it is still computationally expensive, as it doubles the cost of the inference; and (ii) due to non-linear evolution of the 21-cm signal, fixing and pairing becomes less effective with decreasing redshift, which is where current instruments are most sensitive [Acharya et al., 2024].

We build a database of 20 F&P pairs for each of the 50 parameter combinations in the test

$\mu_{\min} = 0$	Sample	F&P	NN
Median FE (%)	7.0	4.9	2.0
Median AE (mK^2)	0.3	0.2	0.1
FE 68 % CL (%)	26.8	21.4	8.2
AE 68 % CL (mK^2)	9.9	7.1	2.8
$\mu_{\min} = 0.97$	Sample	F&P	NN
Median FE (%)	12.1	9.0	2.2
Median AE (mK^2)	0.9	0.6	0.2
FE 68 % CL (%)	30.9	25.0	7.8
AE 68 % CL (mK^2)	26.2	19.4	4.6

Table 4.1: Median and 68 % CL on the fractional error and absolute error of a ∼40k random batch from the test set over 40 redshift bins $\in [5.3, 33]$. The first column compares the mean PS directly to the PS realisation. The second and third columns compare the fixing and pairing method and `21cmPSDenoiser` respectively. The top half of the table ($\mu_{\min} = 0$) compares all cylindrical PS k-modes, while the bottom half does so only over the $\mu_{\min} = 0.97$ region.

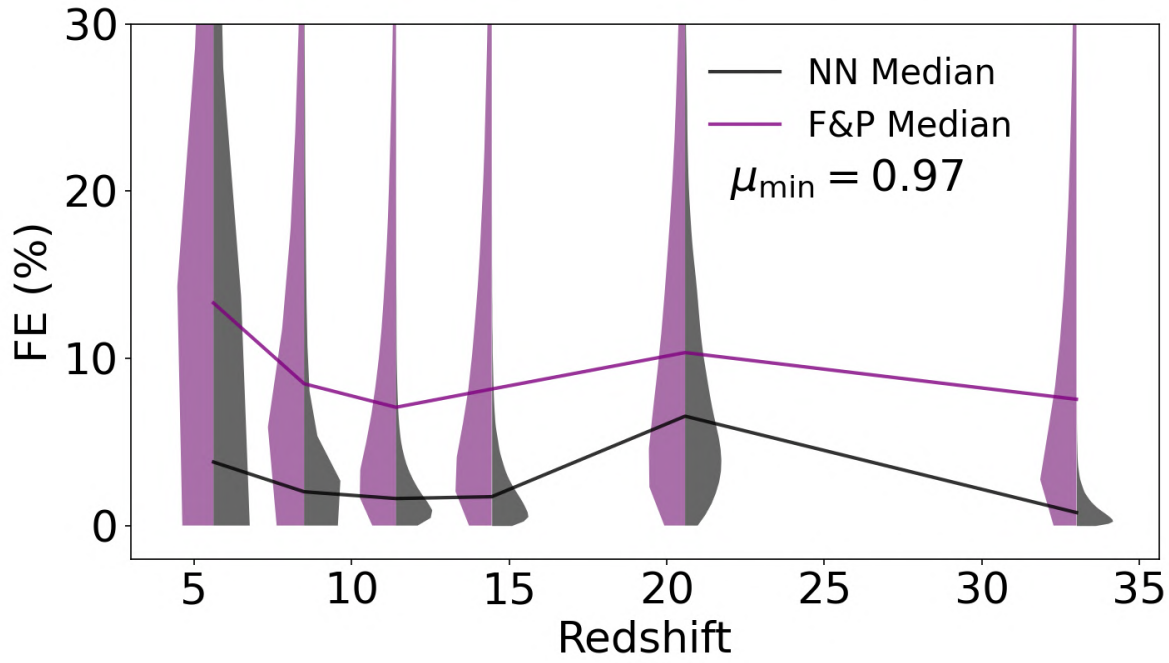

Figure 4.6: PDFs of the fractional error as a function of redshift for fixing and pairing (purple / left violins) and `21cmPSDenoiser` (black / right violins) computed on PS modes above $\mu_{\min} > 0.97$ (see red line in Figure 4.2). The distributions were generated using the 50 parameter samples comprising our test set.

set. In Figure 4.6, we compare the fractional error for the fixing and pairing method (purple / left violins) with `21cmPSDenoiser` (black / right violins). We perform this comparison on power spectra with mean power greater than 0.01 mK^2 for $\sim$ 17k samples cropped at $\mu_{\rm min} = 0.97$. We can see that `21cmPSDenoiser` has median fractional error that is about 5% lower than fixing and pairing across most redshifts[16].

We summarise the results in Table 4.1, where we show the median and 68 % confidence limits (CLs) of the FE and absolute error (AE) of: (i) individual samples; (ii) the mean estimated from fixing and pairing; and (iii) the mean estimated from `21cmPSDenoiser`. The top half of the table shows the FE and AE evaluated over all cylindrical modes, while the bottom half does so over the region above $\mu_{\rm min} = 0.97$. We see that `21cmPSDenoiser` results in a FE that is a factor of ~ 2.5 (~ 4) smaller than F&P over all k-space ($\mu_{\rm min} = 0.97$ region). Additionally, unlike F&P, using `21cmPSDenoiser` in inference comes at essentially no additional computational cost.

We note that there are cases where `21cmPSDenoiser` can produce an output that is farther away from the mean PS than even the noisy sample it was given as input. However, such cases are rare, as demonstrated in this section. In general, the denoiser is more likely to fail when the input is uninformative, as may happen when the neutral fraction is very low and the 21-cm PS has little power. As such, we generally recommend applying `21cmPSDenoiser` on signal with neutral fraction $\gtrsim 0.05$ or with mean power $\gtrsim 10^{-2}$ mK^2 in the input realisation.

4.5 Application to other simulators

As motivated in the introduction, one benefit of our approach is that, unlike emulation, it is model- and simulator-agnostic. `21cmPSDenoiser` can in principle operate on any 2D PS, regardless of what model or simulator was used to make it. In this section, we test `21cmPSDenoiser` on hydrodynamic radiative transfer (RT) simulations from Acharya et al. 2024. Similar to THESAN (e.g., Garaldi et al. 2022, Kannan et al. 2022), these simulations implement moving mesh hydrodynamics with AREPO (e.g., Springel 2010, Weinberger et al. 2020), and radiative transfer of ionising photons with AREPO-RT (e.g., Kannan et al. 2019). Furthermore, as they lack Lyman band and X-ray radiative transfer, these simulations make the simplifying assumption of a homogeneously-saturated spin temperature: $T_S \gg T_R$ in Eq. 4.2. These hydro RT simulations are therefore very different from those used in training `21cmPSDenoiser`, both in terms of the source model as well as the simulator.

From Acharya et al. 2024, we have a total of 40 simulations varying ICs for one parameter set, where five of these also have an additional paired simulation. We post-process the coeval cubes from these simulations into light cones using `tools21cm`[17] (Giri et al. 2020). These resulting light cones have a resolution of 0.373 cMpc and a box size of 95.5 cMpc. `21cmPSDenoiser`, however, is not capable of generalising to a different k-space footprint or resolution. As such, we calculate the 2D PS from these light cones with the goal to match the 2D PS from the training database as closely as possible. This involves taking these light cones and first downsampling them by a factor of four so that they have roughly the same resolution as the `21cmFAST` simulations in the training set. Next, we calculate the 2D PS on the same k-space binning as the training set. However, since these boxes are about three times smaller, this leaves many large-scale mode bins empty. We pad these large-scale bins by copying over the power from the closest non-empty bin at smaller scales. This padded 2D PS is then passed to `21cmPSDenoiser`[18].

[16] The increase in FE at redshift $\sim$ 20 is due to small values of the 21-cm signal following the dark ages and preceding Lyman-α coupling.

[17] https://github.com/sambit-giri/tools21cm

[18] The scripts used to make these power spectra use the publicly available `tuesday` package and can be

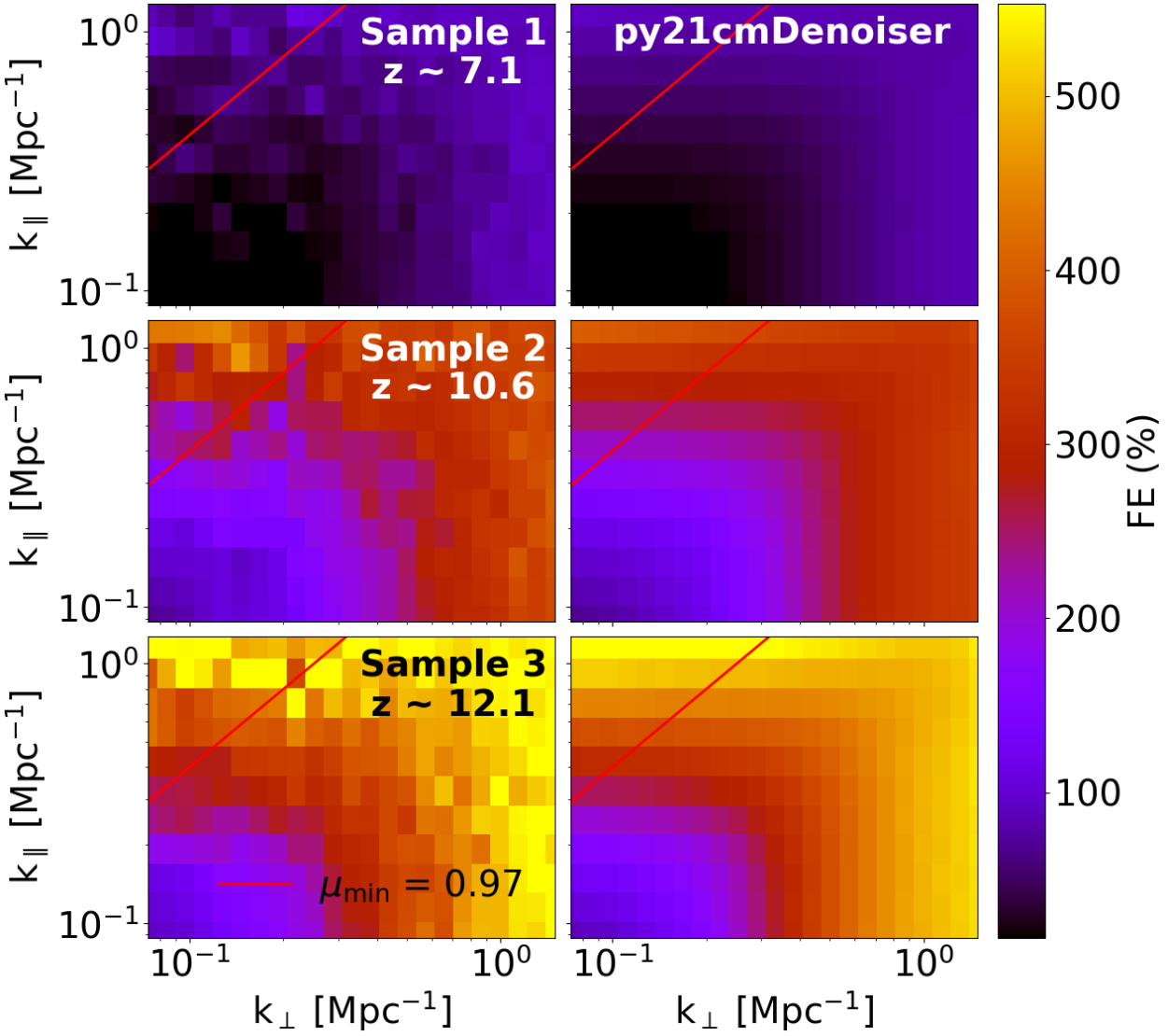

Figure 4.7: 21-cm PS realisations from a hydro RT simulation (left column) and the corresponding mean estimate (right column) obtained after passing the realisation on the left to `21cmPSDenoiser` (which was only trained on `21cmFAST`). Each row corresponds to different initial conditions and redshift.

In Figure 4.7, we show the de-noised PS on the right, corresponding to the single hydro RT realisation input shown on the left. Each row corresponds to a different realisation. Note that in all plots, we crop the k-space footprint to only include the non-empty bins (i.e. we do not consider the padded bins as we have no larger hydro RT simulations against which to compare them). We note that with only 35 realisations and 5 F&P pairs, the estimate of the true target mean is not as accurate as it is for the test set, where we have ~ 200 realisations per parameter. As a consequence, the performance of the NN reported in this section is not as accurate as in previous sections.

We see from Figure 4.7 that `21cmPSDenoiser` is flexible enough to mitigate sample variance on 21-cm power spectra from different astrophysical models and different simulators. We confirm this by looking at Figure 4.8, where we plot the FE of the samples shown in Figure 4.7. The mean PS above $\mu_{\rm min} = 0.97$ is recovered to an median accuracy of about $\sim 4\%$ over all redshifts and PS realisations for the denoiser, while sample variance introduces a median deviation of about $\sim 10\%$ in the same region. Note that the samples shown have an average FE at the redshift plotted that is slightly above the average FE over all available realisations. Figure 4.8 also shows that on the largest and smallest scales, `21cmPSDenoiser` tends to under-predict the mean power: a behaviour not seen in the test set cases. It is reasonable, however, that the out-of-distribution performance is weaker than that seen for the test set. One could further improve the generalization of `21cmPSDenoiser` by fine tuning it on 21-cm power spectra from different models and simulators.

4.6 Application to inference

In this section, we test the performance of `21cmPSDenoiser` by performing inference on a HERA mock observation. We compare the traditional state-of-the-art inference pipeline with the improvements introduced in this work (i.e. left vs right of Figure 4.1).

4.6.1 Mock HERA observation

We choose the mock parameter set $\theta_{\rm mock}$ with $(\log_{10} f_{\rm esc,10}, \log_{10} f_{*,10}, M_{\rm turn}, \log_{10} L_{\rm X,<2keV}/{\rm SFR}, {\rm E}_0) = (-1.23, -1.36, 8.26, 40.59, 1.40{\rm keV})$ out of the 50 parameter sets available in the test set as it has an EoR history that is consistent with that inferred from Lyman-α forest data [Qin et al., 2025], and matches UV luminosity functions at $z = 6$–10 [Bouwens et al., 2015, 2016, Oesch et al., 2018]. In addition to thermal variance from the instrument, an observation of the 21-cm PS is subject to cosmic variance due to the fact that there is only one Universe with its own set of initial conditions to observe i.e. we do not observe the expectation value of the 21-cm PS (over all possible observable universes), but rather a realisation with a finite volume. Cosmic variance has the most effect at the largest scales of an observation as there are fewer Fourier modes to be observed given the finite size of the observable Universe. However, since HERA observations span a much larger field than our forward models, cosmic variance on the scales of the simulation is negligible. Our mock cosmic signal is therefore taken to be the mean 21-cm PS, averaged over the 200 realisations of ICs for $\theta_{\rm mock}$. As our observational summary statistic, we spherically-average this mean cylindrical PS above $\mu_{\rm min} = 0.97$, mimicking the observational footprint (c.f. Figure 4.2), obtaining the 1D PS at $z = 5.6, 6.1, 6.9, 7.9, 9.1, 10.4, 10.8, 16.8,$ and 22.7.

We use `21cmSense` (Pober et al. 2013, 2014, Murray et al. 2024) to forecast the sensitivity for two full seasons of phase II HERA observations where we observe for 94 nights per season

found in the github repo https://github.com/DanielaBreitman/21cmPSDenoiser

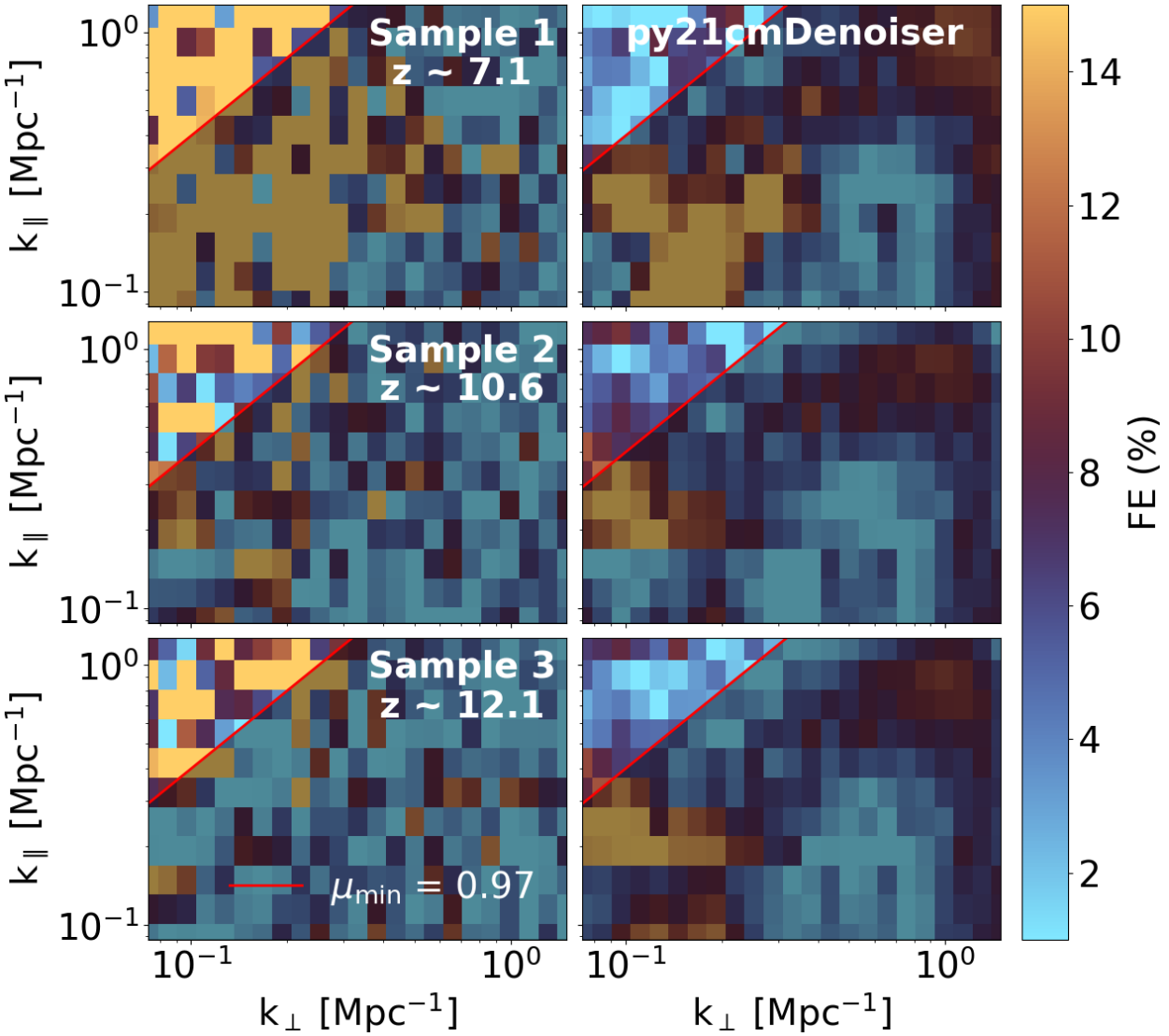

Figure 4.8: Fractional error on the 21-cm power spectra shown in Figure 4.7. We caution that the mean PS is estimated from a relatively small number of realisations (35 unpaired ICs + 5 pairs of ICs).

