



Exploring the Cosmic Dawn through the 21 cm Forest and High-redshift Radio Sources with the SKA

Junsong Cang^{1,2}, Benedetta Ciardi,³ Barun Maity⁴, Atsushi J. Nishizawa^{5,6}, Qi Niu⁷, Yue Shao⁸, Abinash Kumar Shaw⁹, Hayato Shimabukuro^{10,11,12}, Tomáš Šoltinský^{13,14,15}, Tian-Yang Sun⁷, Tsutomu T. Takeuchi^{12,16}, Yidong Xu^{17,18}, Kohji Yoshikawa¹⁹, Bin Yue^{17,18} and Xin Zhang^{7,20,21}

¹*Theoretical and Scientific Data Science, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136 Trieste, Italy*

²*Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy*

³*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany*

⁴*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*

⁵*Gifu Shotoku Gakuen University, 1-1 Takakuwanishi, Yanaizucho, Gifu, 501-6122, Japan*

⁶*Kobayashi Maskawa Institute, Nagoya University, Furocho, Chikusaku, Nagoya, 461-8602, Japan*

⁷*Key Laboratory of Cosmology and Astrophysics (Liaoning), College of Sciences, Northeastern University, Shenyang 110819, China*

⁸*Department of Physics, Liaoning Normal University, Dalian 116029, China*

⁹*Department of Computer Science, University of Nevada, Las Vegas, NV 89154, USA*

¹⁰*South-Western Institute for Astronomy Research (SWIFAR), Yunnan University, Kunming, Yunnan 650500, People's Republic of China*

¹¹*Key Laboratory of Survey Science of Yunnan Province, Yunnan University, Kunming, Yunnan 650500, People's Republic of China*

¹²*Nagoya University, Graduate School of Science, Division of Particle and Astrophysical Science, Chikusa-Ku, Nagoya, 464-8602, Japan*

¹³*INAF-Osservatorio Astronomico di Trieste, Via G.B. Tiepolo, 11, I-34143 Trieste, Italy*

¹⁴*IFPU, Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34151 Trieste, Italy*

¹⁵*INFN, Sezione di Trieste, Via Valerio 2, I-34127 Trieste, Italy*

¹⁶*The Research Center for Statistical Machine Learning, the Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

¹⁷*National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China*

¹⁸*State Key Laboratory of Radio Astronomy and Technology, Beijing 100101, People's Republic of China*

¹⁹*Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan*

²⁰*National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang 110819, China*

²¹*Key Laboratory of Data Analytics and Optimization for Smart Industry (Ministry of Education), Northeastern University, Shenyang 110819, China*

E-mail: shimabukuro@ynu.edu.cn

The 21 cm forest, manifesting as absorption features in the spectra of distant radio sources, is caused by neutral hydrogen in the intervening cosmic structures. It provides a unique opportunity to directly probe the neutral intergalactic medium (IGM) during the epoch of reionization. In particular, this phenomenon offers a distinctive probe of small-scale structures during the epoch of reionization (EoR), and can be used to sensitively constrain the thermal history of the early universe. However, real measurement of the 21 cm forest signals has been hindered by the high sensitivity required to detect the weak individual absorption lines, and the unavailability of high-redshift radio-bright sources at the EoR. In the past decade, significant progress has been made to make the 21 cm forest measurement actually feasible and promising. These include the development of new statistical observables to enhance the sensitivity of measurements with a reasonable observation time, the revision of the expected number counts of background radio-loud quasars in light of recent observations, and the development of deep-learning-based tools to extract physical information more efficiently and accurately. More importantly, methods have been proposed to break the degeneracy between astrophysical effects from the early galaxies and various physical effects that influence the formation of small-scale structures. Therefore, with the advent of the Square Kilometre Array (SKA), it is very promising to utilize the 21 cm forest to probe both the heating effect from early structure formation, or possibly from exotic energy-injection processes, and a range of fundamental physics, such as the dark matter properties, the neutrino mass, the running spectral index, and the relative velocity between dark matter and baryons. In this chapter, we provide a comprehensive overview of recent efforts in the field of 21 cm forest research, and discuss observational strategies and prospects for constraining the first galaxies and the fundamental physics with 21 cm forest observations using the upcoming SKA-Low.

1 Overview of the 21 cm Forest

The 21 cm forest refers to the absorption features caused by neutral hydrogen (HI) observed against high-redshift (high- z) radio-loud sources such as quasars. This phenomenon occurs when cold neutral hydrogen in the intergalactic medium (IGM) or minihalos along the line of sight absorbs radiation from background sources, imprinting characteristic absorption lines onto their spectra. It is a unique method to probe the early universe, complementing approaches such as tomography and power spectrum analyses while providing distinct insights (e.g. Carilli et al., 2002; Furlanetto and Loeb, 2002; Furlanetto, 2006).

An instructive analogy is the well-known Lyman- α forest, which is absorption lines from neutral hydrogen seen in quasar spectra (e.g. Weinberg, 2003; McQuinn, 2016). Like the 21 cm forest, it probes the IGM along the line of sight, but differs in key ways. The Lyman- α forest appears in the UV/optical and traces warmer, ionized or partially neutral gas, while the 21 cm forest, seen in radio, is sensitive to cold neutral hydrogen. Moreover, whereas the Lyman- α forest is mainly used at the end stage of reionization ($z \lesssim 6$), the 21 cm forest can explore higher redshifts during the initial stage of reionization.

One of the key quantities that characterize the 21 cm forest is the optical depth τ_ν , which quantifies the degree of absorption experienced by radio waves due to neutral hydrogen (e.g. Furlanetto and Loeb, 2002; Furlanetto, 2006). The optical depth of a cloud of hydrogen can be expressed as

$$\tau_\nu = \frac{3h_p c^3 A_{10}}{32\pi^{3/2} k_B} \frac{1}{\nu^2} \int_{-\infty}^{+\infty} \frac{n_{\text{HI}}(x)}{b(x) T_S(x)} \exp\left[-\frac{(u(\nu) - v(x))^2}{b^2(x)}\right] dx \quad (1)$$

Here, c is the speed of light, h_p is Planck's constant, A_{10} represents the Einstein coefficient for the hyperfine 21 cm transition, with a value of $2.85 \times 10^{-15} \text{ s}^{-1}$, governing the likelihood of spontaneous emission, and k_B is the Boltzmann constant. $n_{\text{HI}}(x)$ is the neutral hydrogen number density at position x along the line of sight (LoS), $b(x) = (2k_B T_K(x)/m_H)^{1/2}$ is the Doppler parameter, in which m_H is the hydrogen atom mass, $T_K(x)$ is the gas kinetic temperature, and $T_S(x)$ is the spin temperature, which characterizes the relative population of the two hyperfine energy levels of neutral hydrogen. $u(\nu) \equiv c(\nu - \nu_{21 \text{ cm}})/\nu_{21 \text{ cm}}$, where $\nu_{21 \text{ cm}} \approx 1420.4 \text{ MHz}$ is the rest-frame frequency of the 21 cm transition, and $v(x)$ is the velocity of the hydrogen gas projected to the LoS, including the peculiar velocity and the Hubble flow.

When the optical depth is small, the absorption lines are weak and appear only as subtle fluctuations in the background spectrum. Conversely, a larger optical depth produces deeper and more prominent absorption features, reflecting regions of higher density or lower temperature in the neutral hydrogen gas. Because the strength and shape of these lines are sensitive to the local density, temperature, and ionization state of the IGM, analyzing the 21 cm forest enables us to infer the small-scale distribution of matter and to trace the thermal evolution of the IGM during the early stages of cosmic history.

The 21 cm forest is complementary to the 21 cm emission line, which is observed as a diffuse signal from neutral hydrogen itself. While the emission signal is typically measured over large areas of the sky using statistical methods such as power spectrum analysis or tomographic imaging,

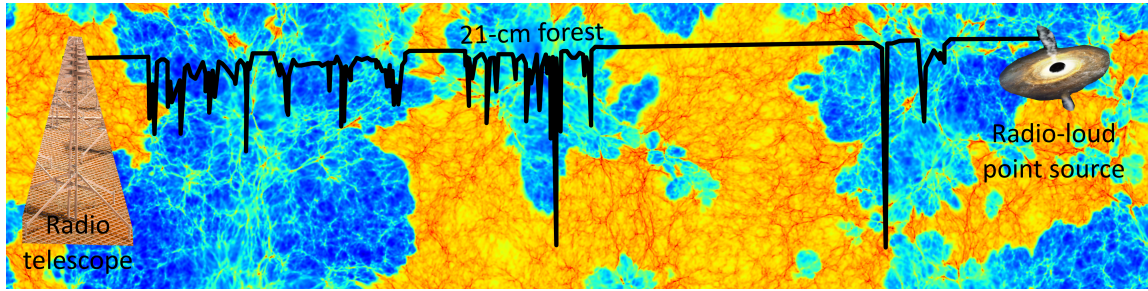


Figure 1: Schematic view of 21 cm forest. Credit: SKAO, Sherwood-relics, International Gemini Observatory, Tomáš Šoltinský.

which thereby offers a global view of the IGM during cosmic dawn and the Epoch of Reionization, the 21 cm forest provides a high-resolution, line-of-sight probe of the intervening gas, revealing detailed information about small-scale structures and local physical conditions. Together, these approaches offer a comprehensive picture: the emission line maps the large-scale distribution of neutral hydrogen, whereas the forest targets the finer details of gas density, temperature, and kinematics along a discrete line of sight.

Observations of the 21 cm forest using next-generation radio telescopes, such as the Square Kilometre Array (SKA), are poised to provide transformative insights into cosmic structure formation and reionization (e.g. Ciardi et al., 2015; de Lera Acedo et al., 2026). These studies promise to deepen our understanding of the early universe by probing the distribution and properties of neutral hydrogen along the line of sight (e.g. Bag et al., 2026; Barkana et al., 2026).

This chapter provides a comprehensive overview of theoretical modeling, numerical simulations, data-analysis techniques, and observational strategies for studying 21 cm forest and its connection to high-redshift radio sources in the context of future SKA-Low observations.

2 Modelling the 21 cm forest

2.1 Analytical vs numerical approach

Absorption features in the spectra of high- z quasars induced by intervening minihalos and dwarf galaxies have been investigated with a semi-analytic approach by e.g. Furlanetto and Loeb (2002), Furlanetto (2006), Meiksin (2011), Xu et al. (2011), and Shimabukuro et al. (2014). Semi-analytic approaches can provide detailed modeling for small-scale structures that are currently impossible to resolve for numerical hydrodynamical simulations, but can be important for 21 cm forest signals, for which the main contributors of the absorption are small-scale structures. We need to model the density, temperature, and ionization profiles not only for the hydrogen gas within dark matter halos, but also for the infalling gas surrounding halos. In particular, the cool and dense gas in the infalling region surrounding halos could have a major contribution to the optical depth when the gas has not been sufficiently heated by the X-rays (Xu et al., 2011). The semi-analytic approach also has the advantage of being able to explore a large parameter space due to its efficient modeling. Similar advantages have fully-analytic models, one of which was developed for the 1D power spectrum of the 21 cm forest by Shao et al. (2025).

On the other hand, numerical hydrodynamical simulations combined with radiative (RT) transfer can provide more detailed models of the IGM during the Epoch of Reionization. Examples of numerical simulations which have been used for 21 cm forest studies include AURORA (Pawlik et al., 2017; Bhagwat et al., 2022), CRASH (Ciardi et al., 2012, 2013), Sherwood-Relics (Puchwein et al., 2023; Šoltinský et al., 2021, 2023) and others (Carilli et al., 2002; Xu et al., 2009; Semelin, 2016; Naruse et al., 2024). From such simulations, we extract, along each line of sight, the hydrogen number density n_{H} , neutral fraction x_{HI} , kinetic and spin temperatures (T_{K} , T_{S}), and the peculiar velocity of the gas v_{pec} . With this information we can calculate the 21 cm forest optical depth in discrete form at pixel i in the LOS with N_{pix} pixels in total as

$$\tau_{21 \text{ cm}, i} = \frac{3h_{\text{p}}c^3 A_{10}}{32\pi^{3/2} v_{21 \text{ cm}}^2 k_{\text{B}}} \frac{\delta v}{H(z)} \sum_{j=1}^{N_{\text{pix}}} \frac{n_{\text{HI}, j}}{b_j T_{\text{S}, j}} \exp\left(-\frac{(v_{\text{H}, i} - u_j)^2}{b_j^2}\right), \quad (2)$$

where δv is the velocity width of the pixels, $H(z)$ is the Hubble parameter, $n_{\text{HI}} = n_{\text{H}} x_{\text{HI}}$, $u_j = v_{\text{H}, j} + v_{\text{pec}, j}$ and v_{H} is the Hubble velocity. However, most simulations include only the ionizing (UV) radiation that drives reionization, while neglecting the contribution of X-ray photons. The latter, however, play a crucial role in determining the thermal state of the IGM and hence the 21 cm forest signal. Owing to their long mean free path,

$$\lambda_{\text{X}} \simeq 1 \text{ cMpc } x_{\text{HI}}^{-1} (1 + \delta)^{-1} \left(\frac{E_{\text{X}}}{0.2 \text{ keV}}\right)^3 \left[\frac{1+z}{10}\right]^{-2}, \quad (3)$$

where δ is the matter overdensity contrast and E_{X} is the energy of the X-ray photon. Since the mean free path increases rapidly with photon energy ($\lambda_{\text{X}} \propto E_{\text{X}}^3$) and decreases in dense or neutral regions, X-ray photons can penetrate far beyond local ionized bubbles, heating large volumes of the IGM. Therefore, their effect is often approximated as a spatially homogeneous background, whose luminosity can be parameterized as (Furlanetto, 2006)

$$L_{\text{X}} = 3.4 \times 10^{40} \text{ erg s}^{-1} f_{\text{X}} \left(\frac{\text{SFR}}{1 M_{\odot} \text{ yr}^{-1}}\right), \quad (4)$$

where the f_{X} is the X-ray background radiation efficiency and the SFR is the star formation rate, which preheats the IGM in the post-processing of cosmological simulations, rather than following the more computationally expensive RT of the photons. Differently from analytic approaches, the numerical simulations mentioned above are computationally expensive, and hence it is difficult to explore a large parameter space with them.

An alternative approach that requires less computational resources is to utilize semi-numerical tools. For example, the 21cmFAST code (Mesinger et al. 2011; Murray et al. 2020, but see also Mack and Wyithe 2012) has been widely used to vary the X-ray background radiation, as well as the reionization and dark matter models and to investigate the impact on the 21 cm forest (e.g. Ewall-Wice et al., 2014; Thyagarajan, 2020; Shao et al., 2023; Šoltinský et al., 2025). Combining the semi-numerical tools with the semi-analytic modeling, a high enough dynamic range can thus be reached to model in detail both the large-scale environment and the small-scale physics driving the signal (Shao et al., 2023).

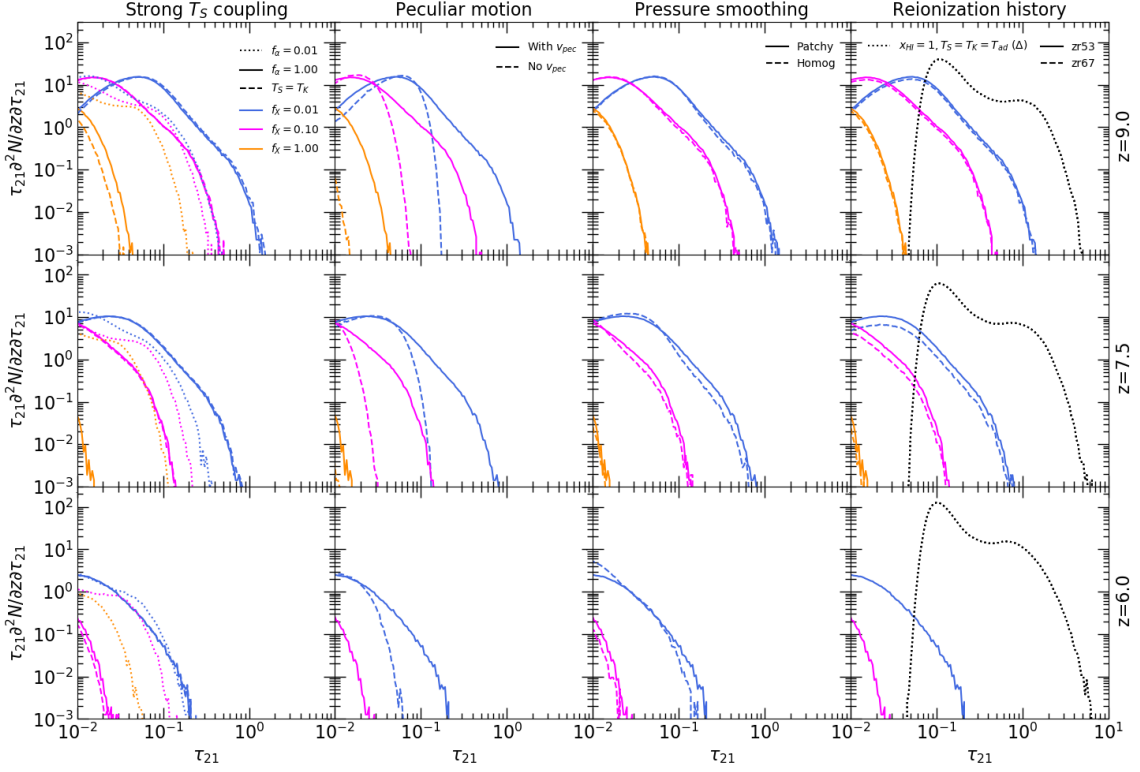


Figure 2: The differential number density of synthetic 21 cm forest absorption features at $z = 9$ (top row panels), 7.5 (middle row panels) and 6 (bottom row panels). Each panel shows cases in which the X-ray efficiencies, $f_X = 0.01$ (blue curves), $f_X = 0.1$ (fuchsia curves) and $f_X = 1$ (orange curves) is assumed. From left to right various effects on the 21 cm forest are tested, namely strength of Ly α coupling of the T_S , redshift space distortions, topology and timing of reionization. This figure is taken from Šoltinský et al. (2021).

2.2 Effect of assumptions implemented in the modelling

In the previous section, we discussed various analytical, semi-analytical, and numerical approaches that have been used to model the 21 cm forest, each with its own advantages and limitations. Despite their methodological differences, these models are generally built upon a set of simplifying assumptions regarding the thermal and ionization state of the intergalactic medium and the relevant radiative processes. In what follows, we describe some of these assumptions and how they affect the detectability of the 21 cm forest signal.

During the end stages of the reionization the universe is expected to be filled with a homogeneous Ly α photons background strong enough for the spin temperature to be coupled to the gas kinetic temperature via the Wouthuysen-Field effect (Wouthuysen, 1952; Field, 1958). Hence, typically it is assumed that $T_S = T_K$, which allows us to avoid the complex calculation of T_S . However, this is not necessarily the case, and in fact the spin temperature could also be partially coupled to the radio background temperature. As a result, the 21 cm forest absorption can be modestly increased or decreased in comparison to a $T_S = T_K$ case, depending on the redshift evolution of T_K (Semelin, 2016; Šoltinský et al., 2021). This is shown in the leftmost column panels in Fig. 2, where the distribution of the absorption lines in the simulated 21 cm forest spectra is shown when the $T_S = T_K$

approximation was assumed (dashed curves), a full computation of T_S was performed (solid curves) and weak Ly α coupling was considered (dotted curves).