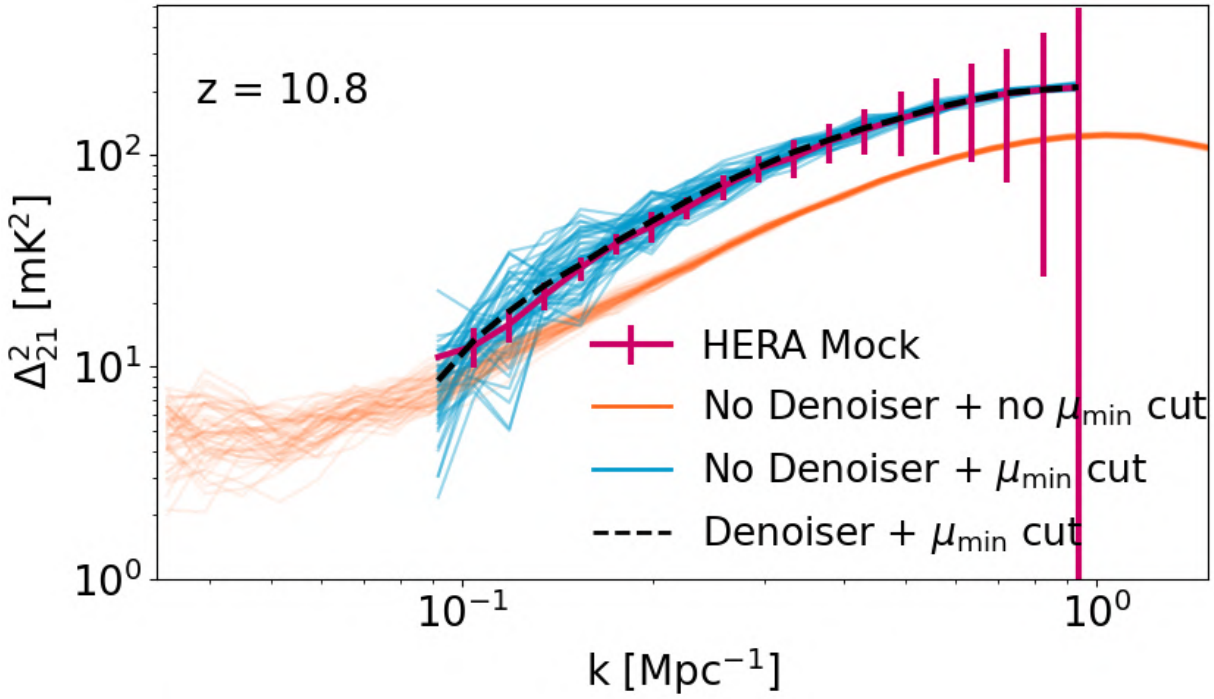

Figure 4.9: Spherically averaged 21-cm power spectra at $z = 10.8$ corresponding to a parameter vector in our test set, $\theta_{\rm mock}$. Pink points and error bars correspond to a mock ~ 2256 h observation with HERA (see text for details). In orange, we plot different realisations, varying ICs at a fixed $\theta_{\rm mock}$, but spherically-averaging the 2D PS down to $\mu_{\rm min} = 0$ as is commonly done when forward-modelling. In blue, we plot these realisations but instead averaging only down to $\mu_{\rm min} = 0.97$ to account for the HERA footprint in cylindrical space (see red line in Figure 4.2). Excising low μ modes removes the bias seen in the orange curves, but dramatically increases the sample variance. The dashed black line corresponds to the output of `21cmPSDenoiser` from a single 2D PS realisation, averaged down to $\mu_{\rm min} = 0.97$. We see that applying `21cmPSDenoiser` mitigates both the bias and the sample variance.

for a total of ~ 2256 integration hours (see HERA Collaboration et al. 2023 and the appendix in Breitman et al. 2024 for more details). The forecast assumes that the number of observed hours is the same in both seasons. The only difference between the two seasons is the number of operating antennas that increased from 140 in the 2022-2023 season to 180 in the 2023-2024 season. To combine the two observations, we sum the total integration times as well as the uv coverage of both seasons. We then use these combined integration time and uv-coverage to obtain the thermal variance of the HERA instrument over both seasons. We also include the cosmic variance of the observation in the error budget, adding it to the thermal noise in quadrature: $\sigma_{\rm sens} = \sqrt{\sigma^2_{\rm thermal} + \sigma^2_{\rm cosmic}}$, where we refer to $\sigma_{\rm sens}$ as the sensitivity.

After this procedure, we are left with HERA mock 1D 21-cm power spectra at all redshifts, together with the associated sensitivities. In Figure 4.9, we plot the mock observation at $z = 10.8$ as the pink points with 1σ error bars. In orange, we show different PS realisations with $\theta = \theta_{\rm mock}$, where all modes (i.e. $\mu_{\rm min} = 0$) of the 2D spectrum have been included in the average, instead of the modes above $\mu_{\rm min} = 0.97$ used for the mock. We see that there is significant scatter between the realisations at the largest scales even when we include all available modes. Aside from the realisation-to-realisation scatter between the orange curves, we find that they are biased by a

factor of ~ 3, revealing the effects of anisotropy when the correct $\mu_{\rm min}$ is neglected.[19]

In blue, we show PS realisations also with $\theta = \theta_{\rm mock}$, but now averaging over only the modes 'observed' in the mock with $\mu_{\rm min} = 0.97$. Unlike the orange curves, these do not display a significant bias with respect to the mock data. However, limiting the modes over which to perform the averaging results in substantial sample variance at small k. The level of sample variance is well in excess of the observational error bars and we thus expect it to limit the inferred parameter constraints (as we confirm below).

The dashed black line shows the mean PS obtained from `21cmPSDenoiser` derived from a single realisation. Similarly to the blue curves, this de-noised PS is averaged only over modes above $\mu_{\rm min} = 0.97$.

The de-noised PS agrees very well with the mock data, with no obvious bias and a dramatic reduction of the effects of sample variance (seen by the reduction in bin-to-bin variance) compared with the blue curves.

4.6.2 Inference set-up

To quantify the qualitative trends seen in Figure 4.9, we use our HERA mock observation to perform inferences under different approximations:

(i) No `21cmPSDenoiser` and no $\mu_{\rm min}$ cut – this corresponds to the current approach of spherically-averaging the forward-modelled 2D PS down to $\mu_{\rm min} = 0$ using a single realisation of the ICs (c.f. orange curves in Figure 4.9);

(ii) No `21cmPSDenoiser` with $\mu_{\rm min} = 0.97$ cut – this also uses a single realisation of the 2D PS, but averages down to the same $\mu_{\rm min} = 0.97$ as is used for the mock data (c.f. blue curves in Figure 4.9);

(iii) `21cmPSDenoiser` with $\mu_{\rm min} = 0.97$ cut – this uses a single realisation of the 2D PS passed to `21cmPSDenoiser` to obtain the mean 2D PS, before averaging down to the correct $\mu_{\rm min} = 0.97$ (c.f. dashed black curve in Figure 4.9).

In all of the inferences, the likelihood is constructed by multiplying individual likelihoods based on four observables:

(i) 21-cm 1D PS: following Figure 4.1, we forward model the 2D PS $\Delta^2_{21,i}(k_\perp, k_\parallel, z)$. We then estimate the mean PS: $\mu(\theta) = \begin{cases} \text{Denoiser}(\Delta^2_{21,i}) \text{ with NN} \\ \Delta^2_{21,i} \text{ otherwise.} \end{cases}$
We then average the mean 2D PS to 1D, weighting each 2D bin by the number of 3D Fourier modes it contains, N_k:

$$\Delta^2_{21}(k_i) = \frac{1}{\sum_{k\in\mathcal{K}_i} N_k(k_\perp, k_\parallel)} \sum_{k\in\mathcal{K}_i} N_k(k_\perp, k_\parallel) \times \Delta^2_{21}(k_\perp, k_\parallel), \tag{4.8}$$

where $\mathcal{K}_i$ is the set of $k = \sqrt{k_\perp^2 + k_\parallel^2} \in k_i$[20]. We then apply a window function W (e.g., Liu et al. 2014a,b, Gorce et al. 2023) calculated with `hera-pspec`[21] to the forward-modelled

[19] We note that we chose $z = 10.8$ explicitly to highlight the impact of PS anisotropy; its impact at other redshifts corresponding to an advanced stage of the EoR and whose bins span a slower redshift evolution of the signal would be smaller (e.g. Mesinger et al. 2011, Mao et al. 2012, Datta et al. 2014).

[20] For more details, see `cylindrical_to_spherical` in `tuesday`.

[21] https://github.com/HERA-Team/hera_pspec/

1D PS. The window function converts the forward-modelled 21-cm PS to an observed PS by including instrumental effects such as the chromaticity of the baselines and the beam, especially affecting large-scale modes. The likelihood function is a Gaussian

$$\log \mathcal{L}(\Delta^2_{21,\mathrm{mock}}|\theta) = -\frac{1}{2}\sum_{f,k} \frac{\left(W\cdot\mu(\theta) - W\cdot\Delta^2_{21,\mathrm{mock}}\right)^2}{\sigma^2}, \tag{4.9}$$

summed over all frequency bands f and k-bins , where the total variance σ^2 includes the sensitivity of the mock observation σ^2_{sens} and a contribution from the forward model σ^2_{fm} coming from either Poisson sample variance or the mean NN error:

$$\sigma^2 = \sigma^2_{\mathrm{sens}} + \sigma^2_{\mathrm{fm}}, \text{ where} \tag{4.10}$$

$$\sigma_{\mathrm{fm}} = \begin{cases} \Delta^2_{21}(\theta) \times \sigma_{\mathrm{denoiser}} \text{ with } \texttt{21cmPSDenoiser} \\ \Delta^2_{21}(\theta)/\sqrt{N} \text{ otherwise.} \end{cases} \tag{4.11}$$

(ii) UV luminosity functions: we include UV LFs at $z = 6, 7, 8$ and 10 based on Hubble data (Bouwens et al. 2015, 2016, Oesch et al. 2018) as in previous works (e.g., HERA Collaboration et al. 2022a, Breitman et al. 2024). This likelihood function is also assumed to be Gaussian.

(iii) Thomson scattering optical depth to the CMB: we include a Gaussian likelihood centred around $\tau_e = 0.0569^{+0.0081}_{-0.0086}$ based on the median and 68% credible interval (CI) from the posterior obtained in Qin et al. 2020 from their re-analysis of Planck Collaboration et al. 2020 data.

(iv) Lyman forest dark fraction: this term compares the proposed model's global neutral fraction at $z = 5.9$ with the upper bound $\overline{x}_{\mathrm{HI}} < 0.06 \pm 0.05$ at 68% CI obtained with the model-independent QSO dark fraction method (McGreer et al. 2015). The likelihood function is unity if the proposed global neutral fraction is below the upper bound at $z = 5.9$, then it decreases as a one-sided Gaussian for higher values of $\overline{x}_{\mathrm{HI}}$.

For computational convenience, we fix E_0 and L_X to the true values in θ_{mock} and perform the inference over the remaining three astrophysical parameters for which we assume flat priors over the following ranges: (i) $\log_{10} f_{*,10} \in [-2, -0.5]$; (ii) $\log_{10} f_{\mathrm{esc},10} \in [-3, 0]$; and (iii) $\mathrm{M}_{\mathrm{turn}}[\mathrm{M}_\odot] \in [8, 10]$.

We run the inferences with the `21cmMC` [22] (e.g., Greig and Mesinger 2015, 2017, 2018) package using the `MultiNest` (e.g. Feroz et al. 2009) [23] sampler. Each inference requires about 15k likelihood evaluations. In the following section, we show and discuss the posterior from each of these inferences.

4.6.3 Inference results

In Figure 4.10 we show the marginal posteriors of our three inferences, together with the true values, θ_{mock} (marked in pink). In orange, we show the posterior for inference (i). As foreshadowed by the orange curves in Figure 4.9, we see that the posterior is indeed biased, due to the mismatch of the 2D PS footprints of the observation and forward model. The bias is at the level

[22] https://github.com/21cmfast/21CMMC

[23] We choose `MultiNest` over `UltraNest` because the former requires significantly fewer likelihood evaluations than the latter and because we expect a simple ellipsoidal posterior.

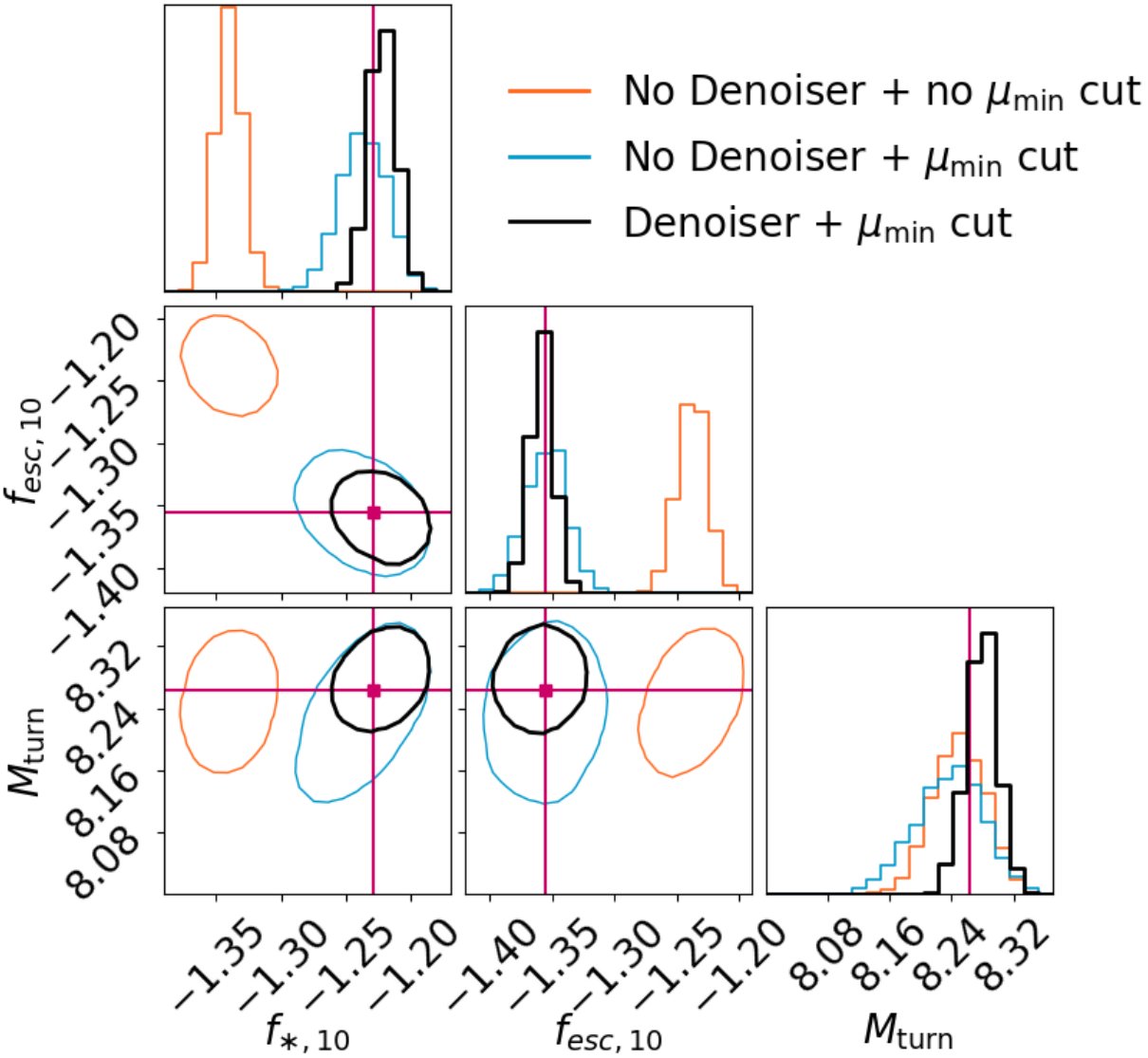

Figure 4.10: 1D and 2D marginal posteriors from three inferences described in the text: (i) traditional state-of-the-art analysis using a single IC realisation and spherically-averaging the 2D PS down to $\mu_{\rm min} = 0$ (orange); (ii) using a single IC realisation but spherically-averaging the 2D PS down to the same $\mu_{\rm min} = 0.97$ used for the mock data (blue); and (iii) applying `21cmPSDenoiser` to a single realisation in order to mitigate sample variance followed by the $\mu_{\rm min}$ cut to mitigate PS anisotropy (black). The 2D contours show 95% CIs. The pink lines show the true parameters $\theta_{\rm mock}$.

of $\sim 10\%$ in the inferred ionising escape fraction and stellar fraction parameters. More dramatic however is the overconfidence of the biased posteriors, with the true values being outside $\sim 10\sigma$ for both $f_{*,10}$ and $f_{\rm esc,10}$.

In blue, we show the posterior for inference (ii). Again, as foreshadowed by the correspondingly-coloured curves in Figure 4.9, this posterior is unbiased; however, it is notably wider than the posterior for inference (iii), shown in black. This highlights the impact of sample variance. Applying `21cmPSDenoiser` to a realisation of the 2D PS during inference tightens the inferred 1D marginal parameter constraints by $\sim 50\%$. While these quantitative results are HERA-specific due to our HERA-motivated choice of $\mu_{\rm min}$, we expect the qualitative trends to hold for any instrument.

4.7 Conclusion

In this work, we study the consequences of two common approximations made on the 21-cm PS likelihood in Bayesian inference problems:

(i) Replacing the 21-cm PS mean with a sample from a single realisation;

(ii) Averaging the 3D Fourier modes within a k bin over all orientations, instead of matching those actually observed by 21-cm PS experiments.

In order to relax both of these assumptions, we developed `21cmPSDenoiser`, a score-based diffusion model that provides an estimate of the mean cylindrical 21-cm PS given a single realisation. Unlike emulators, `21cmPSDenoiser` is not tied to a particular model or simulator since its input is a (model-agnostic) realisation of the 2D 21-cm PS. `21cmPSDenoiser` outperforms state-of-the-art analytical approaches such as fixing and pairing, the benchmark technique for sample variance mitigation. Individual 2D PS realisations can deviate from the mean by over 50% at scales relevant to current interferometers. `21cmPSDenoiser` reduces this deviation to $\sim 2\%$, over a factor of 2 better than fixing and pairing, and at almost no additional cost ($\sim$ 2s per iteration). Moreover, we test `21cmPSDenoiser` on 21-cm power spectra from a completely different simulator and astrophysical model. We find that it produces reasonable power spectra and produces a mean estimate that is ~ 2.5 times more accurate than the 2D PS realisation itself above $\mu_{\rm min} = 0.97$, the region of cylindrical PS space most relevant to current observations.

We test `21cmPSDenoiser` by applying it in a realistic inference context. First, we simulate a realistic HERA mock 21-cm PS observation with `21cmSense`. Then, we run a set of three inferences: (i) a classical inference such as previous state-of-the-art inferences; (ii) an improved inference where we solve only the first issue mentioned above; and (iii) where we solve both issues. We find that inference (i), the typical state-of-the-art inference method, produces a highly overconfident and biased posterior with a bias of over 10σ for two of the three parameters. In inference (ii), we crop the k-space of our cylindrical 21-cm PS closer to that of observations by applying a cut at $\mu_{\rm min} = 0.97$, and, as expected, we obtain an unbiased but wider posterior. Finally, we apply both `21cmPSDenoiser` and the cut at $\mu_{\rm min} = 0.97$ in inference (iii), and obtain an unbiased posterior that is on average 50% narrower for each parameter. We thus explicitly show that `21cmPSDenoiser` would benefit HERA data analysis in the very near future. Our proposed method, however, would also be highly applicable for upcoming experiments such as the SKA, especially for its earliest observation modes that are limited to substation layouts of e.g. 12 m or 18 m that enable the observation of much larger scales than the full array composed of

35 m stations[24].

We leave to future work to apply score-based diffusion models toward other sample-variance limited observations, such as large-scale structure surveys, and quantify the improvement they provide over previous machine learning methods such as convolutional neural networks (e.g., de Santi and Abramo 2022).

Acknowledgements

D.B. thanks Laurence Perrault-Levasseur and Yashar Hezaveh for useful discussions and comments on an early draft of this project, as well as Alexandre Adam for assistance in implementing the score-based diffusion model. We thank Adrian Liu for useful discussions and comments on the manuscript. We gratefully acknowledge computational resources of the Center for High Performance Computing (CHPC) at SNS. A.M. acknowledges support from the Ministry of Universities and Research (MUR) through the PRIN project "Optimal inference from radio images of the epoch of reionization", and the PNRR project "Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing". S. G. M. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101067043. In addition to the packages references in the text, this work made use of the open-source Python packages `NumPy` [Harris et al., 2020], `Matplotlib` [Hunter, 2007], `Astropy` [Astropy Collaboration et al., 2013, 2018, 2022].

Data Availability

Both `21cmPSDenoiser` and `tuesday` are on publicly accessible github repositories, as well as available for installation as a Python package using `pip`.

[24]For an example of SKA sensitivity for the 18 m substation layout, see the 21cmSense tutorial for simulating SKA sensitivities.

Chapter 5

Future outlook

In this thesis, we use machine learning as a tool to improve and facilitate the interpretation of cosmic dawn and epoch of reionisation observations within a Bayesian framework. In particular, in Chapters 2 and 3, we use machine learning to emulate otherwise computationally expensive simulations, thus speeding up the inference runtime by over a factor of four. As discussed in Section 1.4, a wide range of observables can be used to constrain the CD/EoR epochs, making the emulation of additional observables a natural and important direction for future research. Incorporating multiple new observables into the inference framework would sharpen constraints and enhance the constraining power of current 21-cm PS upper limits. Achieving this will require improving simulators so that they can generate a broader spectrum of summary statistics. A particularly promising extension is line-intensity mapping (LIM): integrating LIM into the forward-modelling pipeline would involve developing routines that generate line-intensity maps, which can then be compressed into common forms reported by experiments, such as power spectra. These, in turn, could be easily incorporated into emulators, enabling joint inference across multiple spectral lines and significantly expanding the scope of astrophysical and cosmological constraints.

In preparation for the high-quality 21-cm maps anticipated with the SKA, an important research direction is the emulation of more informative 21-cm statistics that capture the intrinsic non-Gaussianity of the signal. This could range from emulating full lightcones to summary statistics such as bispectra. While existing studies have quantified the improvement in constraining power from these more informative statistics, they typically do not perform synergistic inferences across multiple probes, nor do they address the challenge of combining different summaries extracted from the same instrument, which are not statistically independent. A natural way to overcome this last limitation is to adopt a simulation-based inference framework. Unlike the Gaussian likelihood assumption employed in this thesis, SBI is well suited to handling correlated, non-Gaussian summaries. As upper limits from current experiments continue to improve — and particularly once first detections are made — SBI will become an increasingly essential tool for extracting the full information content of the 21-cm signal.

Perhaps the greatest challenge for inference is model misspecification: Bayesian methods are known to fail if the forward model differs from the true data-generating process. This is a pressing concern, given the unavoidable mismatch between our simplified forward models and the complex physics of the real Universe. Misspecification can arise from incomplete astrophysical prescriptions, numerical approximations, or neglected instrumental systematics. Left unaddressed, it can bias posteriors and produce overconfident or misleading constraints. Standard Bayesian frameworks assume that the forward model is exact, and therefore cannot diagnose or mitigate such issues on their own. In practice, robustness demands cross-checking multiple

models, marginalising over theoretical uncertainties, or adopting more flexible approaches such as simulation-based inference, which can incorporate nuisance parameters or stochastic elements to absorb discrepancies. With the advent of high-precision data from SKA, tackling model misspecification will be critical: the reliability of our conclusions will ultimately be limited by the assumptions embedded in our models.

A detection is no longer a distant prospect, with SKA-Low having observed its first light in March 2025 and already in its commissioning phase. Building on the lessons from SKA pathfinders, an important research priority is to improve forecasts for upcoming observations. For imaging instruments such as SKA, this requires realistically simulating the thermal noise contribution to observed maps within a full imaging pipeline. However, most existing forecasting studies rely on overly idealized imaging setups, limiting their predictive power. In particular, residual systematics—rarely included in forecasts or in forward-modelling pipelines—remain a serious challenge for current experiments. Developing more realistic, end-to-end forecasts that incorporate these complexities will be essential both for interpreting SKA data and for maximizing its scientific return.

The SKA has a relatively large field of view and therefore probes an enormous cosmic volume. As discussed in 4, sample variance in forward models is a major concern for both hydrodynamic and semi-analytical simulations that aim to constrain CD/EoR physics with the 21-cm power spectrum. To enable accurate forecasts on all scales, we need simulations that cover volumes comparable to those observed. To address this, I am currently working on the next iteration of the Evolution of State (EOS) project (see previous iterations: Mesinger et al. 2016, Muñoz et al. 2022). The aim of EOS is to deliver the largest available simulation boxes that incorporate state-of-the-art astrophysics and cosmology. For this new iteration, I am running a 2.1 Gpc box at a resolution of 1.5 Mpc using the new 21cmFASTv4 [Davies et al., 2025]. Even with a semi-analytical code, carrying out a simulation of this size demands immense computational resources. Once completed, this flagship simulation will serve as a benchmark for forecasting upcoming 21-cm surveys and will provide a community resource for years to come.

I believe that community is a cornerstone of scientific progress. For this reason, I remain committed to making all of my code public and to developing well-designed, well-documented Python packages for community use. By sharing ready-to-use, high-quality tools, we allow researchers to focus their efforts on new scientific questions rather than reinventing the wheel, ultimately accelerating progress across the field. I am convinced that embracing this practice is a necessary step for our field to reach its full potential.

To conclude, the 21-cm signal remains one of the most promising probes of the early Universe, with the potential to transform our understanding of cosmic dawn and the epoch of reionisation. In recent years, immense progress has been made in both modelling and data analysis, bringing us to the threshold of a first detection. The work presented in this thesis contributes to this collective effort by developing machine learning tools to accelerate inference, mitigate sample variance, and enable synergistic analyses with multiple observables. As upper limits tighten and we approach the first detection, the challenge will increasingly shift from obtaining data to correctly interpreting it. This will require not only more realistic simulations and end-to-end forecasting pipelines, but also robust inference frameworks that can handle model misspecification and correlated, non-Gaussian statistics. At the same time, community-driven practices such as open-source packages will be essential to maximise scientific return and ensure that the progress made is broadly accessible.

Bibliography

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

T. Abel, G. L. Bryan, and M. L. Norman. The Formation of the First Star in the Universe. *Science*, 295(5552):93–98, Jan. 2002. doi: 10.1126/science.295.5552.93.

A. Acharya, E. Garaldi, B. Ciardi, and Q.-b. Ma. Cosmic variance suppression in radiation-hydrodynamic modelling of the reionization-era 21-cm signal. *MNRAS*, 529(4):3793–3805, Apr. 2024. doi: 10.1093/mnras/stae782.

J. Albert, C. Balázs, A. Fowlie, W. Handley, N. Hunt-Smith, R. Ruiz de Austri, and M. White. A comparison of Bayesian sampling algorithms for high-dimensional particle physics and cosmology applications. *Computer Physics Communications*, 315:109756, Oct. 2025. doi: 10.1016/j.cpc.2025.109756.

J. Alsing, B. Wandelt, and S. Feeney. Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology. *MNRAS*, 477(3):2874–2885, July 2018. doi: 10.1093/mnras/sty819.

B. D. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(82)90051-5.

R. E. Angulo and A. Pontzen. Cosmological N-body simulations with suppressed variance. *MNRAS*, 462(1):L1–L5, Oct. 2016. doi: 10.1093/mnrasl/slw098.

Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil, J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. L. Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher. Astropy: A community Python package for astronomy. *A&A*, 558:A33, Oct. 2013. doi: 10.1051/0004-6361/201322068.

Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. Vand erPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud,

C. Ardelean, T. Babej, Y. P. Bach, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, L. Corrales, D. Crichton, D. D'Avella, C. Deil, É. Depagne, J. P. Dietrich, A. Donath, M. Droettboom, N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž. Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastropietro, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, V. Zabalza, and Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJ*, 156(3):123, Sept. 2018. doi: 10.3847/1538-3881/aabc4f.

Astropy Collaboration, A. M. Price-Whelan, P. L. Lim, N. Earl, N. Starkman, L. Bradley, D. L. Shupe, A. A. Patil, L. Corrales, C. E. Brasseur, M. N"othe, A. Donath, E. Tollerud, B. M. Morris, A. Ginsburg, E. Vaher, B. A. Weaver, J. Tocknell, W. Jamieson, M. H. van Kerkwijk, T. P. Robitaille, B. Merry, M. Bachetti, H. M. G"unther, T. L. Aldcroft, J. A. Alvarado-Montes, A. M. Archibald, A. B'odi, S. Bapat, G. Barentsen, J. Baz'an, M. Biswas, M. Boquien, D. J. Burke, D. Cara, M. Cara, K. E. Conroy, S. Conseil, M. W. Craig, R. M. Cross, K. L. Cruz, F. D'Eugenio, N. Dencheva, H. A. R. Devillepoix, J. P. Dietrich, A. D. Eigenbrot, T. Erben, L. Ferreira, D. Foreman-Mackey, R. Fox, N. Freij, S. Garg, R. Geda, L. Glattly, Y. Gondhalekar, K. D. Gordon, D. Grant, P. Greenfield, A. M. Groener, S. Guest, S. Gurovich, R. Handberg, A. Hart, Z. Hatfield-Dodds, D. Homeier, G. Hosseinzadeh, T. Jenness, C. K. Jones, P. Joseph, J. B. Kalmbach, E. Karamehmetoglu, M. Kaluszy'nski, M. S. P. Kelley, N. Kern, W. E. Kerzendorf, E. W. Koch, S. Kulumani, A. Lee, C. Ly, Z. Ma, C. MacBride, J. M. Maljaars, D. Muna, N. A. Murphy, H. Norman, R. O'Steen, K. A. Oman, C. Pacifici, S. Pascual, J. Pascual-Granado, R. R. Patil, G. I. Perren, T. E. Pickering, T. Rastogi, B. R. Roulston, D. F. Ryan, E. S. Rykoff, J. Sabater, P. Sakurikar, J. Salgado, A. Sanghi, N. Saunders, V. Savchenko, L. Schwardt, M. Seifert-Eckert, A. Y. Shih, A. S. Jain, G. Shukla, J. Sick, C. Simpson, S. Singanamalla, L. P. Singer, J. Singhal, M. Sinha, B. M. SipHocz, L. R. Spitler, D. Stansby, O. Streicher, J. Sumak, J. D. Swinbank, D. S. Taranu, N. Tewary, G. R. Tremblay, M. d. Val-Borro, S. J. Van Kooten, Z. Vasovi'c, S. Verma, J. V. de Miranda Cardoso, P. K. G. Williams, T. J. Wilson, B. Winkel, W. M. Wood-Vasey, R. Xue, P. Yoachim, C. Zhang, A. Zonca, and Astropy Project Contributors. The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *ApJ*, 935(2):167, Aug. 2022. doi: 10.3847/1538-4357/ac7c74.

E. Bañados, B. P. Venemans, C. Mazzucchelli, E. P. Farina, F. Walter, F. Wang, R. Decarli, D. Stern, X. Fan, F. B. Davies, J. F. Hennawi, R. A. Simcoe, M. L. Turner, H.-W. Rix, J. Yang, D. D. Kelson, G. C. Rudie, and J. M. Winters. An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5. *Nature*, 553(7689):473–476, Jan. 2018. doi: 10.1038/nature25180.

R. Barkana and A. Loeb. In the beginning: the first sources of light and the reionization of the universe. *Phys. Rep.*, 349(2):125–238, July 2001. doi: 10.1016/S0370-1573(01)00019-9.

R. Barkana and A. Loeb. Light-cone anisotropy in 21-cm fluctuations during the epoch of reionization. *MNRAS*, 372(1):L43–L47, Oct. 2006. doi: 10.1111/j.1745-3933.2006.00222.x.

G. D. Becker, M. Rauch, and W. L. W. Sargent. The Evolution of Optical Depth in the Lyα Forest: Evidence Against Reionization at z~6. *ApJ*, 662(1):72–93, June 2007. doi: 10.1086/517866.

G. D. Becker, J. S. Bolton, P. Madau, M. Pettini, E. V. Ryan-Weber, and B. P. Venemans. Evidence of patchy hydrogen reionization from an extreme Lyα trough below redshift six. *MNRAS*, 447 (4):3402–3419, Mar. 2015. doi: 10.1093/mnras/stu2646.