Another common approximation is to omit the effect of redshift space distortions (RSD), which can increase the optical depth of individual absorption features and clusters them together. As a consequence, the inclusion of RSD can boost the high τ_{21} tail of the distribution of absorption features by up to an order of magnitude (Semelin, 2016; Šoltinský et al., 2021) as we can see in the middle left column of Fig. 2, as well as the 1D power spectrum at small scales (Šoltinský et al., 2025). We note that hydrodynamical simulations naturally produce the gas peculiar velocity fields, v_{pec} , which can then be used to incorporate the effect of RSD into the 21 cm forest modeling.

On the other hand, Šoltinský et al. (2021) found that the hydrodynamical response of the gas to the inhomogeneous (patchy) reionization has a modest effect (see the middle right column panels in Fig. 2). The structures in the neutral islands of the IGM which have not yet experienced reionization (and thus Jeans pressure smoothing) produce 21 cm forest absorption features which are only slightly stronger than those in a case of homogeneous reionization.

Besides the morphology of the reionization, the timing can affect the detectability of the 21 cm forest too. For example, in the rightmost middle panel of Fig. 2, which corresponds to $z = 7.5$, the model in which the reionization is completed at $z_r = 6.7$ (dashed curves) has slightly suppressed signal relatively to the model with $z_r = 5.3$ (solid curves) because the reionization has progressed faster and more IGM is ionized and hot. For comparison a case of no reionization and no X-ray heating (i.e. $T_S = T_K = T_{\text{ad}} = 2.73 \text{ K} (1 + \delta)^{2/3} (1 + z)^2 / (1 + z_{\text{dec}})$, where T_{ad} is the adiabatic temperature, $z_{\text{dec}} = 147.8$ (Furlanetto et al., 2006), and $x_{\text{HI}} = 1$) is indicated by dotted black curves. Recent observations of the large spatial fluctuations in the Ly α forest opacity (Becker et al., 2015; Bosman et al., 2022), the presence of damping wings in the Ly α forest spectra from neutral islands at $z < 6$ (Spina et al., 2024; Becker et al., 2024; Zhu et al., 2024; Sawyer et al., 2025), Ly α equivalent widths (Nakane et al., 2024), thermal widths of Ly α forest transmission spikes at $z > 5$ (Gaikwad et al., 2020), clustering of Ly α emitters (Weinberger et al., 2019), deficit of Ly α emitting galaxies around extended Ly α absorption troughs (Kashino et al., 2020; Keating et al., 2020; Christenson et al., 2021), long dark gaps in the Ly α forest (Zhu et al., 2021) and Ly β forest (Zhu et al., 2022), mean free path of ionizing photons at $z = 6$ (Becker et al., 2021; Cain et al., 2021; Zhu et al., 2023; Gaikwad et al., 2023), the cross-correlation of the Ly α forest-[O III] emitters (Kakiichi et al., 2025) and the cosmic microwave background measurement of the electron scattering optical depth (Pagano et al., 2020) are consistent with the reionization not being completed before $z \sim 6$ (e.g. Kulkarni et al., 2019; Nasir and D’Aloisio, 2020; Choudhury et al., 2021; Qin et al., 2021). For alternative explanations and the evidence against this see D’Aloisio et al. (2015), Davies and Furlanetto (2016), Chardin et al. (2017), Meiksin (2020) and Zheng et al. (2026). However, if the late-end reionization scenario is indeed correct, we can expect large islands of neutral hydrogen to persist until $z \sim 6$ (e.g. Lidz et al., 2007; Mesinger, 2010), and hence the 21 cm forest absorption might be detectable at such low redshift if the gas is not preheated above $T_S \sim 100 \text{ K}$ (Šoltinský et al., 2021).

Furthermore, one of the open questions in the field is how much do minihalos contribute to the 21 cm forest absorption (Furlanetto, 2006; Xu et al., 2011; Kadota et al., 2023; Naruse et al., 2024), as minihalos are difficult to model in numerical simulations due to the large dynamic range required.

Indeed, large simulation boxes are needed to capture the rarest and most massive halos, but at the same time the modeling of minihalos requires fine resolution. Therefore, most studies employing cosmological simulations focus on the 21 cm forest originating from the diffuse IGM only (e.g. Ciardi et al., 2013; Šoltinský et al., 2021). Additionally, feedback processes may easily suppress the absorption signature from minihalos (Meiksin, 2011; Park et al., 2016; Nakatani et al., 2020). It has been shown, though, that the presence of minihalos impacts the evolution of reionization itself (Ciardi et al., 2006; Yue et al., 2009; Chan et al., 2024), and therefore can indirectly affect also the 21 cm forest.

Finally, often the effect of the background radio-loud quasar radiation on the proximate 21 cm forest is omitted. By postprocessing cosmological hydrodynamical/RT simulations with a multifrequency 1D RT simulations incorporating quasar spectrum from the UV to X-ray photons to model the impact of the quasar radiation on its surroundings, Šoltinský et al. (2023) found that not only the signal is completely suppressed close to the quasar due to the H II bubble pre-ionized by the UV radiation from the galaxies populating this region, but can be partially suppressed far away from the quasar because of the heating driven by the X-ray photons emitted by the quasar itself. It should be noted that, because of their long mean free path, X-ray photons penetrate the IGM beyond the ionization front far away from the background radio-loud quasar.

3 Evaluating 21 cm forest with the SKA

3.1 Direct observation of spectra & sensitivity

Previous works propose a direct counting method to detect the 21 cm absorption lines imprinted by cold neutral hydrogen in and around minihalos and dwarf galaxies along the sightlines to bright, high-redshift radio sources (e.g. Furlanetto and Loeb, 2002; Furlanetto, 2006; Xu et al., 2011). In this approach, the optical depth of a minihalo is determined by the neutral hydrogen number density, the spin temperature of the gas, and the line profile function that describes the velocity dispersion of the gas. The optical depth depends on the impact parameter, which corresponds to the maximum distance along the line of sight at which significant absorption occurs.

By modeling the halo mass function and the internal and ambient gas properties of each minihalo, the cumulative number of absorption features per unit redshift is given by

$$\frac{dN(> \tau)}{dz} = \frac{dr}{dz} \int_{M_{\min}}^{M_{\max}} \frac{dN}{dM} \pi r_{\tau}^2(M, \tau) dM,$$

where $\frac{dr}{dz}$ is the comoving line element, $\frac{dN}{dM}$ is the halo mass function, and $r_{\tau}(M, \tau)$ is the maximum impact parameter for which the optical depth exceeds τ .

This direct counting strategy requires the high sensitivity and kHz-level frequency resolution of instruments to detect weak absorption signals (with $\tau \sim 0.01$ – 0.1). Ultimately, the statistical distribution of these 21 cm absorption lines provides a powerful tool to probe small-scale cosmological fluctuations and to constrain key parameters such as the neutrino mass, the running spectral index, and dark matter particle mass (Shimabukuro et al., 2014). Similarly, Šoltinský et al. (2021) have used the direct counting strategy for the 21 cm absorption lines originating from the diffuse IGM

rather than minihalos. They also suggest that a null-detection of these features in observations by SKA1-low or SKA2-low can be translated to a lower limit on the X-ray background radiation.

3.2 Statistical measurements & sensitivity

In statistical analysis, the brightness temperature of the 21 cm forest. The brightness temperature includes not only information about the optical depth but also contains information about the background radio source. This can be expressed as

$$\delta T_b(\hat{s}, \nu) \approx \frac{T_S(\hat{s}, z) - T_\gamma(\hat{s}, z)}{1 + z} \tau(\hat{s}, z), \quad (5)$$

where \hat{s} denotes the unit vector along the line of sight (i.e., the observing direction), T_S is the spin temperature of neutral hydrogen, and T_γ is the brightness temperature of the background radiation, which is related to the flux density of the background radio source.

Compared with direct detection, statistical analysis methods, such as increased variance (Mack and Wyithe, 2012), power spectrum (Thyagarajan, 2020; Shao et al., 2023), wavelet transform (Shimabukuro et al., 2025) and topological data analysis(Shimabukuro, 2026), have significantly improved the observability of the 21 cm forest. First of all, statistical methods effectively suppress noise through signal superposition. Even if individual absorption lines are difficult to distinguish, they can still extract the overall statistical characteristics and significantly improve the signal-to-noise ratio. Secondly, they are less sensitive to instrumental calibration errors. By analyzing the spatial or frequency correlation of signals, the impact of systematic errors can be reduced. In addition, statistical methods have relatively relaxed requirements for data resolution. Even if the observational resolution is limited, they can still constrain the signal characteristics by measuring the amplitude of fluctuations on different scales.

3.2.1 Power spectrum

The power spectrum is one of the most commonly used statistical analysis methods in cosmology and astronomy(e.g. Acharya et al., 2026a). It quantitatively describes the fluctuation characteristics of signals at different scales through Fourier transform. In the study of the 21 cm forest, Thyagarajan (2020) first proposed the application of one-dimensional (1D) power spectrum along the line of sight, and it has been shown that this 1D power spectrum analysis can effectively suppress the impact of cosmic variance and significantly improve the signal-to-noise ratio of the observed signal. Shao et al. (2023) further found that the 1D power spectrum of the 21 cm forest has an important scientific value, capable of breaking the degeneracy between dark matter properties and intergalactic medium temperature in traditional observations, thereby achieving simultaneous constraints on the mass of dark matter particles and the history of cosmic heating. Similarly, when the 1D power spectrum is computed from the transmitted 21 cm forest flux, Šoltinský et al. (2025) have found that the measurement of this statistical observable can be used to constrain both the thermal and ionization state of the high- z IGM even with an observation of a few radio-background sources. These groundbreaking discoveries provide a new theoretical framework for studying early universe physics using the 21 cm forest.

The observed frequency-dependent brightness temperature $\delta T_b(\hat{s}, \nu)$ can also be represented in terms of the line-of-sight distance r_z , as $\delta T_b(\hat{s}, r_z)$. The Fourier transform of $\delta \tilde{T}'_b(\hat{s}, r_z)$ is defined

as:

$$\delta\tilde{T}'(\hat{s}, k_{\parallel}) = \int \delta T'_b(\hat{s}, r_z) e^{-ik_{\parallel}r_z} dr_z. \quad (6)$$

The corresponding one-dimensional power spectrum is

$$P(\hat{s}, k_{\parallel}) = |\delta\tilde{T}'(\hat{s}, k_{\parallel})|^2 \left(\frac{1}{\Delta r_z} \right). \quad (7)$$

The factor $1/\Delta r_z$ serves as the normalization factor, with Δr_z denoting the length of the sightline in question. By extracting multiple segments of neutral regions from the spectra of different radio sources, we can obtain the expected value of the one-dimensional power spectrum $P(k_{\parallel}) \equiv \langle P(\hat{s}, k_{\parallel}) \rangle$.

The uncertainties in 21 cm forest observations primarily stem from the thermal noise of the interferometer array and the sample variance in power spectrum measurements. The sample variance of the one-dimensional power spectrum can be expressed as $P^S = \sigma_P(k)/\sqrt{N_s \times N_m}$, where $\sigma_P(k)$ is the standard deviation of $P(k)$, N_s represents the number of independent measurements on different neutral patches of Δr_z , and N_m is the number of independent modes in each k -bin from each measurement. Regarding the thermal noise error, [Shao et al. \(2023\)](#) follows the approach taken by [Thyagarajan \(2020\)](#), assuming each spectrum is measured twice separately, or the total integration time is divided into two halves, and the cross-power spectrum is practically measured to avoid noise bias. Under this measurement scheme, the integration time for each measurement of the spectrum is $\delta t_{0.5} = 0.5\delta t$, resulting in an increase in the thermal noise on the spectrum by a factor of $\sqrt{2}$. Then, the uncertainty in the 1D power spectrum due to thermal noise is given by

$$P^N = \left(\frac{\lambda_z^2 T_{\text{sys}}^2}{A_{\text{eff}}} \right) \left(\frac{\Delta r_z}{2\Omega \Delta \nu_z \delta t_{0.5}} \right) \quad (8)$$

where $A_{\text{eff}}/T_{\text{sys}}$ represents the sensitivity of the telescope array, $\Omega = \pi(\theta/2)^2$ is the solid angle of the telescope beam, θ denotes the angular resolution, and $\Delta \nu_z$ is the total observing bandwidth corresponding to Δr_z . For N_s measurements, the thermal noise P^N will further decrease by a factor of $\sqrt{N_s}$.

Figure 3 shows the 1D power spectra that we have simulated. The left panel displays the 1D power spectra corresponding to different heating histories in the CDM model, where $f_X = 0, 0.1, 1, \text{ and } 3$ represent different heating histories. As f_X increases, the temperature of the IGM rises, leading to a significant suppression of the power spectrum across all scales. The right panel illustrates the 1D power spectra for the CDM model without heating and for WDM models with particle masses of 10 keV, 6 keV, and 3 keV, respectively. A lower m_{WDM} reduces small-scale density fluctuations, thereby suppressing the small-scale power spectrum. Unlike heating effects, the influence of dark matter properties is primarily reflected in changes to the slope of the power spectrum, rather than a significant suppression of the overall amplitude. The dotted lines and dashed lines represent the thermal noise corresponding to the sensitivities of the SKA1-LOW and SKA2-LOW, respectively, for which we have adopted parameters of $A_{\text{eff}}/T_{\text{sys}} = 800 \text{ m}^2/\text{K}$ and $4000 \text{ m}^2/\text{K}$. Additionally, [Shao et al. \(2023\)](#) has assumed an integration time of 100 hours on each source, a spectral resolution of

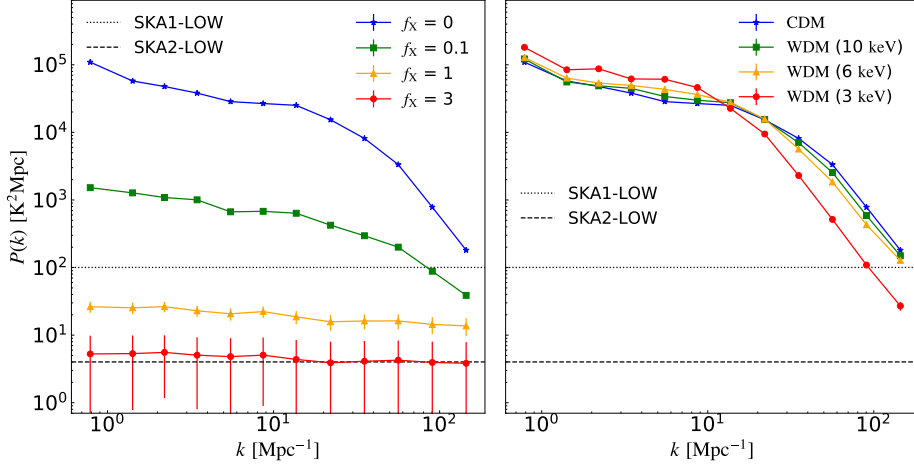


Figure 3: The expected 1D power spectrum of the 21 cm forest for different f_X (left panel) and m_{WDM} (right panel), with $S_{150} = 10$ mJy. This figure is taken from [Shao et al. \(2023\)](#).

1 kHz, and 100 observed neutral segments. Because heating and dark matter effects leave distinct imprints on the 21 cm forest 1D power spectrum, the two can be disentangled, allowing us to simultaneously probe the thermal history of the Universe and the properties of dark matter.

3.2.2 Wavelet Scattering Transform

[Shimabukuro et al. \(2025\)](#) introduced the wavelet scattering transform (WST) to study the 21 cm forest beyond the power spectrum. WST is a multi-scale method that decomposes a signal into scale-dependent coefficients and is well suited to complex data with nested structures and non-Gaussian features. Applied to the 21 cm forest brightness temperature, it captures localized fluctuations and cross-scale interactions through wavelet filters of different sizes. Its nonlinear modulus and averaging operations also make it stable against small perturbations and robust to instrumental and astrophysical noise. The resulting coefficients include first-order terms (S_1), which trace intensity variations at individual scales, and second-order terms (S_2), which quantify hierarchical coupling between scales.

The WST first convolves the 21 cm forest brightness temperature spectrum δT_b with a family of wavelets $\psi_{j,l}(x)$. In the WST, the scale index j determines the dilation of the wavelet filter, that is, how much the wavelet is stretched or compressed in scale, while the orientation index l specifies the direction of the wavelet. Small j corresponds to fine structures or rapid oscillations (small scales), whereas large j captures broad, slowly varying patterns (large scales). Different l values allow the WST to extract directional features from various orientations in the data. However, l is redundant in one-dimensional 21 cm forest spectrum data.

The hierarchical coefficients start with the spatial mean

$$S_0 = \langle \delta T_b \rangle, \quad (9)$$

which encodes global information such as the average brightness temperature. Localised fluctuations at scale j_1 are captured by the first-order coefficients obtained after convolution, modulus, and spatial

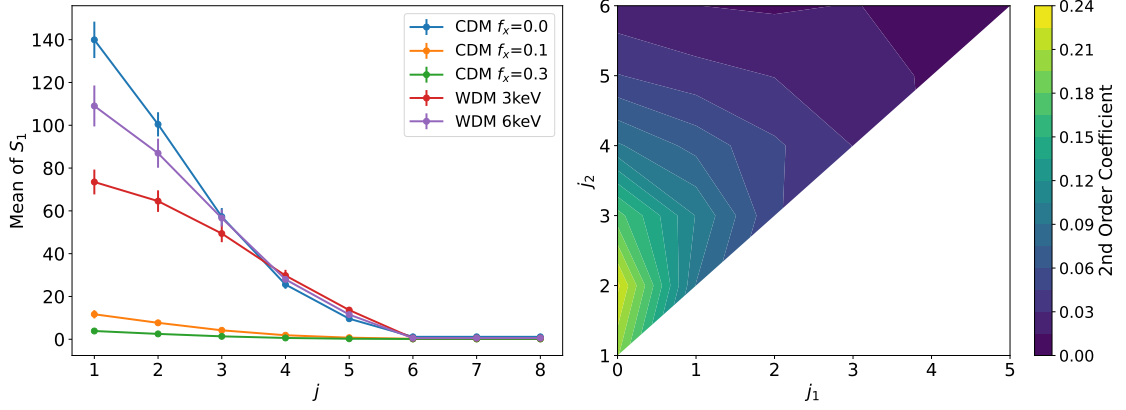


Figure 4: (left) first-order WST coefficient for CDM model, CDM + $f_X = 0.1, 0.3$ models and Warm Dark Matter(WDM) models with different WDM mass. (right) second-order WST coefficient for CDM model. This figure is taken from Shimabukuro et al. (2025).

averaging,

$$S_1^{j_1} = \left\langle \left| \delta T_b * \psi_{j_1}(x) \right| \right\rangle. \quad (10)$$

To probe cross-scale structure, the modulus field $|\delta T_b * \psi_{j_1}|$ is convolved with a second wavelet of scale j_2 ; taking the modulus again and averaging yields the second-order coefficients

$$S_2^{j_1, j_2} = \left\langle \left| \left| \delta T_b * \psi_{j_1}(x) \right| * \psi_{j_2}(x) \right| \right\rangle, \quad (11)$$

which quantify how structures at scale j_1 modulate those at scale j_2 and thus encode higher-order, non-Gaussian statistics.