P. S. Behroozi and J. Silk. A Simple Technique for Predicting High-redshift Galaxy Evolution. *ApJ*, 799(1):32, Jan. 2015. doi: 10.1088/0004-637X/799/1/32.

C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *ApJS*, 208(2):20, Oct. 2013. doi: 10.1088/0067-0049/208/2/20.

H. T. J. Bevins, W. J. Handley, A. Fialkov, E. de Lera Acedo, and K. Javid. GLOBALEMU: a novel and robust approach for emulating the sky-averaged 21-cm signal from the cosmic dawn and epoch of reionization. *MNRAS*, 508(2):2923–2936, Dec. 2021. doi: 10.1093/mnras/stab2737.

S. Bharadwaj and S. S. Ali. The cosmic microwave background radiation fluctuations from HI perturbations prior to reionization. *MNRAS*, 352(1):142–146, July 2004. doi: 10.1111/j.1365-2966.2004.07907.x.

J. S. Bolton, M. G. Haehnelt, S. J. Warren, P. C. Hewett, D. J. Mortlock, B. P. Venemans, R. G. McMahon, and C. Simpson. How neutral is the intergalactic medium surrounding the redshift z = 7.085 quasar ULAS J1120+0641? *MNRAS*, 416(1):L70–L74, Sept. 2011. doi: 10.1111/j.1745-3933.2011.01100.x.

J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. *ApJ*, 379:440, Oct. 1991. doi: 10.1086/170520.

S. E. I. Bosman, X. Fan, L. Jiang, S. Reed, Y. Matsuoka, G. Becker, and M. Haehnelt. New constraints on Lyman-α opacity with a sample of 62 quasarsat z > 5.7. *MNRAS*, 479(1):1055–1076, Sept. 2018. doi: 10.1093/mnras/sty1344.

S. E. I. Bosman, F. B. Davies, G. D. Becker, L. C. Keating, R. L. Davies, Y. Zhu, A.-C. Eilers, V. D'Odorico, F. Bian, M. Bischetti, S. V. Cristiani, X. Fan, E. P. Farina, M. G. Haehnelt, J. F. Hennawi, G. Kulkarni, A. Mesinger, R. A. Meyer, M. Onoue, A. Pallottini, Y. Qin, E. Ryan-Weber, J.-T. Schindler, F. Walter, F. Wang, and J. Yang. Hydrogen reionization ends by z = 5.3: Lyman-α optical depth measured by the XQR-30 sample. *MNRAS*, 514(1):55–76, July 2022. doi: 10.1093/mnras/stac1046.

R. J. Bouwens, G. D. Illingworth, P. A. Oesch, M. Trenti, I. Labbé, L. Bradley, M. Carollo, P. G. van Dokkum, V. Gonzalez, B. Holwerda, M. Franx, L. Spitler, R. Smit, and D. Magee. UV Luminosity Functions at Redshifts z $\sim$ 4 to z $\sim$ 10: 10,000 Galaxies from HST Legacy Fields. *ApJ*, 803(1): 34, Apr. 2015. doi: 10.1088/0004-637X/803/1/34.

R. J. Bouwens, P. A. Oesch, I. Labbé, G. D. Illingworth, G. G. Fazio, D. Coe, B. Holwerda, R. Smit, M. Stefanon, P. G. van Dokkum, M. Trenti, M. L. N. Ashby, J. S. Huang, L. Spitler, C. Straatman, L. Bradley, and D. Magee. The Bright End of the z $\sim$ 9 and z $\sim$ 10 UV Luminosity Functions Using All Five CANDELS Fields*. *ApJ*, 830(2):67, Oct. 2016. doi: 10.3847/0004-637X/830/2/67.

J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh. An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature*, 555(7694):67–70, Mar. 2018. doi: 10.1038/nature25792.

D. Breitman, A. Mesinger, S. G. Murray, D. Prelogović, Y. Qin, and R. Trotta. 21CMEMU: an emulator of 21CMFAST summary observables. *MNRAS*, 527(4):9833–9852, Feb. 2024. doi: 10.1093/mnras/stad3849.

D. Breitman, A. Mesinger, S. G. Murray, and A. Acharya. Sample Variance Denoising in Cylindrical 21-cm Power Spectra. *arXiv e-prints*, art. arXiv:2507.12545, July 2025a. doi: 10.48550/arXiv.2507.12545.

D. Breitman, A. Mesinger, S. G. Murray, I. Nikolic, and R. Trotta. 21cmEMUv3: a score-based diffusion emulator of 21cmFAST summary observables including Pop II and Pop III stars. *arXiv e-prints*, Nov. 2025b.

V. Bromm and R. B. Larson. The First Stars. *ARA&A*, 42(1):79–118, Sept. 2004. doi: 10.1146/annurev.astro.42.053102.134034.

J. Buchner. A statistical test for Nested Sampling algorithms. *Statistics and Computing*, 26(1-2):383–392, Jan. 2016. doi: 10.1007/s11222-014-9512-y.

J. Buchner. Collaborative Nested Sampling: Big Data versus Complex Physical Models. *PASP*, 131 (1004):108005, Oct. 2019. doi: 10.1088/1538-3873/aae7fc.

J. Buchner. UltraNest - a robust, general purpose Bayesian inference engine. *The Journal of Open Source Software*, 6(60):3001, Apr. 2021. doi: 10.21105/joss.03001.

C. H. Bye, S. K. N. Portillo, and A. Fialkov. 21cmVAE: A Very Accurate Emulator of the 21 cm Global Signal. *ApJ*, 930(1):79, May 2022. doi: 10.3847/1538-4357/ac6424.

J. Cang, A. Mesinger, S. G. Murray, D. Breitman, Y. Qin, and R. Trotta. The EDGES measurement disfavors an excess radio background during the cosmic dawn. *arXiv e-prints*, art. arXiv:2411.08134, Nov. 2024. doi: 10.48550/arXiv.2411.08134.

E. Chapman and V. Jelić. Foregrounds and their mitigation. In *The Cosmic 21-cm Revolution*, 2514-3433, pages 6–1 to 6–29. IOP Publishing, 2019. ISBN 978-0-7503-2236-2. doi: 10.1088/2514-3433/ab4a73ch6. URL https://dx.doi.org/10.1088/2514-3433/ab4a73ch6.

CHIME Collaboration, M. Amiri, K. Bandura, T. Chen, M. Deng, M. Dobbs, M. Fandino, S. Foreman, M. Halpern, A. S. Hill, G. Hinshaw, C. Höfer, J. Kania, T. L. Landecker, J. MacEachern, K. Masui, J. Mena-Parra, N. Milutinovic, A. Mirhosseini, L. Newburgh, A. Ordog, U.-L. Pen, T. Pinsonneault-Marotte, A. Polzin, A. Reda, A. Renard, J. R. Shaw, S. R. Siegel, S. Singh, K. Vanderlinde, H. Wang, D. V. Wiebe, D. Wulf, and CHIME Collaboration. Detection of Cosmological 21 cm Emission with the Canadian Hydrogen Intensity Mapping Experiment. *ApJ*, 947(1):16, Apr. 2023. doi: 10.3847/1538-4357/acb13f.

T. R. Choudhury and A. Ferrara. Experimental constraints on self-consistent reionization models. *MNRAS*, 361(2):577–594, Aug. 2005. doi: 10.1111/j.1365-2966.2005.09196.x.

T. R. Choudhury, A. Paranjape, and S. E. I. Bosman. Studying the Lyman α optical depth fluctuations at z $\sim$ 5.5 using fast semi-numerical methods. *MNRAS*, 501(4):5782–5796, Mar. 2021. doi: 10.1093/mnras/stab045.

B. Clément, J. G. Cuby, F. Courbin, A. Fontana, W. Freudling, J. Fynbo, J. Gallego, P. Hibon, J. P. Kneib, O. Le Fèvre, C. Lidman, R. McMahon, B. Milvang-Jensen, P. Moller, A. Moorwood, K. K. Nilsson, L. Pentericci, B. Venemans, V. Villar, and J. Willis. Evolution of the observed Lyα luminosity function from z = 6.5 to z = 7.7: evidence for the epoch of reionization? *A&A*, 538: A66, Feb. 2012. doi: 10.1051/0004-6361/201117312.

A. Cohen, A. Fialkov, R. Barkana, and R. A. Monsalve. Emulating the global 21-cm signal from Cosmic Dawn and Reionization. *MNRAS*, 495(4):4845–4859, July 2020. doi: 10.1093/mnras/staa1530.

A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger. Fast and credible likelihood-free cosmology with truncated marginal neural ratio estimation. *J. Cosmology Astropart. Phys.*, 2022(9):004, Sept. 2022. doi: 10.1088/1475-7516/2022/09/004.

K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062, Dec. 2020. doi: 10.1073/pnas.1912789117.

A. Das, A. Mesinger, A. Pallottini, A. Ferrara, and J. H. Wise. High-mass X-ray binaries and the cosmic 21-cm signal: impact of host galaxy absorption. *MNRAS*, 469(1):1166–1174, July 2017. doi: 10.1093/mnras/stx943.

S. Das, T. Louis, M. R. Nolta, G. E. Addison, E. S. Battistelli, J. R. Bond, E. Calabrese, D. Crichton, M. J. Devlin, S. Dicker, J. Dunkley, R. Dünner, J. W. Fowler, M. Gralla, A. Hajian, M. Halpern, M. Hasselfield, M. Hilton, A. D. Hincks, R. Hlozek, K. M. Huffenberger, J. P. Hughes, K. D. Irwin, A. Kosowsky, R. H. Lupton, T. A. Marriage, D. Marsden, F. Menanteau, K. Moodley, M. D. Niemack, L. A. Page, B. Partridge, E. D. Reese, B. L. Schmitt, N. Sehgal, B. D. Sherwin, J. L. Sievers, D. N. Spergel, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. Thornton, H. Trac, and E. Wollack. The Atacama Cosmology Telescope: temperature and gravitational lensing power spectrum measurements from three seasons of data. *J. Cosmology Astropart. Phys.*, 2014(4):014, Apr. 2014. doi: 10.1088/1475-7516/2014/04/014.

K. K. Datta, G. Mellema, Y. Mao, I. T. Iliev, P. R. Shapiro, and K. Ahn. Light-cone effect on the reionization 21-cm power spectrum. *MNRAS*, 424(3):1877–1891, Aug. 2012. doi: 10.1111/j.1365-2966.2012.21293.x.

K. K. Datta, H. Jensen, S. Majumdar, G. Mellema, I. T. Iliev, Y. Mao, P. R. Shapiro, and K. Ahn. Light cone effect on the reionization 21-cm signal - II. Evolution, anisotropies and observational implications. *MNRAS*, 442(2):1491–1506, Aug. 2014. doi: 10.1093/mnras/stu927.

J. E. Davies, A. Mesinger, and S. Murray. Efficient simulation of discrete galaxy populations and associated radiation fields during the first billion years. *arXiv e-prints*, art. arXiv:2504.17254, Apr. 2025. doi: 10.48550/arXiv.2504.17254.

P. Dayal, A. Ferrara, J. S. Dunlop, and F. Pacucci. Essential physics of early galaxy formation. *MNRAS*, 445(3):2545–2557, Dec. 2014. doi: 10.1093/mnras/stu1848.

R. de Belsunce, S. Gratton, W. Coulton, and G. Efstathiou. Inference of the optical depth to reionization from low multipole temperature and polarization Planck data. *MNRAS*, 507(1): 1072–1091, Oct. 2021. doi: 10.1093/mnras/stab2215.

A. de Oliveira-Costa, M. Tegmark, B. M. Gaensler, J. Jonas, T. L. Landecker, and P. Reich. A model of diffuse Galactic radio emission from 10 MHz to 100 GHz. *MNRAS*, 388(1):247–260, July 2008. doi: 10.1111/j.1365-2966.2008.13376.x.

N. S. M. de Santi and L. R. Abramo. Improving cosmological covariance matrices with machine learning. *J. Cosmology Astropart. Phys.*, 2022(9):013, Sept. 2022. doi: 10.1088/1475-7516/2022/09/013.

D. R. DeBoer, A. R. Parsons, J. E. Aguirre, P. Alexander, Z. S. Ali, A. P. Beardsley, G. Bernardi, J. D. Bowman, R. F. Bradley, C. L. Carilli, C. Cheng, E. de Lera Acedo, J. S. Dillon, A. Ewall-Wice, G. Fadana, N. Fagnoni, R. Fritz, S. R. Furlanetto, B. Glendenning, B. Greig, J. Grobbelaar, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, M. Kariseb, S. A. Kohn, T. Lekalake, A. Liu, A. Loots, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. Martinot, N. Mathison, E. Matsetela, A. Mesinger, M. F. Morales, A. R. Neben, N. Patra, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Ringuette, J. Robnett, K. Rosie, R. Sell, C. Smith, A. Syce, M. Tegmark, N. Thyagarajan, P. K. G. Williams, and H. Zheng. Hydrogen Epoch of Reionization Array (HERA). *PASP*, 129(974):045001, Apr. 2017. doi: 10.1088/1538-3873/129/974/045001.

T. Developers. Tensorflow, May 2022. URL https://doi.org/10.5281/zenodo.6574269. Specific TensorFlow versions can be found in the "Versions" list on the right side of this page.
See the full list of authors on GitHub.

J. S. Dillon and S. Murray. HERA Memo #122: H6C (2023 season) Summary of Season Flags. Technical report, Nov. 2021. HERA Memo #122.

S. Dodelson and F. Schmidt. *Modern Cosmology*. 2020. doi: 10.1016/C2017-0-01943-2.

V. D'Odorico, E. Bañados, G. D. Becker, M. Bischetti, S. E. I. Bosman, G. Cupani, R. Davies, E. P. Farina, A. Ferrara, C. Feruglio, C. Mazzucchelli, E. Ryan-Weber, J. T. Schindler, A. Sodini, B. P. Venemans, F. Walter, H. Chen, S. Lai, Y. Zhu, F. Bian, S. Campo, S. Carniani, S. Cristiani, F. Davies, R. Decarli, A. Drake, A. C. Eilers, X. Fan, P. Gaikwad, S. Gallerani, B. Greig, M. G. Haehnelt, J. Hennawi, L. Keating, G. Kulkarni, A. Mesinger, R. A. Meyer, M. Neeleman, M. Onoue, A. Pallottini, Y. Qin, S. Rojas-Ruiz, S. Satyavolu, A. Sebastian, R. Tripodi, F. Wang, M. Wolfson, J. Yang, and M. V. Zanchettin. XQR-30: The ultimate XSHOOTER quasar sample at the reionization epoch. *MNRAS*, 523(1):1399–1420, July 2023. doi: 10.1093/mnras/stad1468.

J. Dorigo Jones, S. M. Bahauddin, D. Rapetti, J. Mirocha, and J. O. Burns. 21CMLSTM: A Fast Memory-based Emulator of the Global 21 cm Signal with Unprecedented Accuracy. *ApJ*, 977 (1):19, Dec. 2024. doi: 10.3847/1538-4357/ad8b20.

A. B. Drake, B. Guiderdoni, J. Blaizot, L. Wisotzki, E. C. Herenz, T. Garel, J. Richard, R. Bacon, D. Bina, S. Cantalupo, T. Contini, M. den Brok, T. Hashimoto, R. A. Marino, R. Pelló, J. Schaye, and K. B. Schmidt. MUSE deep-fields: the Ly α luminosity function in the Hubble Deep Field-South at $2.91 < z < 6.64$. *MNRAS*, 471(1):267–278, Oct. 2017. doi: 10.1093/mnras/stx1515.

D. J. Eisenstein and W. Hu. Power Spectra for Cold Dark Matter and Its Variants. *ApJ*, 511(1): 5–15, Jan. 1999. doi: 10.1086/306640.

J. D. Emberson, R. M. Thomas, and M. A. Alvarez. The Opacity of the Intergalactic Medium during Reionization: Resolving Small-scale Structure. *ApJ*, 763(2):146, Feb. 2013. doi: 10.1088/0004-637X/763/2/146.

N. Fagnoni, E. de Lera Acedo, D. R. DeBoer, Z. Abdurashidova, J. E. Aguirre, P. Alexander, Z. S. Ali, Y. Balfour, A. P. Beardsley, G. Bernardi, T. S. Billings, J. D. Bowman, R. F. Bradley, P. Bull, J. Burba, C. L. Carilli, C. Cheng, M. Dexter, J. S. Dillon, A. Ewall-Wice, R. Fritz, S. R. Furlanetto, K. Gale-Sides, B. Glendenning, D. Gorthi, B. Greig, J. Grobbelaar, Z. Halday, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Josaitis, A. Julius, N. S. Kern, J. Kerrigan, H. Kim, P. Kittiwisit, S. A. Kohn, M. Kolopanis, A. Lanman, P. L. Plante, T. Lekalake, A. Liu, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. E. Martinot, E. Matsetela, J. Mena Parra, A. Mesinger, M. Molewa, M. F. Morales, T. Mosiane, A. R. Neben, B. Nikolic, A. R. Parsons, N. Patra, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Robnett, K. Rosie, P. Sims, C. Smith, A. Syce, N. Thyagarajan, P. K. G. Williams, and H. Zheng. Understanding the HERA Phase I receiver system with simulations and its impact on the detectability of the EoR delay power spectrum. *MNRAS*, 500(1):1232–1242, Jan. 2021. doi: 10.1093/mnras/staa3268.

X. Fan, M. A. Strauss, R. H. Becker, R. L. White, J. E. Gunn, G. R. Knapp, G. T. Richards, D. P. Schneider, J. Brinkmann, and M. Fukugita. Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z∼6 Quasars. II. A Sample of 19 Quasars. *AJ*, 132 (1):117–136, July 2006. doi: 10.1086/504836.

F. Feroz, M. P. Hobson, and M. Bridges. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398(4):1601–1614, Oct. 2009. doi: 10.1111/j.1365-2966.2009.14548.x.

A. Fialkov, R. Barkana, E. Visbal, D. Tseliakhovich, and C. M. Hirata. The 21-cm signature of the first stars during the Lyman-Werner feedback era. *MNRAS*, 432(4):2909–2916, July 2013. doi: 10.1093/mnras/stt650.

D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP*, 125(925):306, Mar. 2013. doi: 10.1086/670067.

T. Fragos, B. Lehmer, M. Tremmel, P. Tzanavaris, A. Basu-Zych, K. Belczynski, A. Hornschemeier, L. Jenkins, V. Kalogera, A. Ptak, and A. Zezas. X-Ray Binary Evolution Across Cosmic Time. *ApJ*, 764(1):41, Feb. 2013. doi: 10.1088/0004-637X/764/1/41.

S. R. Furlanetto. The global 21-centimeter background from high redshifts. *MNRAS*, 371(2):867–878, Sept. 2006. doi: 10.1111/j.1365-2966.2006.10725.x.

S. R. Furlanetto, S. P. Oh, and F. H. Briggs. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Phys. Rep.*, 433(4-6):181–301, Oct. 2006. doi: 10.1016/j.physrep.2006.08.002.

E. Garaldi, R. Kannan, A. Smith, V. Springel, R. Pakmor, M. Vogelsberger, and L. Hernquist. The THESAN project: properties of the intergalactic medium and its connection to reionization-era galaxies. *MNRAS*, 512(4):4909–4933, June 2022. doi: 10.1093/mnras/stac257.

E. M. George, C. L. Reichardt, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H. M. Cho, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, J. Dudley, N. W. Halverson, N. L. Harrington, G. P. Holder, W. L. Holzapfel, Z. Hou, J. D. Hrubes, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, M. Lueker, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, M. Millea, L. M. Mocanu, J. J. Mohr, T. E. Montroy, S. Padin, T. Plagge, C. Pryke, J. E. Ruhl, K. K. Schaffer, L. Shaw, E. Shirokoff, H. G. Spieler, Z. Staniszewski, A. A. Stark, K. T. Story, A. van Engelen, K. Vanderlinde, J. D. Vieira, R. Williamson, and O. Zahn. A Measurement of Secondary Cosmic Microwave Background Anisotropies from the 2500 Square-degree SPT-SZ Survey. *ApJ*, 799 (2):177, Feb. 2015. doi: 10.1088/0004-637X/799/2/177.

R. Ghara, T. R. Choudhury, and K. K. Datta. 21 cm signal from cosmic dawn: imprints of spin temperature fluctuations and peculiar velocities. *MNRAS*, 447(2):1806–1825, Feb. 2015. doi: 10.1093/mnras/stu2512.

R. Ghara, S. K. Giri, G. Mellema, B. Ciardi, S. Zaroubi, I. T. Iliev, L. V. E. Koopmans, E. Chapman, S. Gazagnes, B. K. Gehlot, A. Ghosh, V. Jelić, F. G. Mertens, R. Mondal, J. Schaye, M. B. Silva, K. M. B. Asad, R. Kooistra, M. Mevius, A. R. Offringa, V. N. Pandey, and S. Yatawatta. Constraining the intergalactic medium at z = 9.1 using LOFAR Epoch of Reionization observations. *MNRAS*, 493(4):4728–4747, Feb. 2020. doi: 10.1093/mnras/staa487.

R. Ghara, S. Zaroubi, B. Ciardi, G. Mellema, S. K. Giri, F. G. Mertens, M. Mevius, L. V. E. Koopmans, I. T. Iliev, A. Acharya, S. A. Brackenhoff, E. Ceccotti, K. Chege, I. Georgiev, S. Ghosh, I. Hothi, C. Höfer, Q. Ma, S. Munshi, A. R. Offringa, A. K. Shaw, V. N. Pandey, S. Yatawatta, and M. Choudhury. Constraints on the state of the IGM at $z \sim 8 - 10$ using redshifted 21-cm observations with LOFAR. *arXiv e-prints*, art. arXiv:2505.00373, May 2025. doi: 10.48550/arXiv.2505.00373.

S. Giri, G. Mellema, and H. Jensen. Tools21cm: A python package to analyse the large-scale 21-cm signal from the Epoch of Reionization and Cosmic Dawn. *The Journal of Open Source Software*, 5(52):2363, Aug. 2020. doi: 10.21105/joss.02363.

S. K. Giri, A. Schneider, F. Maion, and R. E. Angulo. Suppressing variance in 21 cm signal simulations during reionization. *A&A*, 669:A6, Jan. 2023. doi: 10.1051/0004-6361/202244986.

A. Gorce, S. Ganjam, A. Liu, S. G. Murray, Z. Abdurashidova, T. Adams, J. E. Aguirre, P. Alexander, Z. S. Ali, R. Baartman, Y. Balfour, A. P. Beardsley, G. Bernardi, T. S. Billings, J. D. Bowman, R. F. Bradley, P. Bull, J. Burba, S. Carey, C. L. Carilli, C. Cheng, D. R. DeBoer, E. d. L. Acedo, M. Dexter, J. S. Dillon, N. Eksteen, J. Ely, A. Ewall-Wice, N. Fagnoni, R. Fritz, S. R. Furlanetto, K. Gale-Sides, B. Glendenning, D. Gorthi, B. Greig, J. Grobbelaar, Z. Halday, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, M. Kariseb, N. S. Kern, J. Kerrigan, P. Kittiwisit, S. A. Kohn, M. Kolopanis, A. Lanman, P. La Plante, A. Loots, D. H. E. MacMahon, L. Malan, C. Malgas, K. Malgas, B. Marero, Z. E. Martinot, A. Mesinger, M. Molewa, M. F. Morales, T. Mosiane, A. R. Neben, B. Nikolic, H. Nuwegeld, A. R. Parsons, N. Patra, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Robnett, K. Rosie, P. Sims, H. Swarts, N. Thyagarajan, P. van Wyngaarden, P. K. G. Williams, and H. Zheng. Impact of instrument and data characteristics in the interferometric reconstruction of the 21 cm power spectrum. *MNRAS*, 520(1):375–391, Mar. 2023. doi: 10.1093/mnras/stad090.

B. Greig and A. Mesinger. 21CMMC: an MCMC analysis tool enabling astrophysical parameter studies of the cosmic 21 cm signal. *MNRAS*, 449(4):4246–4263, June 2015. doi: 10.1093/mnras/stv571.

B. Greig and A. Mesinger. Simultaneously constraining the astrophysics of reionization and the epoch of heating with 21CMMC. *MNRAS*, 472(3):2651–2669, Dec. 2017. doi: 10.1093/mnras/stx2118.

B. Greig and A. Mesinger. 21CMMC with a 3D light-cone: the impact of the co-evolution approximation on the astrophysics of reionization and cosmic dawn. *MNRAS*, 477(3):3217–3229, July 2018. doi: 10.1093/mnras/sty796.

B. Greig, A. Mesinger, and E. Bañados. Constraints on reionization from the z = 7.5 QSO ULASJ1342+0928. *MNRAS*, 484(4):5094–5101, Apr. 2019. doi: 10.1093/mnras/stz230.

B. Greig, A. Mesinger, F. B. Davies, F. Wang, J. Yang, and J. F. Hennawi. IGM damping wing constraints on reionization from covariance reconstruction of two $z \gtrsim 7$ QSOs. *MNRAS*, 512 (4):5390–5403, June 2022. doi: 10.1093/mnras/stac825.

B. Greig, D. Prelogović, Y. Qin, Y.-S. Ting, and A. Mesinger. Inferring astrophysical parameters using the 2D cylindrical power spectrum from reionization. *MNRAS*, 533(2):2530–2545, Sept. 2024. doi: 10.1093/mnras/stae1984.

Z. Haiman and G. L. Bryan. Was Star Formation Suppressed in High-Redshift Minihalos? *ApJ*, 650(1):7–11, Oct. 2006. doi: 10.1086/506580.

Y. Harikane, K. Nakajima, M. Ouchi, H. Umeda, Y. Isobe, Y. Ono, Y. Xu, and Y. Zhang. Pure spectroscopic constraints on uv luminosity functions and cosmic star formation history from 25 galaxies at zspec = 8.61–13.20 confirmed with jwst/nirspec. *The Astrophysical Journal*, 960(1):56, dec 2023. doi: 10.3847/1538-4357/adob7e. URL https://dx.doi.org/10.3847/1538-4357/ad0b7e.

C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2.

C. Heinrich and W. Hu. Reionization effective likelihood from Planck 2018 data. *Phys. Rev. D*, 104 (6):063505, Sept. 2021. doi: 10.1103/PhysRevD.104.063505.

HERA Collaboration, Z. Abdurashidova, J. E. Aguirre, P. Alexander, Z. S. Ali, Y. Balfour, R. Barkana, A. P. Beardsley, G. Bernardi, T. S. Billings, J. D. Bowman, R. F. Bradley, P. Bull, J. Burba, S. Carey, C. L. Carilli, C. Cheng, D. R. DeBoer, M. Dexter, E. de Lera Acedo, J. S. Dillon, J. Ely, A. Ewall-Wice, N. Fagnoni, A. Fialkov, R. Fritz, S. R. Furlanetto, K. Gale-Sides, B. Glendenning, D. Gorthi, B. Greig, J. Grobbelaar, Z. Halday, B. J. Hazelton, S. Heimersheim, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, N. S. Kern, J. Kerrigan, P. Kittiwisit, S. A. Kohn, M. Kolopanis, A. Lanman, P. La Plante, T. Lekalake, D. Lewis, A. Liu, Y.-Z. Ma, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. E. Martinot, E. Matsetela, A. Mesinger, J. Mirocha, M. Molewa, M. F. Morales, T. Mosiane, J. B. Muñoz, S. G. Murray, A. R. Neben, B. Nikolic, C. D. Nunhokee, A. R. Parsons, N. Patra, S. Pieterse, J. C. Pober, Y. Qin, N. Razavi-Ghods, I. Reis, J. Ringuette, J. Robnett, K. Rosie, M. G. Santos, S. Sikder, P. Sims, C. Smith, A. Syce, N. Thyagarajan, P. K. G. Williams, and H. Zheng. HERA Phase I Limits on the Cosmic 21 cm Signal: Constraints on Astrophysics and Cosmology during the Epoch of Reionization. *ApJ*, 924(2):51, Jan. 2022a. doi: 10.3847/1538-4357/ac2ffc.

HERA Collaboration, Z. Abdurashidova, J. E. Aguirre, P. Alexander, Z. S. Ali, Y. Balfour, A. P. Beardsley, G. Bernardi, T. S. Billings, J. D. Bowman, R. F. Bradley, P. Bull, J. Burba, S. Carey, C. L. Carilli, C. Cheng, D. R. DeBoer, M. Dexter, E. de Lera Acedo, T. Dibblee-Barkman, J. S. Dillon, J. Ely, A. Ewall-Wice, N. Fagnoni, R. Fritz, S. R. Furlanetto, K. Gale-Sides, B. Glendenning, D. Gorthi, B. Greig, J. Grobbelaar, Z. Halday, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, N. S. Kern, J. Kerrigan, P. Kittiwisit, S. A. Kohn, M. Kolopanis, A. Lanman, P. La Plante, T. Lekalake, D. Lewis, A. Liu, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. E. Martinot, E. Matsetela, A. Mesinger, M. Molewa, M. F. Morales, T. Mosiane, S. G. Murray, A. R. Neben, B. Nikolic, C. D. Nunhokee, A. R. Parsons, N. Patra, R. Pascua, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Ringuette, J. Robnett, K. Rosie, P. Sims, S. Singh, C. Smith, A. Syce, N. Thyagarajan, P. K. G. Williams, H. Zheng, and HERA Collaboration. First Results from HERA Phase I: Upper Limits on the Epoch of Reionization 21 cm Power Spectrum. *ApJ*, 925(2): 221, Feb. 2022b. doi: 10.3847/1538-4357/ac1c78.

HERA Collaboration, Z. Abdurashidova, T. Adams, J. E. Aguirre, P. Alexander, Z. S. Ali, R. Baartman, Y. Balfour, R. Barkana, A. P. Beardsley, G. Bernardi, T. S. Billings, J. D. Bowman, R. F. Bradley, D. Breitman, P. Bull, J. Burba, S. Carey, C. L. Carilli, C. Cheng, S. Choudhuri, D. R. DeBoer, E. de Lera Acedo, M. Dexter, J. S. Dillon, J. Ely, A. Ewall-Wice, N. Fagnoni, A. Fialkov, R. Fritz, S. R. Furlanetto, K. Gale-Sides, H. Garsden, B. Glendenning, A. Gorce, D. Gorthi, B. Greig, J. Grobbelaar, Z. Halday, B. J. Hazelton, S. Heimersheim, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, N. S. Kern, J. Kerrigan, P. Kittiwisit, S. A. Kohn, M. Kolopanis, A. Lanman, P. La Plante, D. Lewis, A. Liu, A. Loots, Y.-Z. Ma, D. H. E. MacMahon, L. Malan, K. Malgas, C. Malgas, M. Maree, B. Marero, Z. E. Martinot, L. McBride, A. Mesinger, J. Mirocha, M. Molewa, M. F. Morales, T. Mosiane, J. B. Muñoz, S. G. Murray, V. Nagpal, A. R. Neben, B. Nikolic, C. D. Nunhokee, H. Nuwegeld, A. R. Parsons, R. Pascua, N. Patra, S. Pieterse, Y. Qin, N. Razavi-Ghods, J. Robnett, K. Rosie, M. G. Santos, P. Sims, S. Singh, C. Smith, H. Swarts, J. Tan, N. Thyagarajan, M. J. Wilensky, P. K. G. Williams, P. van Wyngaarden, and H. Zheng. Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations. *ApJ*, 945(2):124, Mar. 2023. doi: 10.3847/1538-4357/acaf50.

J. Ho, A. Jain, and P. Abbeel. Denoising Diffusion Probabilistic Models. *arXiv e-prints*, art. arXiv:2006.11239, June 2020. doi: 10.48550/arXiv.2006.11239.

A. Hoag, M. Bradač, K. Huang, C. Mason, T. Treu, K. B. Schmidt, M. Trenti, V. Strait, B. C. Lemaux, E. Q. Finney, and M. Paddock. Constraining the Neutral Fraction of Hydrogen in the IGM at Redshift 7.5. *ApJ*, 878(1):12, June 2019. doi: 10.3847/1538-4357/ab1de7.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 11 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

L. N. Holzbauer and S. R. Furlanetto. Fluctuations in the high-redshift lyman–werner and lya radiation backgrounds. *Monthly Notices of the Royal Astronomical Society*, 419(1):718–731, 12 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.19752.x. URL https://doi.org/10.1111/j.1365-2966.2011.19752.x.

K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(89)90020-8. URL https://www.sciencedirect.com/science/article/pii/0893608089900208.

L. Hui and N. Y. Gnedin. Equation of state of the photoionized intergalactic medium. *MNRAS*, 292(1):27–42, Nov. 1997. doi: 10.1093/mnras/292.1.27.

J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, 9 (3):90–95, May 2007. doi: 10.1109/MCSE.2007.55.

A. Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005. URL http://jmlr.org/papers/v6/hyvarinen05a.html.

I. T. Iliev, G. Mellema, U. L. Pen, H. Merz, P. R. Shapiro, and M. A. Alvarez. Simulating cosmic reionization at large scales - I. The geometry of reionization. *MNRAS*, 369(4):1625–1638, July 2006. doi: 10.1111/j.1365-2966.2006.10502.x.

A. K. Inoue, I. Iwata, and J.-M. Deharveng. The escape fraction of ionizing photons from galaxies at z = 0–6. *Monthly Notices of the Royal Astronomical Society: Letters*, 371(1):L1–L5, 09 2006. ISSN 1745-3925. doi: 10.1111/j.1745-3933.2006.00195.x. URL https://doi.org/10.1111/j.1745-3933.2006.00195.x.

W. D. Jennings, C. A. Watkinson, F. B. Abdalla, and J. D. McEwen. Evaluating machine learning techniques for predicting power spectra from reionization simulations. *MNRAS*, 483(3):2907–2922, Mar. 2019. doi: 10.1093/mnras/sty3168.

H. Jensen, K. K. Datta, G. Mellema, E. Chapman, F. B. Abdalla, I. T. Iliev, Y. Mao, M. G. Santos, P. R. Shapiro, S. Zaroubi, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, G. J. A. Harker, V. Jelić, S. Kazemi, L. V. E. Koopmans, P. Labropoulos, O. Martinez, A. R. Offringa, V. N. Pandey, J. Schaye, R. M. Thomas, V. Veligatla, H. Vedantham, and S. Yatawatta. Probing reionization with LOFAR using 21-cm redshift space distortions. *MNRAS*, 435(1):460–474, Oct. 2013. doi: 10.1093/mnras/stt1341.