In Fig.4, we show the first and second orders of the WST coefficient. The left panel shows the scale-dependent first-order scattering coefficients, $\langle S_1 \rangle$, as a function of the wavelet scale $j (= j_1)$ for five cosmological models: cold dark matter (CDM) with $f_X = 0.0, 0.1, 0.3$, and warm dark matter (WDM) with particle masses of 3 keV and 6 keV. In all cases, $\langle S_1 \rangle$ falls off rapidly with increasing j , indicating that small-scale fluctuations dominate the WST response; models with more free-streaming (higher f_X or lighter WDM) yield systematically lower amplitudes at each j . The right panel presents the second-order scattering coefficients, S_2 , as a function of the scale pair (j_1, j_2) . The strongest cross-scale coupling occurs at the smallest scales (low j_1 and j_2), and S_2 declines steeply as either index increases, reflecting the decreasing interaction between widely separated scales.

3.2.3 Sample size and statistical problem

As discussed above, the 21 cm forest is expected to provide a powerful probe of large-scale structure and galaxy formation in the high- z Universe. However, even with the SKA, the number of sufficiently bright background radio sources at such redshifts is expected to be limited (Takeuchi et al., 2016, and references therein). The total number of usable sightlines may reach only a few tens to ~ 100 over the full survey. In contrast, the high frequency resolution of the SKA yields a very large

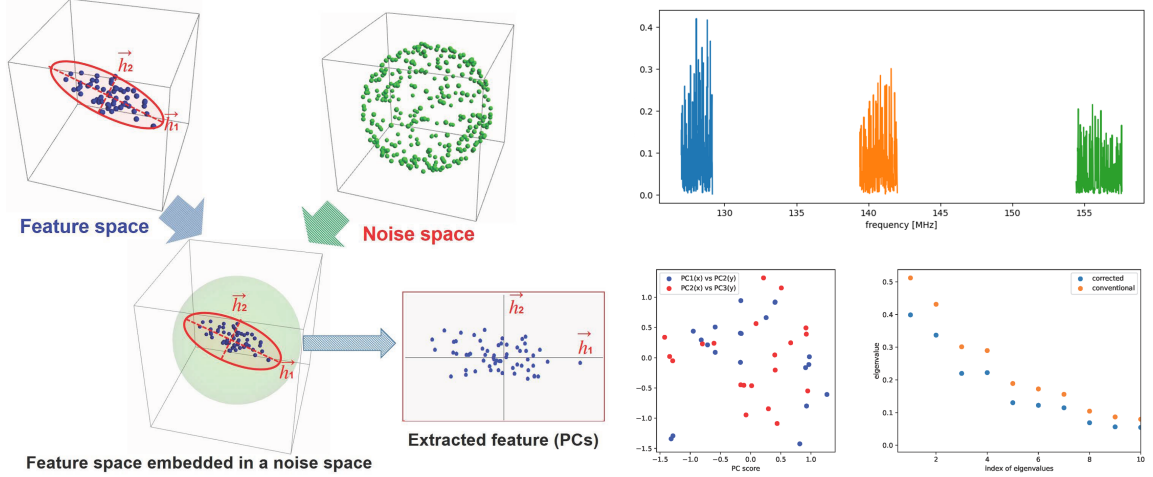


Figure 5: A schematic description of the high-dimensional principal component analysis (PCA) of the 21 cm forest. Left panel: Structure of the HDLSS data and their analysis. Even if we have significant important features, because of the existence of a huge noise sphere, they are completely embedded in the observed data. The high-dimensional PCA can subtract the noise sphere and enables to extract the desired features (Takeuchi et al., 2024). Right panel: Application to the simulated 21 cm forest. Top panel shows the optical depth of the realization of the 21 cm forest based on the Illustris TNG simulation. Bottom-left panel shows the result of the high-dimensional PCA applied to the simulated 21 cm forest data. Bottom-right panes compares the eigenvalues (contributions) obtained from the classical and high-dimensional PCA.

spectral data dimension. Thus, the observational dataset naturally falls into a regime where the sample size n is much smaller than the data dimension d .

Consider a d -dimensional parent population with covariance matrix $\tilde{\Sigma}_d > \tilde{O}$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$. From this population we draw n independent samples $\vec{x}_1, \dots, \vec{x}_n$ ($d > n$) and construct the data matrix $\tilde{X} = (\vec{x}_1, \dots, \vec{x}_n)$. The sample covariance matrix is $\tilde{S} = \frac{1}{n} \tilde{X} \tilde{X}^\top$, which becomes rank-deficient when $n \ll d$. In practice, principal component analysis (PCA)ⁿ is often used to extract dominant structures, but in this regime the leading eigenvalues can be strongly biased by high-dimensional noise.

A useful correction is obtained by considering the dual representation $\tilde{S}_D = \frac{1}{n} \tilde{X}^\top \tilde{X}$ and modifying the sample eigenvalues $\hat{\lambda}_i$ as

$$\check{\lambda}_i = \hat{\lambda}_i - \frac{1}{n-i} \left(\text{tr} \tilde{S}_D - \sum_{j=1}^i \hat{\lambda}_j \right) \quad (i = 1, \dots, n-1). \quad (12)$$

This estimator removes the high-dimensional noise contribution and provides a more stable characterization of physically meaningful variance components.

To illustrate this point, we constructed a mock sample of 21 cm forest spectra from the Illustris TNG simulation¹. Even when significant physical features are present, they can be masked by high-dimensional noise in conventional PCA. Figure 5 shows that the corrected analysis enables more

¹<https://www.tng-project.org/media/>.

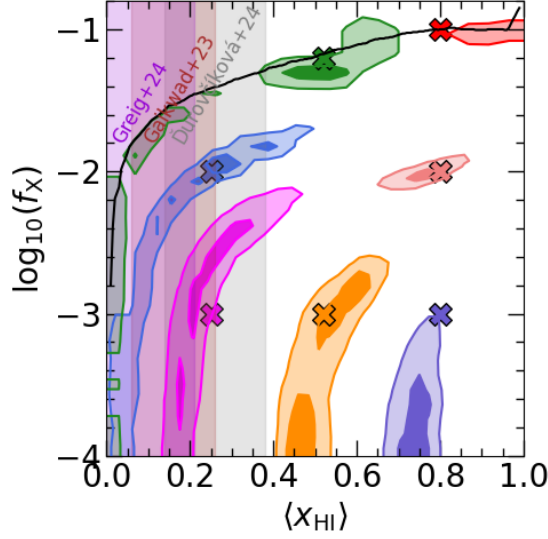


Figure 6: The posterior distributions of $\log_{10}f_X$ and $\langle x_{\text{HI}} \rangle$ from a Bayesian statistics analyses based on 1D power spectrum of 21 cm forest corresponding to 1 and 2σ confidence regions. True values of the $\log_{10}f_X$ and $\langle x_{\text{HI}} \rangle$ are indicated by the crosses with corresponding colour to the contours. This figure was taken from Šoltinský et al. (2025) and modified.

robust extraction of such features, which will be important for exploring the spatial distribution and evolution of the 21 cm forest in the SKA era.

3.2.4 Practical limitations: frequency-dependent systematics and validation

The sensitivity estimates discussed above can be optimistic if they neglect frequency-dependent instrumental and analysis systematics (e.g., residual bandpass structure, beam chromaticity and the associated sidelobe-induced mode-mixing of foreground emission, and RFI-related spectral structure(e.g. Burba et al., 2026)). Such effects are known to complicate low-frequency 21 cm analyses with precursor/pathfinder facilities (e.g., GMRT/uGMRT, MWA, LOFAR, and HERA; Paciga et al. 2013; Beardsley et al. 2016; Ewall-Wice et al. 2016; Patil et al. 2017; Mertens et al. 2020; Abdurashidova et al. 2022, 2023) . Mitigating these effects will require improved bandpass/beam characterization and end-to-end validation tests (e.g., injection and null tests). While the brightest targets may, in principle, enable high-spectral-resolution follow-up, we emphasize that in the near term (early commissioning and science verification) the most robust SKA results are expected to come from large-scale spectral statistics of the 21 cm forest. Such measurements can constrain the mean 21 cm optical depth and hence place lower limits on the IGM spin temperature, as well as provide coarse constraints on the reionization history (e.g., Thyagarajan, 2020).

4 Physics from the 21 cm forest

4.1 Astrophysics

The hot and ionized bubbles of the IGM have been explored extensively through the observations of the Ly α forest signal. On the other hand, the studies of the cold and neutral islands of the

intergalactic gas lag behind. Fortunately, the 21 cm forest signal has a potential to be one of the few probes of such regions in the IGM. Given its sensitivity to the kinetic and spin temperature of the IGM (see. Eq. 1 and 2) it has been suggested that this signal can be used to constrain the thermal state of the neutral IGM at $z \geq 6$ and the radiation that pre-heats these regions including the X-ray background radiation (Xu et al., 2009, 2011; Ewall-Wice et al., 2014; Shao et al., 2023). An example of such constraining power, which was already mentioned in Sec. 3.2.1, is presented in Fig. 6 for which mock observations of 1D power spectrum of the 21 cm forest from 10 PSO J0309+27-like quasars (Belladitta et al., 2020) each over 50 hr by the SKA1-low was assumed. However, if the neutral gas is pre-heated to $\gtrsim 100$ K, the 21 cm forest absorption is suppressed to $< 1\%$ leaving the signal likely undetectable (Ciardi et al., 2013; Šoltinský et al., 2021). Given this, a null-detection of the 21 cm forest signatures can be translated to lower limits on the X-ray driven pre-heating (Mack and Wyithe, 2012; Šoltinský et al., 2021, 2025). For instance, if the observations described above were to result in a null detection, lower limits on f_X defined by the solid black curve in Fig. 6 can be acquired. Such results would be important as the X-ray background radiation and the history of its sources are still largely unconstrained.

Recent JWST observations provide relevant context for this uncertainty. Deep JWST spectroscopy has revealed a higher-than-expected incidence of AGN signatures among galaxies at $z \gtrsim 4-7$, including faint broad-line (type-1) AGNs identified with NIRSpec (e.g., Harikane et al., 2023), and has also reported compelling evidence for accreting black holes in very high-redshift galaxies (e.g., Maiolino et al., 2024; Goulding et al., 2023). At the same time, many JWST-identified AGN candidates are generally undetected or only weakly constrained in X-rays, suggesting that the emerging high- z AGN population does not yet translate into a well-determined early X-ray background (e.g., Maiolino et al., 2025; Mazzolari et al., 2024). This motivates treating the amplitude and spectral properties of X-ray preheating (often summarized by f_X) as an uncertain but testable ingredient, and highlights the 21 cm forest as an independent probe of the neutral-IGM thermal history.

In addition, the ionization state of the IGM, and hence the reionization history, can be probed with the 21 cm forest too (Xu et al., 2009; Šoltinský et al., 2025). Such constraints would be complementary to numerous measurements based on the Ly α forest (e.g. Gaikwad et al., 2023; Ďurovčíková et al., 2024; Greig et al., 2024) and CMB (e.g. Pagano et al., 2020) observations as we can see in Fig. 6.

The 21 cm forest absorption features can be suppressed by not only the background X-ray radiation but also by the X-ray photons emitted by the background radio-loud quasar as mentioned in Sec. 2.2. The extent of such quasar near-zone in the 21 cm forest spectra is sensitive to the integrated lifetime of the background quasar as opposed to the Ly α forest near-zone size, which is dictated by the latest quasar accretion episode. Therefore, the 21 cm forest in the proximity of the background radio-loud quasars can be used to test models of supermassive black hole growth (Šoltinský et al., 2023).

Besides constraining the mechanics driving the quasars, Ewall-Wice et al. (2014) suggest that the 21 cm forest features in the tomographic 21 cm power spectrum provide information on the high- z radio-loud quasar population. More information on these sources can be found in Sec. 5.

4.2 Cosmology

The 21 cm forest is an essential tool for probing small-scale cosmological structures and dark matter properties. By analyzing the absorption lines produced by neutral hydrogen in minihalos, this technique enables exploration of early universe dynamics and the thermal history of the IGM. The 21 cm forest is particularly useful for studying small-scale structures that are beyond the reach of other methods such as the Lyman- α forest. One of its key applications is the investigation of ultralight dark matter (ULDM), including axion-like particles (ALPs), which suppress small-scale structures via quantum pressure effects. The number of 21 cm absorption lines can be used to constrain ULDM masses as low as 10^{-18} eV, far surpassing the capabilities of Lyman- α forest observations (Shimabukuro et al., 2020a). Additionally, scenarios involving axion-like particles in post-inflationary Peccei–Quinn symmetry breaking can also be probed, in which isocurvature perturbations enhance small-scale structures and increase the number of 21 cm absorption lines. This allows for the detection of ALP masses ranging from 10^{-18} eV to 10^{-6} eV, depending on the temperature dependence of the axion mass (Shimabukuro et al., 2020b).

Beyond dark matter, the 21 cm forest also provides insights into astrophysical and dynamical processes that affect structure formation. The relative velocity between dark matter and baryons, a result of different evolution rates before photon decoupling, suppresses the growth of minihalos. This suppression reduces the number of 21 cm absorption lines, and its scale dependence thereby makes it an important observable for testing early structure formation models (Shimabukuro et al., 2023). Primordial black holes (PBHs) introduce shot noise isocurvature modes, which enhance the formation of minihalos and increase the number of absorption lines. However, PBHs with masses larger than $10M_{\odot}$ can heat the IGM via accretion, counteracting this enhancement. This complex interplay allows the 21 cm forest to provide constraints on PBH abundances, complementing other observational techniques (Villanueva-Domingo and Ichiki, 2023). Furthermore, the 21 cm forest can also probe the presence of long-lived scalar field configurations, such as oscillons arising from ultralight axion-like fields, which generate additional small-scale fluctuations and enhance the number of 21 cm absorption lines. By analyzing the statistical distribution of these lines, the 21 cm forest offers a way to detect non-trivial dark matter structures and distinguish them from standard cold dark matter models (Kawasaki et al., 2021).

4.3 Exotic physics

In addition to astrophysical sources, exotic energy-injection processes—including dark matter annihilation (Cang et al., 2025; Nishizawa et al., 2025), dark matter decay (Sun et al., 2025; Facchinetti et al., 2024), Hawking radiation (Cang et al., 2022), and PBH accretion (Villanueva-Domingo and Ichiki, 2023)—can inject energetic particles into the IGM.

Electromagnetic cascades of these particles can heat and ionize the IGM, raising its temperature and ionization fraction. As illustrated in Vasiliev and Shchekinov (2013); Villanueva-Domingo and Ichiki (2023), additional heating and ionization from these processes tend to suppress both the optical depth and the 21 cm absorption signal.

While energy injection from astrophysical sources and decaying dark matter generally scales with the dark matter density ρ_c (Sun et al., 2025), the annihilation rate scales as ρ_c^2 and can be significantly

enhanced by structure formation at redshifts targeted by the 21 cm forest (Cang et al., 2025). Machine learning–based analyses have recently highlighted the potential of convolutional neural networks to constrain such annihilation scenarios directly from simulated 21 cm signals (Nishizawa et al., 2025). Energy injection from PBHs may receive characteristic boosts at relevant redshifts, either through violent explosions of light PBHs ($\lesssim 10^{15}$ g) reaching the end of their lives via Hawking evaporation (Cang et al., 2022), or through amplified accretion luminosity of massive PBHs (with initial mass $\gtrsim 10 M_{\odot}$) that undergo substantial mass growth via accretion (Zhang et al., 2025). These characteristic boosts and evolutionary features can enhance their impact on the IGM and potentially help break degeneracies with background astrophysical processes. Complementary to these approaches, studies of small-scale dark matter clumping (Sikder et al., 2025) show that the enhanced substructure can further amplify the annihilation signal, leaving distinctive imprints on the large-scale 21 cm power spectrum.

Conversely, interactions between dark matter and baryons can transfer heat from the IGM to the dark matter, thereby lowering the IGM temperature (Muñoz and Loeb, 2018; Barkana, 2018; Flitter and Kovetz, 2024). Such dark matter–induced cooling has been proposed to explain the unusually deep global 21 cm signal reported by EDGES (Bowman et al., 2018). In the context of the 21 cm forest, this cooling is expected to increase the optical depth and enhance the absorption features.

4.4 Parameter inference

In order to effectively utilize the 21 cm forest for investigating astrophysics, cosmology, and exotic physics, parameter estimation is a crucial step. In this section, we explore different parameter estimation techniques to accurately analyze the 21 cm forest signal, particularly in complex physical environments. These methods include Fisher matrix analysis, Bayesian inference, and deep learning, which help extract key information from the 21 cm signal (e.g. Acharya et al., 2026b).

To put different approaches on a clear footing, parameter inference methods for the 21 cm forest can be broadly grouped into (i) explicit-likelihood approaches, which require specifying an analytic likelihood (e.g., Fisher forecasts and MCMC with an assumed likelihood for binned bandpowers), and (ii) implicit-likelihood (likelihood-free) approaches, commonly referred to as simulation-based inference (SBI), which learn posteriors/likelihoods from forward simulations.

In this context, neural networks such as normalizing flows and U-Nets are not alternatives to SBI but are widely used building blocks within SBI frameworks: normalizing flows for density estimation, and U-Nets for data compression and feature extraction.

A common explicit-likelihood choice is a Gaussian likelihood for the (binned) power spectrum,

$$P(D|\theta) \propto \exp\left[-\frac{1}{2} \sum_{k,k'} (P_{\text{obs}}(k) - P_{\text{model}}(k|\theta)) \mathbf{C}_{kk'}^{-1} (P_{\text{obs}}(k') - P_{\text{model}}(k'|\theta))\right], \quad (13)$$

where the covariance $\mathbf{C}_{kk'}$ includes both thermal noise and sample (cosmic) variance. This Gaussian, bandpower-based likelihood is an approximation whose validity depends primarily on the effective number of independent modes/sightlines and the degree of non-Gaussianity in the signal and its covariance, rather than on S/N alone; in noise-dominated regimes the likelihood often

becomes more nearly Gaussian, whereas sparse sampling and non-Gaussian covariances can bias Gaussian-likelihood inference.

Studies in the 21 cm literature have shown that SBI can yield accurate posteriors with training sets smaller than the number of forward-model evaluations required for traditional MCMC (Saxena et al., 2023; Prelogović and Mesinger, 2023). For the 21 cm forest, however, generating the simulation training set can remain costly due to the required dynamic range and cross-scale modelling. This motivates amortized pipelines that combine SBI with fast emulators for the forward model.

As an example of amortized inference using neural density estimators within SBI frameworks, Sun et al. (2025) proposed a method that combines generative normalizing flows (GNF) with inference normalizing flows (INF) for data augmentation and parameter inference. This method first utilizes GNF to generate high-fidelity 1D power spectra based on a small number of samples, ensuring that the coefficient of determination (R^2) between the generated results corresponding to each parameter and the simulation results exceeds $R^2 > 0.99$. Subsequently, the observed or fiducial model's $P(k)$ is mapped to the posterior distribution through a bijection transformation by an INF:

$$Q(\boldsymbol{\theta} | P) = \pi[f_P^{-1}(\boldsymbol{\theta})] |\det J_{f_P^{-1}}|, \quad (14)$$

achieving millisecond-level parameter inference.