R. Kannan, M. Vogelsberger, F. Marinacci, R. McKinnon, R. Pakmor, and V. Springel. AREPO-RT: radiation hydrodynamics on a moving mesh. *MNRAS*, 485(1):117–149, May 2019. doi: 10.1093/mnras/stz287.

R. Kannan, A. Smith, E. Garaldi, X. Shen, M. Vogelsberger, R. Pakmor, V. Springel, and L. Hernquist. The THESAN project: predictions for multitracer line intensity mapping in the epoch of reionization. *MNRAS*, 514(3):3857–3878, Aug. 2022. doi: 10.1093/mnras/stac1557.

H. D. Kaur, N. Gillet, and A. Mesinger. Minimum size of 21-cm simulations. *MNRAS*, 495(2): 2354–2362, June 2020. doi: 10.1093/mnras/staa1323.

B. Kawar, M. Elad, S. Ermon, and J. Song. Denoising Diffusion Restoration Models. *arXiv e-prints*, art. arXiv:2201.11793, Jan. 2022. doi: 10.48550/arXiv.2201.11793.

N. S. Kern, A. Liu, A. R. Parsons, A. Mesinger, and B. Greig. Emulating Simulations of Cosmic Dawn for 21 cm Power Spectrum Constraints on Cosmology, Reionization, and X-Ray Heating. *ApJ*, 848(1):23, Oct. 2017. doi: 10.3847/1538-4357/aa8bb4.

T. Kimm and R. Cen. Escape fraction of ionizing photons during reionization: Effects due to supernova feedback and runaway ob stars. *The Astrophysical Journal*, 788(2):121, may 2014. doi: 10.1088/0004-637X/788/2/121. URL https://doi.org/10.1088/0004-637X/788/2/121.

T. Kimm, H. Katz, M. Haehnelt, J. Rosdahl, J. Devriendt, and A. Slyz. Feedback-regulated star formation and escape of LyC photons from mini-haloes during reionization. *MNRAS*, 466(4): 4826–4846, Apr. 2017. doi: 10.1093/mnras/stx052.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

P. Kittiwisit, J. D. Bowman, S. G. Murray, B. K. Gehlot, D. C. Jacobs, and A. P. Beardsley. Measurements of one-point statistics in 21-cm intensity maps via foreground avoidance strategy. *MNRAS*, 517(2):2138–2150, Dec. 2022. doi: 10.1093/mnras/stac2826.

A. Konno, M. Ouchi, Y. Ono, K. Shimasaku, T. Shibuya, H. Furusawa, K. Nakajima, Y. Naito, R. Momose, S. Yuma, and M. Iye. Accelerated Evolution of the Lyα Luminosity Function at z $>\sim$7 Revealed by the Subaru Ultra-deep Survey for Lyα Emitters at z = 7.3. *ApJ*, 797(1):16, Dec. 2014. doi: 10.1088/0004-637X/797/1/16.

L. Koopmans, J. Pritchard, G. Mellema, J. Aguirre, K. Ahn, R. Barkana, I. van Bemmel, G. Bernardi, A. Bonaldi, F. Briggs, A. G. de Bruyn, T. C. Chang, E. Chapman, X. Chen, B. Ciardi, P. Dayal, A. Ferrara, A. Fialkov, F. Fiore, K. Ichiki, I. T. Illiev, S. Inoue, V. Jelic, M. Jones, J. Lazio, U. Maio, S. Majumdar, K. J. Mack, A. Mesinger, M. F. Morales, A. Parsons, U. L. Pen, M. Santos, R. Schneider, B. Semelin, R. S. de Souza, R. Subrahmanyan, T. Takeuchi, H. Vedantham, J. Wagg, R. Webster, S. Wyithe, K. K. Datta, and C. Trott. The Cosmic Dawn and Epoch of Reionisation with SKA. In *Advancing Astrophysics with the Square Kilometre Array (AAemu4)*, page 1, Apr. 2015. doi: 10.22323/1.215.0001.

M. Kuhlen and C.-A. Faucher-Giguère. Concordance models of reionization: implications for faint galaxies and escape fraction evolution. *MNRAS*, 423(1):862–876, June 2012. doi: 10.1111/j.1365-2966.2012.20924.x.

J. U. Lange. NAUTILUS: boosting Bayesian importance nested sampling with deep learning. *MNRAS*, 525(2):3181–3194, Oct. 2023. doi: 10.1093/mnras/stad2441.

H. Lazare, D. Sarkar, and E. D. Kovetz. HERA bound on X-ray luminosity weakens when accounting for Population III stars. *arXiv e-prints*, art. arXiv:2307.15577, July 2023. doi: 10.48550/arXiv.2307.15577.

J. S. W. Lewis, P. Ocvirk, D. Aubert, J. G. Sorce, P. R. Shapiro, N. Deparis, T. Dawoodbhoy, R. Teyssier, G. Yepes, S. Gottlöber, K. Ahn, I. T. Iliev, and J. Chardin. Galactic ionizing photon budget during the epoch of reionization in the Cosmic Dawn II simulation. *MNRAS*, 496(4): 4342–4357, Aug. 2020. doi: 10.1093/mnras/staa1748.

H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein. Visualizing the loss landscape of neural nets. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. doi: 10.48550/arXiv.1712.09913. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/a41b3bb3e6b050b6c9067c67f663b915-Paper.pdf.

A. Lidz. Physics of the Intergalactic Medium During the Epoch of Reionization. In A. Mesinger, editor, *Understanding the Epoch of Cosmic Reionization: Challenges and Progress*, volume 423 of *Astrophysics and Space Science Library*, page 23, Jan. 2016. doi: 10.1007/978-3-319-21957-8_2.

A. Lidz, O. Zahn, M. McQuinn, M. Zaldarriaga, S. Dutta, and L. Hernquist. Higher Order Contributions to the 21 cm Power Spectrum. *ApJ*, 659(2):865–876, Apr. 2007. doi: 10.1086/511670.

A. Liu and J. R. Shaw. Data Analysis for Precision 21 cm Cosmology. *PASP*, 132(1012):062001, June 2020. doi: 10.1088/1538-3873/ab5bfd.

A. Liu and M. Tegmark. A method for 21 cm power spectrum estimation in the presence of foregrounds. *Phys. Rev. D*, 83(10):103006, May 2011. doi: 10.1103/PhysRevD.83.103006.

A. Liu, A. R. Parsons, and C. M. Trott. Epoch of reionization window. I. Mathematical formalism. *Phys. Rev. D*, 90(2):023018, July 2014a. doi: 10.1103/PhysRevD.90.023018.

A. Liu, A. R. Parsons, and C. M. Trott. Epoch of reionization window. II. Statistical methods for foreground wedge reduction. *Phys. Rev. D*, 90(2):023019, July 2014b. doi: 10.1103/PhysRevD.90.023019.

T. Louis, A. La Posta, Z. Atkins, H. T. Jense, I. Abril-Cabezas, G. E. Addison, P. A. R. Ade, S. Aiola, T. Alford, D. Alonso, M. Amiri, R. An, J. E. Austermann, E. Barbavara, N. Battaglia, E. S. Battistelli, J. A. Beall, R. Bean, A. Beheshti, B. Beringue, T. Bhandarkar, E. Biermann, B. Bolliet, J. R. Bond, E. Calabrese, V. Capalbo, F. Carrero, S.-F. Chen, G. Chesmore, H.-m. Cho, S. K. Choi, S. E. Clark, N. F. Cothard, K. Coughlin, W. Coulton, D. Crichton, K. T. Crowley, O. Darwish, M. J. Devlin, S. Dicker, C. J. Duell, S. M. Duff, A. J. Duivenvoorden, J. Dunkley, R. Dunner, C. Embil Villagra, M. Fankhanel, G. S. Farren, S. Ferraro, A. Foster, R. Freundt, B. Fuzia, P. A. Gallardo, X. Garrido, M. Gerbino, S. Giardiello, A. Gill, J. Givans, V. Gluscevic, S. Goldstein, J. E. Golec, Y. Gong, Y. Guan, M. Halpern, I. Harrison, M. Hasselfield, E. Healy, S. Henderson, B. Hensley, C. Hervías-Caimapo, J. C. Hill, G. C. Hilton, M. Hilton, A. D. Hincks, R. Hložek, S.-P. P. Ho, J. Hood, E. Hornecker, Z. B. Huber, J. Hubmayr, K. M. Huffenberger, J. P. Hughes, M. Ikape, K. Irwin, G. Isopi, N. Joshi, B. Keller, J. Kim, K. Knowles, B. J. Koopman, A. Kosowsky, D. Kramer, A. Kusiak, A. Lague, V. Lakey, E. Lee, Y. Li, Z. Li, M. Limon, M. Lokken, M. Lungu, N. MacCrann, A. MacInnis, M. S. Madhavacheril, D. Maldonado, F. Maldonado, M. Mallaby-Kay, G. A. Marques, J. van Marrewijk, F. McCarthy, J. McMahon, Y. Mehta, F. Menanteau, K. Moodley, T. W. Morris, T. Mroczkowski, S. Naess, T. Namikawa, F. Nati, S. K. Nerval, L. Newburgh, A. Nicola, M. D. Niemack, M. R. Nolta, J. Orlowski-Scherer, L. Pagano, L. A. Page, S. Pandey, B. Partridge, K. Perez Sarmiento, H. Prince, R. Puddu, F. J. Qu, D. C. Ragavan, B. Ried Guachalla, K. K. Rogers, F. Rojas, T. Sakuma, E. Schaan, B. L. Schmitt, N. Sehgal, S. Shaikh, B. D. Sherwin, C. Sierra, J. Sievers, C. Sifón, S. Simon, R. Sonka, D. N. Spergel, S. T. Staggs, E. Storer, K. Surrao, E. R. Switzer, N. Tampier, R. Thornton, H. Trac, C. Tucker, J. Ullom, L. R. Vale, A. Van Engelen, J. Van Lanen, C. Vargas, E. M. Vavagiakis, K. Wagoner, Y. Wang, L. Wenzl, E. J. Wollack, and K. Zheng. The Atacama Cosmology Telescope: DR6 Power Spectra, Likelihoods and ΛCDM Parameters. *arXiv e-prints*, art. arXiv:2503.14452, Mar. 2025. doi: 10.48550/arXiv.2503.14452.

X. Ma, E. Quataert, A. Wetzel, P. F. Hopkins, C.-A. Faucher-Giguère, and D. Kereš. No missing photons for reionization: moderate ionizing photon escape fractions from the FIRE-2 simulations. *MNRAS*, 498(2):2001–2017, Oct. 2020. doi: 10.1093/mnras/staa2404.

P. Madau and M. Dickinson. Cosmic Star-Formation History. *ARA&A*, 52:415–486, Aug. 2014. doi: 10.1146/annurev-astro-081811-125615.

P. Madau, A. Meiksin, and M. J. Rees. 21 Centimeter Tomography of the Intergalactic Medium at High Redshift. *ApJ*, 475(2):429–444, Feb. 1997. doi: 10.1086/303549.

B. Maity and T. R. Choudhury. Constraining the reionization and thermal history of the Universe using a seminumerical photon-conserving code SCRIPT. *MNRAS*, 515(1):617–630, Sept. 2022. doi: 10.1093/mnras/stac1847.

F. Mannucci, G. Cresci, R. Maiolino, A. Marconi, and A. Gnerucci. A fundamental relation between mass, star formation rate and metallicity in local and high-redshift galaxies. *Monthly Notices of the Royal Astronomical Society*, 408(4):2115–2127, 10 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.17291.x. URL https://doi.org/10.1111/j.1365-2966.2010.17291.x.

Y. Mao, P. R. Shapiro, G. Mellema, I. T. Iliev, J. Koda, and K. Ahn. Redshift-space distortion of the 21-cm background from the epoch of reionization - I. Methodology re-examined. *MNRAS*, 422 (2):926–954, May 2012. doi: 10.1111/j.1365-2966.2012.20471.x.

C. A. Mason, T. Treu, M. Dijkstra, A. Mesinger, M. Trenti, L. Pentericci, S. de Barros, and E. Vanzella. The Universe Is Reionizing at z $\sim$ 7: Bayesian Inference of the IGM Neutral Fraction Using Lyα Emission from Galaxies. *ApJ*, 856(1):2, Mar. 2018. doi: 10.3847/1538-4357/aaboa7.

I. D. McGreer, A. Mesinger, and V. D'Odorico. Model-independent evidence in favour of an end to reionization by z $\approx$ 6. *MNRAS*, 447(1):499–505, Feb. 2015. doi: 10.1093/mnras/stu2449.

M. McQuinn. The Evolution of the Intergalactic Medium. *ARA&A*, 54:313–362, Sept. 2016. doi: 10.1146/annurev-astro-082214-122355.

G. Mellema, L. V. E. Koopmans, F. A. Abdalla, G. Bernardi, B. Ciardi, S. Daiboo, A. G. de Bruyn, K. K. Datta, H. Falcke, A. Ferrara, I. T. Iliev, F. Iocco, V. Jelić, H. Jensen, R. Joseph, P. Labroupoulos, A. Meiksin, A. Mesinger, A. R. Offringa, V. N. Pandey, J. R. Pritchard, M. G. Santos, D. J. Schwarz, B. Semelin, H. Vedantham, S. Yatawatta, and S. Zaroubi. Reionization and the Cosmic Dawn with the Square Kilometre Array. *Experimental Astronomy*, 36(1-2):235–318, Aug. 2013. doi: 10.1007/s10686-013-9334-5.

R. Meriot, B. Semelin, and D. Cornu. Comparison of Bayesian inference methods using the Loreli II database of hydro-radiative simulations of the 21-cm signal. *arXiv e-prints*, art. arXiv:2411.03093, Nov. 2024. doi: 10.48550/arXiv.2411.03093.

F. G. Mertens, M. Mevius, L. V. E. Koopmans, A. R. Offringa, G. Mellema, S. Zaroubi, M. A. Brentjens, H. Gan, B. K. Gehlot, V. N. Pandey, A. M. Sardarabadi, H. K. Vedantham, S. Yatawatta, K. M. B. Asad, B. Ciardi, E. Chapman, S. Gazagnes, R. Ghara, A. Ghosh, S. K. Giri, I. T. Iliev, V. Jelić, R. Kooistra, R. Mondal, J. Schaye, and M. B. Silva. Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at z $\approx$ 9.1 from LOFAR. *MNRAS*, 493(2):1662–1685, Apr. 2020. doi: 10.1093/mnras/staa327.

F. G. Mertens, M. Mevius, L. V. E. Koopmans, A. R. Offringa, S. Zaroubi, A. Acharya, S. A. Brackenhoff, E. Ceccotti, E. Chapman, K. Chege, B. Ciardi, R. Ghara, S. Ghosh, S. K. Giri, I. Hothi, C. Höfer, I. T. Iliev, V. Jelić, Q. Ma, G. Mellema, S. Munshi, V. N. Pandey, and S. Yatawatta. Deeper multi-redshift upper limits on the Epoch of Reionization 21-cm signal power spectrum from LOFAR between z=8.3 and z=10.1. *arXiv e-prints*, art. arXiv:2503.05576, Mar. 2025. doi: 10.48550/arXiv.2503.05576.

A. Mesinger. *The Cosmic 21-cm Revolution; Charting the first billion years of our universe*. 2019. doi: 10.1088/2514-3433/ab4a73.

A. Mesinger and S. Furlanetto. Efficient Simulations of Early Structure Formation and Reionization. *ApJ*, 669(2):663–675, Nov. 2007. doi: 10.1086/521806.

A. Mesinger, S. Furlanetto, and R. Cen. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *MNRAS*, 411(2):955–972, Feb. 2011. doi: 10.1111/j.1365-2966.2010.17731.x.

A. Mesinger, A. Aykutalp, E. Vanzella, L. Pentericci, A. Ferrara, and M. Dijkstra. Can the intergalactic medium cause a rapid drop in Lyα emission at z > 6? *MNRAS*, 446(1):566–577, Jan. 2015. doi: 10.1093/mnras/stu2089.

A. Mesinger, B. Greig, and E. Sobacchi. The Evolution Of 21 cm Structure (EOS): public, large-scale simulations of Cosmic Dawn and reionization. *MNRAS*, 459(3):2342–2353, July 2016. doi: 10.1093/mnras/stw831.

J. Mirocha, S. R. Furlanetto, and G. Sun. The global 21-cm signal in the context of the high- z galaxy luminosity function. *MNRAS*, 464(2):1365–1379, Jan. 2017. doi: 10.1093/mnras/stw2412.

S. Mitra, T. R. Choudhury, and A. Ferrara. Cosmic reionization after Planck. *MNRAS*, 454:L76–L80, Nov. 2015. doi: 10.1093/mnrasl/slv134.

H. Mo, F. C. van den Bosch, and S. White. *Galaxy Formation and Evolution*. 2010. doi: 10.1017/CBO9780511807244.

R. Mondal, S. Bharadwaj, S. Majumdar, A. Bera, and A. Acharyya. The effect of non-Gaussianity on error predictions for the Epoch of Reionization (EoR) 21-cm power spectrum. *MNRAS*, 449: L41–L45, Apr. 2015. doi: 10.1093/mnrasl/slv015.