As shown in Figure 7, under low heating ($T_K \approx 60$ K), the posterior matches the Fisher ellipse. However, in strongly heated, non-Gaussian regimes, the Fisher matrix fails to provide accurate errors, while the normalizing flow approach reliably recovers the true m_{WDM} and T_K . This provides an efficient and reliable analysis pipeline for the ongoing SKA 21 cm forest observations. Deep learning enhances computational efficiency, overcoming the resource demands of simulation-based methods, while maintaining excellent performance in non-Gaussian environments for robust parameter inference.

Another likelihood-free inference approach employs a U-Net architecture to denoise the 21 cm forest flux spectrum, followed by XGBoost regression for parameter inference. Patil et al. (2026) have shown that such parameter inference analysis based on the 1D power spectra acquired from a U-Net denoised 21 cm forest outperforms traditional Bayesian methods. Moreover, omitting the power spectrum entirely and using latent features of the 21 cm forest flux spectrum yields even tighter constraints on the IGM properties and/or alleviates the requirements on the integration time of the observations.

Although deep learning and likelihood-free inference methods provide powerful tools, challenges remain in areas such as model interpretability, handling complex astrophysical signals, and reducing computational resource demands. Future research will focus on refining these techniques to improve parameter-estimation accuracy and to develop efficient, reliable analysis pipelines for ongoing SKA 21 cm forest observations.

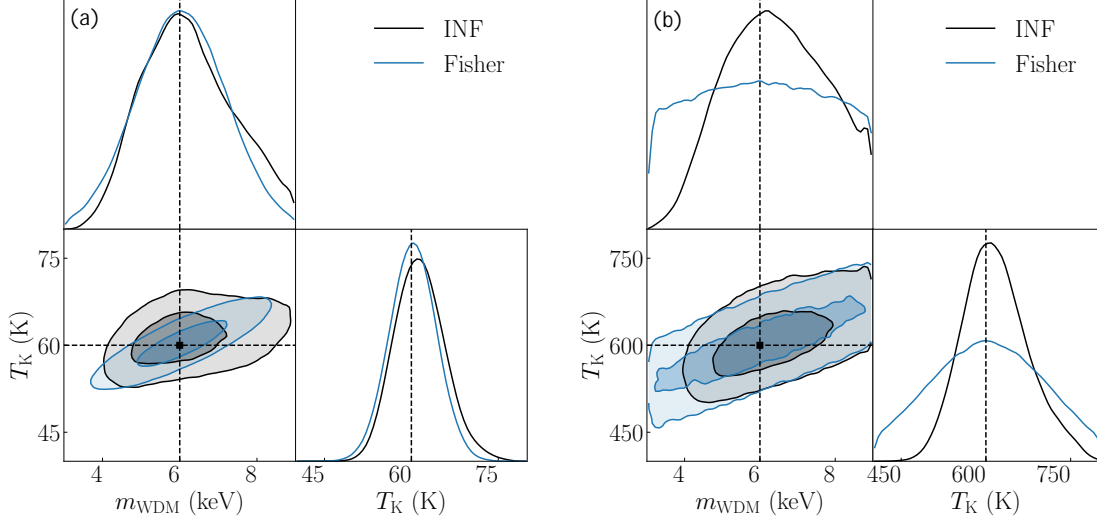


Figure 7: The posterior distributions for m_{WDM} and T_K with the inference normalizing flow and Fisher matrix under different heating levels. The figure is taken from Sun et al. (2025).

5 High-redshift radio sources

5.1 General high- z sources

5.1.1 Fast Radio Bursts as an independent probe

Fast Radio Bursts (FRBs) are highly energetic transients of cosmic origin that are widely distributed across the whole sky (e.g. Caleb et al., 2026; Curtin et al., 2026). The FRB continuum signal experiences dispersion while passing through the intervening ionized medium, including both Galactic and extragalactic components, which is quantified as the dispersion measure (DM). Although the origin of these events is still mysterious, the DM measurements are accurate enough to quantify the amount of plasma they pass through.

During the epoch of reionization (EoR), when the intergalactic medium (IGM) transitioned from a neutral to an ionized state, the DM measurements of FRBs could serve as a potential alternative probe of the ionized IGM, complementary to the 21 cm forest that traces the neutral hydrogen component (Fialkov and Loeb, 2016; Yoshiura and Takahashi, 2018). Recent studies have found that it is possible to constrain the physical state of the IGM (characterized by the mean IGM temperature, the midpoint and width of reionization, and Thomson scattering optical depth) with only $\sim 100 - 1000$ FRBs observed during EoR (Pagano and Fronenberg, 2021; Heimersheim et al., 2022; Maity, 2024). Furthermore, for a given cosmology, FRBs have the potential to constrain the

astrophysics of reionization (Maity, 2024; Shaw et al., 2024).

In addition, it has been demonstrated that the sky-averaged DM and its angular variance can potentially distinguish between different reionization histories, as they are directly connected to the typical sizes of ionized bubbles. We require ~ 1000 FRBs distributed over the entire EoR window, under ideal conditions, to place constraints on the typical bubble sizes using the variances (Shaw et al., 2024), and reionization topologies can be further probed with an order of magnitude higher number of FRBs (Ziegler et al., 2024).

However, the question remains regarding the detection prospects of high-redshift FRBs ($z \geq 5$). At present, highly magnetized and rapidly rotating magnetars are considered the most promising progenitors of FRBs, which are expected to arise from the deaths of massive Pop III stars. Given the increasing evidence for these massive stars during the EoR and the improved sensitivity of the upcoming SKA telescopes, there is a strong potential to detect a large number of FRBs during the EoR (Fialkov and Loeb, 2017; Hashimoto et al., 2020; Gupta et al., 2025), which would provide an ionized-gas counterpart to the neutral-gas information revealed by the 21 cm forest, together offering a more complete picture of the early Universe.

5.1.2 Radio galaxies

In 2018, the EDGES experiment reported the first detection of the global 21 cm differential brightness temperature (Bowman et al., 2018), revealing an absorption trough at redshift $z \sim 17$ with a depth of $\delta T_b \approx -500$ mK. This signal is roughly twice as deep as that predicted by standard models, which assume a radio background dominated by the CMB. Early radio-loud galaxies capable of producing an excess radio background comparable to the CMB could potentially explain such a deep signal (Feng and Holder, 2018; Reis et al., 2020; Fialkov and Barkana, 2019; Mirocha and Furlanetto, 2019), and consequently leave detectable imprints on 21 cm forest signals.

However, for typical atomic-cooling galaxies (ACGs) to achieve the depth observed by EDGES, the required radio background strength would exceed the current ARCADE2 upper limits (Fixsen et al., 2011) by orders of magnitude (Mirocha and Furlanetto, 2019; Reis et al., 2020). Furthermore, the corresponding star-formation rate density would also surpass the limits inferred from UV luminosity functions by nearly two orders of magnitude (Mirocha and Furlanetto, 2019).

In contrast, molecular-cooling galaxies (MCGs), which reside in minihalos with masses $\lesssim 10^6 M_\odot$, as opposed to ACGs that inhabit halos with masses $\gtrsim 10^8 M_\odot$ (Muñoz et al., 2022), are more susceptible to feedback from Lyman–Werner radiation. This radiation photodissociates H_2 molecules, suppressing further star formation and thereby causing the radio emission from MCGs to decline naturally once a Lyman–Werner background is established. As a result, MCGs are considered a plausible candidate population capable of generating a radio background strong enough to explain the EDGES depth without violating ARCADE2 constraints (Cang et al., 2025).

Nevertheless, several concerns have been raised regarding the original EDGES analysis of sky-temperature data (Hills et al., 2018; Sims and Pober, 2020; Cang et al., 2025), including the modeling of foregrounds, calibration residuals, and the parameterization of the cosmic 21 cm signal. Most recently, Cang et al. (2025) re-analyzed the EDGES sky-temperature data, jointly

forward-modeling the foregrounds, calibration residuals, and the cosmic signal expected from MCG radio galaxies. Their analysis found that an excess radio background is in fact decisively disfavored by the data. Moreover, a subsequent measurement by SARAS3 (Bevins et al., 2022) found no evidence for the 21 cm signal initially reported by EDGES. Therefore, the existence of a strong 21 cm global signal—and consequently of early radio galaxies responsible for it—requires further experimental verification and more careful characterization of foregrounds and instrumental systematics.

5.2 Background sources for 21 cm forest observations

We now turn to bright radio-loud quasars, which are the primary background sources for 21 cm forest observations. Among the various types of high-redshift sources discussed above, bright radio-loud quasars are of particular importance for 21 cm forest observations. Detecting the 21 cm forest requires such luminous background radio sources, since the absorption features are imprinted on their spectra as the signal passes through intervening neutral hydrogen. The limited number of suitable sources has long been one of the main challenges for practical 21 cm forest detection.

The number of radio-loud quasars identified at $z \geq 5.5$ (i.e., before the completion of reionization) has recently increased. Specifically, this number has more than quadrupled since the previous edition of the SKA Science Book (2015), now exceeding 30 known sources (e.g., Fan et al., 2001; McGreer et al., 2006; Jiang et al., 2009; Willott et al., 2010; Zeimann et al., 2011; Bañados et al., 2015, 2021, 2023; Wang et al., 2019; Belladitta et al., 2020; Liu et al., 2021; Ighina et al., 2021, 2023, 2024, 2025; Endsley et al., 2023; Gloudemans et al., 2022, 2023; Wolf et al., 2024). In addition, both the redshift and brightness ranges of radio-loud quasars have expanded. For example, the most distant radio-loud quasar has been detected at $z \sim 7$ (Bañados et al., 2024), while the brightest one shows a flux density of 110.6 ± 13.8 mJy at 150 MHz (Bañados et al., 2018). Alternative potential background sources include gamma-ray burst afterglows (Ioka and Mészáros, 2005; Ciardi et al., 2015) and radio galaxies (Drouart et al., 2020). With these newly discovered background radio sources, the prospects for detecting the 21 cm forest have significantly improved.