R. Mondal, A. Fialkov, C. Fling, I. T. Iliev, R. Barkana, B. Ciardi, G. Mellema, S. Zaroubi, L. V. E. Koopmans, F. G. Mertens, B. K. Gehlot, R. Ghara, A. Ghosh, S. K. Giri, A. Offringa, and V. N. Pandey. Tight constraints on the excess radio background at z = 9.1 from LOFAR. *MNRAS*, 498 (3):4178–4191, Nov. 2020. doi: 10.1093/mnras/staa2422.

R. Mondal, G. Mellema, A. K. Shaw, M. Kamran, and S. Majumdar. The Epoch of Reionization 21-cm bispectrum: the impact of light-cone effects and detectability. *MNRAS*, 508(3):3848–3859, Dec. 2021. doi: 10.1093/mnras/stab2900.

R. Mondal, G. Mellema, S. G. Murray, and B. Greig. The multifrequency angular power spectrum in parameter studies of the cosmic 21-cm signal. *MNRAS*, 514(1):L31–L35, July 2022. doi: 10.1093/mnrasl/slac053.

D. J. Mortlock, S. J. Warren, B. P. Venemans, M. Patel, P. C. Hewett, R. G. McMahon, C. Simpson, T. Theuns, E. A. Gonzáles-Solares, A. Adamson, S. Dye, N. C. Hambly, P. Hirst, M. J. Irwin, E. Kuiper, A. Lawrence, and H. J. A. Röttgering. A luminous quasar at a redshift of z = 7.085. *Nature*, 474(7353):616–619, June 2011. doi: 10.1038/nature10159.

J. B. Muñoz, Y. Qin, A. Mesinger, S. G. Murray, B. Greig, and C. Mason. The impact of the first galaxies on cosmic dawn and reionization. *MNRAS*, 511(3):3657–3681, Apr. 2022. doi: 10.1093/mnras/stac185.

F. Munshi, A. M. Brooks, E. Applebaum, D. R. Weisz, F. Governato, and T. R. Quinn. Going, going, gone dark: Quantifying the scatter in the faintest dwarf galaxies. *arXiv e-prints*, art. arXiv:1705.06286, May 2017. doi: 10.48550/arXiv.1705.06286.

F. Munshi, A. M. Brooks, E. Applebaum, C. R. Christensen, T. Quinn, and S. Sligh. Quantifying Scatter in Galaxy Formation at the Lowest Masses. *ApJ*, 923(1):35, Dec. 2021. doi: 10.3847/1538-4357/acodb6.

S. Munshi, F. G. Mertens, L. V. E. Koopmans, A. R. Offringa, B. Semelin, D. Aubert, R. Barkana, A. Bracco, S. A. Brackenhoff, B. Cecconi, E. Ceccotti, S. Corbel, A. Fialkov, B. K. Gehlot, R. Ghara, J. N. Girard, J. M. Grießmeier, C. Höfer, I. Hothi, R. Mériot, M. Mevius, P. Ocvirk, A. K. Shaw, G. Theureau, S. Yatawatta, P. Zarka, and S. Zaroubi. First upper limits on the 21 cm signal power spectrum from cosmic dawn from one night of observations with NenuFAR. *Astronomy & Astrophysics*, 681:A62, Jan. 2024. ISSN 0004-6361, 1432-0746. URL http://arxiv.org/abs/2311.05364.

S. Murray, B. Greig, A. Mesinger, J. Muñoz, Y. Qin, J. Park, and C. Watkinson. 21cmFAST v3: A Python-integrated C code for generating 3D realizations of the cosmic 21cm signal. *The Journal of Open Source Software*, 5(54):2582, Oct. 2020. doi: 10.21105/joss.02582.

S. Murray, J. Pober, and M. Kolopanis. 21cmSense v2: A modular, open-source 21 cm sensitivity calculator. *The Journal of Open Source Software*, 9(97):6501, May 2024. doi: 10.21105/joss.06501.

S. G. Murray. *Next-Generation Tools For Next-Generation Surveys*. PhD thesis, University of Western Australia, Apr. 2017.

S. G. Murray. powerbox: A python package for creating structured fields with isotropic power spectra. *Journal of Open Source Software*, 3(28):850, 2018. doi: 10.21105/joss.00850. URL https://doi.org/10.21105/joss.00850.

S. G. Murray, J. D. Bowman, P. H. Sims, N. Mahesh, A. E. E. Rogers, R. A. Monsalve, T. Samson, and A. K. Vydula. A Bayesian calibration framework for EDGES. *MNRAS*, 517(2):2264–2284, Dec. 2022. doi: 10.1093/mnras/stac2600.

S. J. Mutch, P. M. Geil, G. B. Poole, P. W. Angel, A. R. Duffy, A. Mesinger, and J. S. B. Wyithe. Dark-ages reionization and galaxy formation simulation - III. Modelling galaxy formation and the epoch of reionization. *MNRAS*, 462(1):250–276, Oct. 2016. doi: 10.1093/mnras/stw1506.

I. Nikolić, A. Mesinger, Y. Qin, and A. Gorce. Inferring reionization and galaxy properties from the patchy kinetic Sunyaev-Zel'dovich signal. *arXiv e-prints*, art. arXiv:2307.01265, July 2023. doi: 10.48550/arXiv.2307.01265.

C. D. Nunhokee, D. Null, C. M. Trott, N. Barry, Y. Qin, R. B. Wayth, J. L. B. Line, C. H. Jordan, B. Pindor, J. H. Cook, J. Bowman, A. Chokshi, J. Ducharme, K. Elder, Q. Guo, B. Hazelton, W. Hidayat, T. Ito, D. Jacobs, E. Jong, M. Kolopanis, T. Kunicki, E. Lilleskov, M. F. Morales, J. C. Pober, A. Selvaraj, R. Shi, K. Takahashi, S. J. Tingay, R. L. Webster, S. Yoshiura, and Q. Zheng. Limits on the 21 cm power spectrum at z=6.5-7.0 from MWA observations. *arXiv e-prints*, art. arXiv:2505.09097, May 2025. doi: 10.48550/arXiv.2505.09097.

P. Ocvirk, D. Aubert, J. G. Sorce, P. R. Shapiro, N. Deparis, T. Dawoodbhoy, J. Lewis, R. Teyssier, G. Yepes, S. Gottlöber, K. Ahn, I. T. Iliev, and Y. Hoffman. Cosmic dawn II (CoDa II): a new radiation-hydrodynamics simulation of the self-consistent coupling of galaxy formation and reionization. *Monthly Notices of the Royal Astronomical Society*, 496(4):4087–4107, may 2020. doi: 10.1093/mnras/staa1266. URL https://doi.org/10.1093%2Fmnras%2Fstaa1266.

P. A. Oesch, R. J. Bouwens, G. D. Illingworth, I. Labbé, and M. Stefanon. The Dearth of z $\sim$ 10 Galaxies in All HST Legacy Fields—The Rapid Evolution of the Galaxy Population in the First 500 Myr. *ApJ*, 855(2):105, Mar. 2018. doi: 10.3847/1538-4357/aab03f.

T. Okamoto, L. Gao, and T. Theuns. Mass loss of galaxies due to an ultraviolet background. *MNRAS*, 390(3):920–928, Nov. 2008. doi: 10.1111/j.1365-2966.2008.13830.x.

J. B. Oke and J. E. Gunn. Secondary standard stars for absolute spectrophotometry. *ApJ*, 266: 713–717, Mar. 1983. doi: 10.1086/160817.

M. Ouchi, K. Shimasaku, H. Furusawa, T. Saito, M. Yoshida, M. Akiyama, Y. Ono, T. Yamada, K. Ota, N. Kashikawa, M. Iye, T. Kodama, S. Okamura, C. Simpson, and M. Yoshida. Statistics of 207 Lyα Emitters at a Redshift Near 7: Constraints on Reionization and Galaxy Formation Models. *ApJ*, 723(1):869–894, Nov. 2010. doi: 10.1088/0004-637X/723/1/869.

J.-P. Paardekooper, S. Khochfar, and C. Dalla Vecchia. The First Billion Years project: the escape fraction of ionizing photons in the epoch of reionization. *MNRAS*, 451(3):2544–2563, Aug. 2015. doi: 10.1093/mnras/stv1114.

F. Pacucci, A. Mesinger, S. Mineo, and A. Ferrara. The X-ray spectra of the first galaxies: 21 cm signatures. *MNRAS*, 443(1):678–686, Sept. 2014. doi: 10.1093/mnras/stu1240.

G. Papamakarios, D. C. Sterratt, and I. Murray. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. *arXiv e-prints*, art. arXiv:1805.07226, May 2018. doi: 10.48550/arXiv.1805.07226.

J. Park, A. Mesinger, B. Greig, and N. Gillet. Inferring the astrophysics of reionization and cosmic dawn from galaxy luminosity functions and the 21-cm signal. *MNRAS*, 484(1):933–949, Mar. 2019. doi: 10.1093/mnras/stz032.

A. R. Parsons, J. C. Pober, J. E. Aguirre, C. L. Carilli, D. C. Jacobs, and D. F. Moore. A Per-baseline, Delay-spectrum Technique for Accessing the 21 cm Cosmic Reionization Signature. *ApJ*, 756 (2):165, Sept. 2012. doi: 10.1088/0004-637X/756/2/165.

R. Pascua, Z. E. Martinot, A. Liu, J. E. Aguirre, N. S. Kern, J. S. Dillon, M. J. Wilensky, N. Fagnoni, E. de Lera Acedo, and D. R. DeBoer. A Generalized Method for Characterizing 21 cm Power Spectrum Signal Loss from Temporal Filtering of Drift-scanning Visibilities. *ApJ*, 985(1):127, May 2025. doi: 10.3847/1538-4357/adc37d.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv e-prints*, art. arXiv:1912.01703, Dec. 2019. doi: 10.48550/arXiv.1912.01703.

A. H. Pawlik, J. Schaye, and E. Van Scherpenzeel. Keeping the universe ionized: photoheating and the clumping factor of the high-redshift intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 394(4):1812–1824, 04 2009. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.14486.x. URL https://doi.org/10.1111/j.1365-2966.2009.14486.x.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard,

M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, Sept. 2020. doi: 10.1051/0004-6361/201833910.

J. C. Pober. The impact of foregrounds on redshift space distortion measurements with the highly redshifted 21-cm line. *MNRAS*, 447(2):1705–1712, Feb. 2015. doi: 10.1093/mnras/stu2575.

J. C. Pober, A. R. Parsons, D. R. DeBoer, P. McDonald, M. McQuinn, J. E. Aguirre, Z. Ali, R. F. Bradley, T.-C. Chang, and M. F. Morales. The Baryon Acoustic Oscillation Broadband and Broad-beam Array: Design Overview and Sensitivity Forecasts. *AJ*, 145(3):65, Mar. 2013. doi: 10.1088/0004-6256/145/3/65.

J. C. Pober, A. Liu, J. S. Dillon, J. E. Aguirre, J. D. Bowman, R. F. Bradley, C. L. Carilli, D. R. DeBoer, J. N. Hewitt, D. C. Jacobs, M. McQuinn, M. F. Morales, A. R. Parsons, M. Tegmark, and D. J. Werthimer. What Next-generation 21 cm Power Spectrum Measurements can Teach us About the Epoch of Reionization. *ApJ*, 782(2):66, Feb. 2014. doi: 10.1088/0004-637X/782/2/66.

A. Pontzen, A. Slosar, N. Roth, and H. V. Peiris. Inverted initial conditions: Exploring the growth of cosmic structure and voids. *Phys. Rev. D*, 93(10):103519, May 2016. doi: 10.1103/PhysRevD.93.103519.

D. Prelogović and A. Mesinger. Exploring the likelihood of the 21-cm power spectrum with simulation-based inference. *MNRAS*, 524(3):4239–4255, Sept. 2023. doi: 10.1093/mnras/stad2027.

D. Prelogović and A. Mesinger. How informative are summaries of the cosmic 21 cm signal? *A&A*, 688:A199, Aug. 2024. doi: 10.1051/0004-6361/202449309.

D. Prelogović, A. Mesinger, S. Murray, G. Fiameni, and N. Gillet. Machine learning astrophysics from 21 cm lightcones: impact of network architectures and signal contamination. *MNRAS*, 509(3):3852–3867, Jan. 2022. doi: 10.1093/mnras/stab3215.

J. R. Pritchard and S. R. Furlanetto. 21-cm fluctuations from inhomogeneous X-ray heating before reionization. *MNRAS*, 376(4):1680–1694, Apr. 2007. doi: 10.1111/j.1365-2966.2007.11519.x.

J. R. Pritchard and A. Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, Aug. 2012. doi: 10.1088/0034-4885/75/8/086901.

Y. Qin, V. Poulin, A. Mesinger, B. Greig, S. Murray, and J. Park. Reionization inference from the CMB optical depth and E-mode polarization power spectra. *MNRAS*, 499(1):550–558, Nov. 2020. doi: 10.1093/mnras/staa2797.

Y. Qin, A. Mesinger, S. E. I. Bosman, and M. Viel. Reionization and galaxy inference from the high-redshift Ly α forest. *MNRAS*, 506(2):2390–2407, Sept. 2021. doi: 10.1093/mnras/stab1833.

Y. Qin, A. Mesinger, D. Prelogović, G. Becker, M. Bischetti, S. Bosman, F. Davies, V. D'Odorico, P. Gaikwad, M. Haehnelt, L. Keating, S. Lai, E. Ryan-Weber, S. Satyavolu, F. Walter, and Y. Zhu. Percent-level timing of reionisation: Self-consistent, implicit-likelihood inference from XQR-30+ Lyα forest data. *PASA*, 42:e049, Apr. 2025. doi: 10.1017/pasa.2025.35.

G. Rácz, A. Kiessling, I. Csabai, and I. Szapudi. Complementary cosmological simulations. *A&A*, 672:A59, Apr. 2023. doi: 10.1051/0004-6361/202245211.

E. Rath, R. Pascua, A. T. Josaitis, A. Ewall-Wice, N. Fagnoni, E. de Lera Acedo, Z. E. Martinot, Z. Abdurashidova, T. Adams, J. E. Aguirre, R. Baartman, A. P. Beardsley, L. M. Berkhout, G. Bernardi, T. S. Billings, J. D. Bowman, P. Bull, J. Burba, R. Byrne, S. Carey, K. F. Chen, S. Choudhuri, T. Cox, D. R. DeBoer, M. Dexter, J. S. Dillon, S. Dynes, N. Eksteen, J. Ely, R. Fritz, S. R. Furlanetto, K. Gale-Sides, H. Garsden, B. K. Gehlot, A. Ghosh, A. Gorce, D. Gorthi, Z. Halday, B. J. Hazelton, J. N. Hewitt, J. Hickish, T. Huang, D. C. Jacobs, N. S. Kern, J. Kerrigan, P. Kittiwisit, M. Kolopanis, A. Lanman, A. Liu, Y. Z. Ma, D. H. E. MacMahon, L. Malan, C. Malgas, K. Malgas, B. Marero, L. McBride, A. Mesinger, N. Mohamed-Hinds, M. Molewa, M. F. Morales, S. G. Murray, B. Nikolic, H. Nuwegeld, A. R. Parsons, N. Patra, P. L. Plante, Y. Qin, N. Razavi-Ghods, D. Riley, J. Robnett, K. Rosie, M. G. Santos, P. Sims, S. Singh, D. Storer, H. Swarts, J. Tan, M. J. Wilensky, P. K. G. Williams, P. van Wyngaarden, and H. Zheng. Investigating mutual coupling in the hydrogen epoch of reionization array and mitigating its effects on the 21-cm power spectrum. *MNRAS*, 541(2):1125–1144, Aug. 2025. doi: 10.1093/mnras/staf1012.

C. L. Reichardt, S. Patil, P. A. R. Ade, A. J. Anderson, J. E. Austermann, J. S. Avva, E. Baxter, J. A. Beall, A. N. Bender, B. A. Benson, F. Bianchini, L. E. Bleem, J. E. Carlstrom, C. L. Chang, P. Chaubal, H. C. Chiang, T. L. Chou, R. Citron, C. C. Moran, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, W. Everett, J. Gallicchio, E. M. George, A. Gilbert, N. Gupta, N. W. Halverson, N. Harrington, J. W. Henning, G. C. Hilton, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, N. Huang, J. Hubmayr, K. D. Irwin, L. Knox, A. T. Lee, D. Li, A. Lowitz, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, M. Millea, L. M. Mocanu, J. J. Mohr, J. Montgomery, A. Nadolski, T. Natoli, J. P. Nibarger, G. Noble, V. Novosad, Y. Omori, S. Padin, C. Pryke, J. E. Ruhl, B. R. Saliwanchik, J. T. Sayre, K. K. Schaffer, E. Shirokoff, C. Sievers, G. Smecher, H. G. Spieler, Z. Staniszewski, A. A. Stark, C. Tucker, K. Vanderlinde, T. Veach, J. D. Vieira, G. Wang, N. Whitehorn, R. Williamson, W. L. K. Wu, and V. Yefremenko. An Improved Measurement of the Secondary Cosmic Microwave Background Anisotropies from the SPT-SZ + SPTpol Surveys. *ApJ*, 908(2):199, Feb. 2021. doi: 10.3847/1538-4357/abd407.

O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints*, art. arXiv:1505.04597, May 2015. doi: 10.48550/arXiv.1505.04597.

H. E. Ross, S. K. Giri, G. Mellema, K. L. Dixon, R. Ghara, and I. T. Iliev. Redshift-space distortions in simulations of the 21-cm signal from the cosmic dawn. *MNRAS*, 506(3):3717–3733, Sept. 2021. doi: 10.1093/mnras/stab1822.

M. G. Santos, L. Ferramacho, M. B. Silva, A. Amblard, and A. Cooray. Fast large volume simulations of the 21-cm signal from the reionization and pre-reionization epochs. *MNRAS*, 406(4): 2421–2432, Aug. 2010. doi: 10.1111/j.1365-2966.2010.16898.x.

A. Saxena, A. Cole, S. Gazagnes, P. D. Meerburg, C. Weniger, and S. J. Witte. Constraining the X-ray heating and reionization using 21-cm power spectra with Marginal Neural Ratio Estimation. *arXiv e-prints*, art. arXiv:2303.07339, Mar. 2023. doi: 10.48550/arXiv.2303.07339.

D. Schaerer. On the properties of massive Population III stars and metal-free stellar populations. *A&A*, 382:28–42, Jan. 2002. doi: 10.1051/0004-6361:20011619.

A. T. P. Schauer, S. C. O. Glover, R. S. Klessen, and P. Clark. The influence of streaming velocities and lyman–werner radiation on the formation of the first stars. *Monthly Notices of the Royal Astronomical Society*, 507(2):1775–1787, 07 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab1953. URL https://doi.org/10.1093/mnras/stab1953.

C. J. Schmit and J. R. Pritchard. Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks. *MNRAS*, 475(1):1213–1223, Mar. 2018. doi: 10.1093/mnras/stx3292.

A. Schneider, S. K. Giri, S. Amodeo, and A. Refregier. Constraining baryonic feedback and cosmology with weak-lensing, X-ray, and kinematic Sunyaev-Zeldovich observations. *MNRAS*, 514(3):3802–3814, Aug. 2022. doi: 10.1093/mnras/stac1493.

B. Schosser, C. Heneka, and T. Plehn. Optimal, fast, and robust inference of reionization-era cosmology with the 21cmPIE-INN. *SciPost Physics Core*, 8(2):037, Apr. 2025. doi: 10.21468/SciPostPhysCore.8.2.037.

R. Scoccimarro. Transients from initial conditions: a perturbative analysis. *MNRAS*, 299(4):1097–1118, Oct. 1998. doi: 10.1046/j.1365-8711.1998.01845.x.

A. E. Shapley, C. C. Steidel, M. Pettini, D. K. Erb, and K. L. Adelberger. The Direct Detection of Lyman-Continuum Emission from Star-forming Galaxies at z~3. *Astrophys. J.*, 651:688–703, 2006. doi: 10.1086/507511.

A. K. Shaw, S. Bharadwaj, and R. Mondal. The impact of non-Gaussianity on the error covariance for observations of the Epoch of Reionization 21-cm power spectrum. *MNRAS*, 487(4):4951–4964, Aug. 2019. doi: 10.1093/mnras/stz1561.

T. Shibuya, M. Ouchi, Y. Harikane, and K. Nakajima. Morphologies of ∼190,000 Galaxies at z = 0-10 Revealed with HST Legacy Data. III. Continuum Profile and Size Evolution of Lyα Emitters. *ApJ*, 871(2):164, Feb. 2019. doi: 10.3847/1538-4357/aaf64b.

H. Shimabukuro and B. Semelin. Analysing the 21 cm signal from the epoch of reionization with artificial neural networks. *MNRAS*, 468(4):3869–3877, July 2017. doi: 10.1093/mnras/stx734.

S. Singh, N. T. Jishnu, R. Subrahmanyan, N. Udaya Shankar, B. S. Girish, A. Raghunathan, R. Somashekar, K. S. Srivani, and M. Sathyanarayana Rao. On the detection of a cosmic dawn signal in the radio background. *Nature Astronomy*, 6:607–617, Feb. 2022. doi: 10.1038/s41550-022-01610-5.

E. Sobacchi and A. Mesinger. The depletion of gas in high-redshift dwarf galaxies from an inhomogeneous reionization. *MNRAS*, 432:L51–L55, May 2013. doi: 10.1093/mnrasl/slt035.

J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv e-prints*, art. arXiv:1503.03585, Mar. 2015. doi: 10.48550/arXiv.1503.03585.

Y. Song and S. Ermon. Generative Modeling by Estimating Gradients of the Data Distribution. *arXiv e-prints*, art. arXiv:1907.05600, July 2019. doi: 10.48550/arXiv.1907.05600.

Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution, 2020.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv e-prints*, art. arXiv:2011.13456, Nov. 2020. doi: 10.48550/arXiv.2011.13456.

V. Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *MNRAS*, 401(2):791–851, Jan. 2010. doi: 10.1111/j.1365-2966.2009.15715.x.

V. Springel and L. Hernquist. The history of star formation in a Λ cold dark matter universe. *MNRAS*, 339(2):312–334, Feb. 2003. doi: 10.1046/j.1365-8711.2003.06207.x.

C. C. Steidel, M. Pettini, and K. L. Adelberger. Lyman-continuum emission from galaxies at z ~ 3.4 *. *The Astrophysical Journal*, 546(2):665, jan 2001. doi: 10.1086/318323. URL https://doi.org/10.1086/318323.

G. Sun and S. R. Furlanetto. Constraints on the star formation efficiency of galaxies during the epoch of reionization. *MNRAS*, 460(1):417–433, July 2016. doi: 10.1093/mnras/stw980.

M. Tegmark, J. Silk, M. J. Rees, A. Blanchard, T. Abel, and F. Palla. How Small Were the First Cosmological Objects? *ApJ*, 474:1, Jan. 1997. doi: 10.1086/303434.

R. M. Thomas, S. Zaroubi, B. Ciardi, A. H. Pawlik, P. Labropoulos, V. Jelić, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, G. J. A. Harker, L. V. E. Koopmans, G. Mellema, V. N. Pandey, J. Schaye, and S. Yatawatta. Fast large-scale reionization simulations. *MNRAS*, 393(1):32–48, Feb. 2009. doi: 10.1111/j.1365-2966.2008.14206.x.

S. J. Tingay, R. Goeke, J. D. Bowman, D. Emrich, S. M. Ord, D. A. Mitchell, M. F. Morales, T. Booler, B. Crosse, R. B. Wayth, C. J. Lonsdale, S. Tremblay, D. Pallot, T. Colegate, A. Wicenec, N. Kudryavtseva, W. Arcus, D. Barnes, G. Bernardi, F. Briggs, S. Burns, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. Desouza, B. M. Gaensler, L. J. Greenhill, P. J. Hall, B. J. Hazelton, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, M. J. Lynch, B. Mckinley, S. R. Mcwhirter, E. Morgan, D. Oberoi, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya-Shankar, F. Schlagenhaufer, K. S. Srivani, J. Stevens, R. Subrahmanyan, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. *PASA*, 30: e007, Jan. 2013. doi: 10.1017/pasa.2012.007.

H. Tiwari, A. K. Shaw, S. Majumdar, M. Kamran, and M. Choudhury. Improving constraints on the reionization parameters using 21-cm bispectrum. *J. Cosmology Astropart. Phys.*, 2022(4): 045, Apr. 2022. doi: 10.1088/1475-7516/2022/04/045.

H. Trac, N. Chen, I. Holst, M. A. Alvarez, and R. Cen. AMBER: A Semi-numerical Abundance Matching Box for the Epoch of Reionization. *ApJ*, 927(2):186, Mar. 2022. doi: 10.3847/1538-4357/ac5116.

C. M. Trott. Exploring the evolution of reionization using a wavelet transform and the light cone effect. *MNRAS*, 461(1):126–135, Sept. 2016. doi: 10.1093/mnras/stw1310.

C. M. Trott, C. H. Jordan, S. Midgley, N. Barry, B. Greig, B. Pindor, J. H. Cook, G. Sleap, S. J. Tingay, D. Ung, P. Hancock, A. Williams, J. Bowman, R. Byrne, A. Chokshi, B. J. Hazelton, K. Hasegawa, D. Jacobs, R. C. Joseph, W. Li, J. L. B. Line, C. Lynch, B. McKinley, D. A. Mitchell, M. F. Morales, M. Ouchi, J. C. Pober, M. Rahimi, K. Takahashi, R. B. Wayth, R. L. Webster, M. Wilensky, J. S. B. Wyithe, S. Yoshiura, Z. Zhang, and Q. Zheng. Deep multiredshift limits on Epoch of Reionization 21 cm power spectra from four seasons of Murchison Widefield Array observations. *MNRAS*, 493(4):4711–4727, Apr. 2020. doi: 10.1093/mnras/staa414.

J. Tumlinson and J. M. Shull. Zero-metallicity stars and the effects of the first stars on reionization. *The Astrophysical Journal*, 528(2):L65, dec 1999. doi: 10.1086/312432. URL https://doi.org/10.1086/312432.

M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows, M. Brentjens, A. Nelles, R. Beck, H. Falcke, R. Fender, J. Hörandel, L. V. E. Koopmans, G. Mann, G. Miley, H. Röttgering, B. W. Stappers, R. A. M. J. Wijers, S. Zaroubi, M. van den Akker, A. Alexov, J. Anderson, K. Anderson, A. van Ardenne, M. Arts, A. Asgekar, I. M. Avruch, F. Batejat, L. Bähren, M. E. Bell, M. R. Bell, I. van Bemmel, P. Bennema, M. J. Bentum, G. Bernardi, P. Best, L. Bîrzan, A. Bonafede, A. J. Boonstra, R. Braun, J. Bregman, F. Breitling, R. H. van de Brink, J. Broderick, P. C. Broekema, W. N. Brouw, M. Brüggen, H. R. Butcher, W. van Cappellen, B. Ciardi, T. Coenen, J. Conway, A. Coolen, A. Corstanje, S. Damstra, O. Davies, A. T. Deller, R. J. Dettmar, G. van Diepen, K. Dijkstra, P. Donker, A. Doorduin, J. Dromer, M. Drost, A. van Duin, J. Eislöffel, J. van Enst, C. Ferrari, W. Frieswijk, H. Gankema, M. A. Garrett, F. de Gasperin, M. Gerbers, E. de Geus, J. M. Grießmeier, T. Grit, P. Gruppen, J. P. Hamaker, T. Hassall, M. Hoeft, H. A. Holties, A. Horneffer, A. van der Horst, A. van Houwelingen, A. Huijgen, M. Iacobelli, H. Intema, N. Jackson, V. Jelic, A. de Jong, E. Juette, D. Kant, A. Karastergiou, A. Koers, H. Kollen, V. I. Kondratiev, E. Kooistra, Y. Koopman, A. Koster, M. Kuniyoshi, M. Kramer, G. Kuper, P. Lambropoulos, C. Law, J. van Leeuwen, J. Lemaitre, M. Loose, P. Maat, G. Macario, S. Markoff, J. Masters, R. A. McFadden, D. McKay-Bukowski, H. Meijering, H. Meulman, M. Mevius, E. Middelberg, R. Millenaar, J. C. A. Miller-Jones, R. N. Mohan, J. D. Mol, J. Morawietz, R. Morganti, D. D. Mulcahy, E. Mulder, H. Munk, L. Nieuwenhuis, R. van Nieuwpoort, J. E. Noordam, M. Norden, A. Noutsos, A. R. Offringa, H. Olofsson, A. Omar, E. Orrú, R. Overeem, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, J. P. de Reijer, J. Reitsma, G. A. Renting, P. Riemers, E. Rol, J. W. Romein, J. Roosjen, M. Ruiter, A. Scaife, K. van der Schaaf, B. Scheers, P. Schellart, A. Schoenmakers, G. Schoonderbeek, M. Serylak, A. Shulevski, J. Sluman, O. Smirnov, C. Sobey, H. Spreeuw, M. Steinmetz, C. G. M. Sterks, H. J. Stiepel, K. Stuurwold, M. Tagger, Y. Tang, C. Tasse, I. Thomas, S. Thoudam, M. C. Toribio, B. van der Tol, O. Usov, M. van Veelen, A. J. van der Veen, S. ter Veen, J. P. W. Verbiest, R. Vermeulen, N. Vermaas, C. Vocks, C. Vogt, M. de Vos, E. van der Wal, R. van Weeren, H. Weggemans,

P. Weltevrede, S. White, S. J. Wijnholds, T. Wilhelmsson, O. Wucknitz, S. Yatawatta, P. Zarka, A. Zensus, and J. van Zwieten. LOFAR: The LOw-Frequency ARray. *A&A*, 556:A2, Aug. 2013. doi: 10.1051/0004-6361/201220873.

F. Villaescusa-Navarro, S. Naess, S. Genel, A. Pontzen, B. Wandelt, L. Anderson, A. Font-Ribera, N. Battaglia, and D. N. Spergel. Statistical Properties of Paired Fixed Fields. *ApJ*, 867(2):137, Nov. 2018. doi: 10.3847/1538-4357/aae52b.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, Feb. 2020. doi: 10.1038/s41592-019-0686-2.

E. Visbal, R. Barkana, A. Fialkov, D. Tseliakhovich, and C. M. Hirata. The signature of the first stars in atomic hydrogen at redshift 20. *Nature*, 487(7405):70–73, July 2012. doi: 10.1038/nature11177.

F. Wang, F. B. Davies, J. Yang, J. F. Hennawi, X. Fan, A. J. Barth, L. Jiang, X.-B. Wu, D. M. Mudd, E. Banados, F. Bian, R. Decarli, A.-C. Eilers, E. P. Farina, B. Venemans, F. Walter, and M. Yue. A Significantly Neutral Intergalactic Medium Around the Luminous z=7 Quasar J0252-0503. *arXiv e-prints*, art. arXiv:2004.10877, Apr. 2020. doi: 10.48550/arXiv.2004.10877.

C. A. Watkinson, B. Greig, and A. Mesinger. Epoch of reionization parameter estimation with the 21-cm bispectrum. *MNRAS*, 510(3):3838–3848, Mar. 2022. doi: 10.1093/mnras/stab3706.

R. Weinberger, V. Springel, and R. Pakmor. The AREPO Public Code Release. *ApJS*, 248(2):32, June 2020. doi: 10.3847/1538-4365/ab908c.

J. Wolcott-Green, Z. Haiman, and G. L. Bryan. Beyond J_{crit}: a critical curve for suppression of H_2-cooling in protogalaxies. *MNRAS*, 469(3):3329–3336, Aug. 2017. doi: 10.1093/mnras/stx167.

H. Xu, J. H. Wise, M. L. Norman, K. Ahn, and B. W. O'Shea. Galaxy Properties and UV Escape Fractions during the Epoch of Reionization: Results from the Renaissance Simulations. *ApJ*, 833(1):84, Dec. 2016. doi: 10.3847/1538-4357/833/1/84.

J. Yang, F. Wang, X. Fan, J. F. Hennawi, F. B. Davies, M. Yue, E. Banados, X.-B. Wu, B. Venemans, A. J. Barth, F. Bian, K. Boutsia, R. Decarli, E. P. Farina, R. Green, L. Jiang, J.-T. Li, C. Mazzucchelli, and F. Walter. Pōniuā'ena: A Luminous z = 7.5 Quasar Hosting a 1.5 Billion Solar Mass Black Hole. *ApJ*, 897(1):L14, July 2020. doi: 10.3847/2041-8213/ab9c26.

S. Yoshiura, T. Minoda, and T. Takahashi. EDGES and JWST with 21cm global signal emulator. *arXiv e-prints*, art. arXiv:2305.11441, May 2023. doi: 10.48550/arXiv.2305.11441.

B. Yue, A. Ferrara, and Y. Xu. On the faint-end of the high-z galaxy luminosity function. *MNRAS*, 463(2):1968–1979, Dec. 2016. doi: 10.1093/mnras/stw2145.

O. Zahn, A. Mesinger, M. McQuinn, H. Trac, R. Cen, and L. E. Hernquist. Comparison of reionization models: radiative transfer simulations and approximate, seminumeric models. *MNRAS*, 414(1):727–738, June 2011. doi: 10.1111/j.1365-2966.2011.18439.x.

P. Zarka, J. N. Girard, M. Tagger, and L. Denis. LSS/NenuFAR: The LOFAR Super Station project in Nançay. In S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud, and H. Wozniak, editors, *SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pages 687–694, Dec. 2012.

Y. G. Zhang, A. Liu, and A. R. Parsons. Unlocking Sensitivity for Visibility-based Estimators of the 21 cm Reionization Power Spectrum. *ApJ*, 852(2):110, Jan. 2018. doi: 10.3847/1538-4357/aaa029.

X. Zhao, Y. Mao, C. Cheng, and B. D. Wandelt. Simulation-based Inference of Reionization Parameters from 3D Tomographic 21 cm Light-cone Images. *ApJ*, 926(2):151, Feb. 2022a. doi: 10.3847/1538-4357/ac457d.

X. Zhao, Y. Mao, and B. D. Wandelt. Implicit Likelihood Inference of Reionization Parameters from the 21 cm Power Spectrum. *ApJ*, 933(2):236, July 2022b. doi: 10.3847/1538-4357/ac778e.

X. Zhao, Y. Mao, S. Zuo, and B. D. Wandelt. Simulation-based Inference of Reionization Parameters from 3D Tomographic 21 cm Light-cone Images. II. Application of Solid Harmonic Wavelet Scattering Transform. *ApJ*, 973(1):41, Sept. 2024. doi: 10.3847/1538-4357/ad5ffo.