The SKA itself also holds great promise for detecting radio-loud quasars. Recently, the Australian Square Kilometre Array Pathfinder (ASKAP) discovered six new radio-loud quasars at $z > 5.5$ through the combination of the Rapid ASKAP Continuum Survey (RACS) and deep, wide-area optical/near-infrared surveys (Ighina et al., 2025), demonstrating the SKA’s powerful potential in this field. Using a physically motivated model, Niu et al. (2025) predicted the observational prospects of SKA-Low for high- z radio-loud quasars. As shown in Figure 8, their results indicate that a one-year sky survey with SKA-Low could detect approximately 20 radio-loud quasars at $z \sim 9$, each bright enough that, in principle, high-spectral-resolution follow-up could resolve individual 21 cm absorption lines. These quasars will serve as crucial background sources for future studies of the 21 cm forest.

It is worth noting that the discovery of new quasars relies on spectroscopic observations in the optical and near-infrared bands to determine their redshifts. For quasars in the epoch of reionization, the Lyman- α break in their spectra is redshifted into the infrared due to cosmic expansion. Meanwhile, spectral features in the near-infrared are crucial for efficiently identifying quasars among the

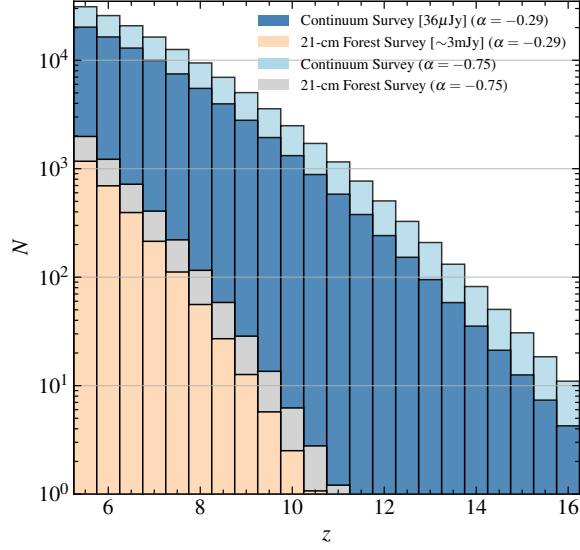


Figure 8: Predicted redshift distribution of radio-loud quasars observable with one year of SKA-Low observations. The “21 cm Forest Survey” refers to high-spectral-resolution observations suitable for 21 cm forest analysis. The figure is taken from Niu et al. (2025).

numerous point sources detected in radio observations (e.g., Liu et al., 2021; Yang et al., 2023; Euclid Collaboration et al., 2024). Therefore, future observations of high- z radio-loud quasars will require a combination of SKA and advanced infrared telescopes. Niu et al. (2025) assessed the observational prospects of Euclid (Euclid Collaboration et al., 2022) and the Nancy Grace Roman Space Telescope (Spergel et al., 2015) for radio-loud quasars at $z \sim 9$. Their results show that both space-based infrared surveys can detect the majority of radio-loud quasars at $z \sim 9$ that SKA-Low could observe with resolved individual 21 cm lines. The synergy between SKA and space-based infrared telescopes will provide a crucial foundation for identifying radio-loud quasars that serve as background sources for 21 cm forest studies at the highest redshifts.

We may also note that JWST has detected more high- z AGNs than previously expected in the pre-JWST era (Harikane et al., 2023), and black holes with higher masses compared to those predicted by the local $M_{\bullet}-M_{*}$ relation (Maiolino et al., 2024). These black holes have masses of $\sim 10^6-10^8 M_{\odot}$, indicating that SMBH seeds formed efficiently and grew rapidly during the Cosmic Dawn and EoR (Goulding et al., 2023; Jeon et al., 2025). Although attempts to detect the radio emission from JWST-detected AGNs have not revealed strong signals (Mazzolari et al., 2024; Gloudemans et al., 2025), this does not rule out the possibility that some of them are radio-loud. Even if only a small fraction of these high- z black holes enter a radio-loud phase, they could serve as potential background sources for 21 cm forest observations.

6 Summary & Discussion

The 21 cm forest, seen as neutral hydrogen (HI) absorption lines in the spectra of distant radio sources, is a unique probe of small-scale cosmic structure and the thermal history of the early universe. Recent progress has made both theoretical and observational studies increasingly feasible.

Its modeling includes analytic, semi-analytic, hydrodynamical, and semi-numerical approaches: analytic methods efficiently explore large parameter spaces, while numerical simulations capture the detailed physics of the intergalactic medium (IGM). In particular, hydrodynamic simulations with radiative transfer allow sophisticated modeling of absorption features and their astrophysical dependence. Statistical tools such as the 1D power spectrum and WST are particularly valuable for data analysis. The power spectrum improves signal-to-noise and helps separate astrophysical and dark matter effects, whereas WST captures non-Gaussian structures, localized fluctuations, and cross-scale interactions. The 21 cm forest can constrain the thermal and ionization state of the IGM, supermassive black hole growth, reionization history, and cosmological physics such as ultralight dark matter, primordial black holes, and baryon-dark matter relative velocities. The SKA will greatly enhance the sensitivity of these observations, enabling major progress in early-universe studies.

We should prioritize refining observational strategies and analysis pipelines to optimize SKA data for 21 cm forest measurements. Because the forest signal appears as an extremely weak, narrow-band absorption feature on bright continuum spectra, the main near-term limitations are likely to come from spectral systematics rather than raw sensitivity alone. Key priorities therefore include controlling frequency-dependent instrumental effects—such as residual bandpass/gain ripples, beam chromaticity, polarization leakage, ionospheric variability, and spectral structure from imperfect RFI flagging—through improved calibration, robust quality assessment, and end-to-end validation, including signal-injection tests, null tests, and independent pipeline cross-checks. Accurate modeling of the intrinsic spectra of background radio sources is also essential to distinguish genuine absorption from calibration-induced structure. Given these limitations, AA* and science-verification observations will likely first provide robust constraints from large-scale spectral statistics, such as the mean optical depth, the IGM spin temperature, and the ionization history, before enabling detailed individual-line studies and tighter model discrimination.

In parallel, a statistical sample of high-redshift FRBs offers an independent probe of the IGM evolution and morphology, complementing 21 cm observations. With the sensitivity of the SKA, detecting such events during the reionization epoch may become feasible. This will require close integration with multi-wavelength observations, especially from JWST, Euclid, and the Nancy Grace Roman Space Telescope. Combining SKA radio data with deep optical/IR imaging and spectroscopy will allow robust identification, redshift measurement, and physical characterization of high-redshift radio-loud quasars, galaxies, and FRB hosts. Such synergy is crucial for constructing reliable background-source samples for 21 cm forest studies and for cross-validating interpretations of absorption features. Optical/IR data constrain the ionization state of the IGM and reionization history, while the 21 cm forest probes small-scale structure and thermal conditions. Future work should focus on coordinated surveys, joint analysis pipelines, and statistical frameworks for multi-wavelength cross-correlation.

Finally, expanding parameter-estimation methods with Bayesian inference and deep-learning-based likelihood-free approaches will greatly improve both the accuracy and efficiency of data analysis. Traditional methods such as Fisher analysis and MCMC remain powerful, but are limited by Gaussian assumptions and computational cost. Future work should therefore emphasize flexible Bayesian

frameworks that can handle non-Gaussian, multi-modal, and otherwise complex likelihoods, especially in regimes of strong heating or low signal-to-noise. At the same time, deep-learning approaches enable rapid posterior estimation in high-dimensional spaces without requiring explicit likelihoods. Integrating these methods into analysis pipelines will allow more robust extraction of astrophysical and cosmological information from SKA data, helping to fully realize the potential of the 21 cm forest as a probe of the early Universe.

7 Author contributions

H.S. and Y.X. coordinated this chapter. H.S. contributed to Secs. 1, 3.1, 3.2, 4.2, 4.4, and 6. T.Š. contributed to Secs. 1, 2.1, 2.2, 3.1, 3.2, 4.1, 4.4, and 5.2. Y.X. contributed to the abstract and Secs. 1, 2, 3.2, and 4.4. T.-Y.S. and Y.S. contributed to Secs. 3.2 and 4.4. Q.N. and B.Y. contributed to Sec. 5.2. X.Z. contributed to Secs. 3.2, 4.4, and 5.2. T.T.T. and K.Y. contributed to Sec. 3.2. A.K.S. and B.M. contributed to Sec. 5.1. J.C. contributed to Secs. 4.3 and 5.1.2. A.J.N. contributed to Sec. 4.3. B.C. contributed to Secs. 2.1 and 2.2.

Acknowledgement

HS is supported by the National SKA Program of China (No. 2020SKA0110401), the NSFC (Grant No. 12103044), and the Yunnan Provincial Key Laboratory of Survey Science (Project No. 202449CE340002). TŠ acknowledges support from the Istituto Nazionale di Astrofisica Osservatorio Astronomico di Trieste (INAF-OATs) through the Theory grant “Cosmological Investigation of the Cosmic Web” (C93C23006820005), and from the Istituto Nazionale di Fisica Nucleare (INFN) INDARK grant. YX acknowledges support from the National SKA Program of China (No. 2020SKA0110401) and the National Key R&D Program of China (No. 2022YFF0504300). BY acknowledges support from the National SKA Program of China (No. 2020SKA0110402) and the NSFC International (Regional) Cooperation and Exchange Project (No. 12361141814). TTT acknowledges JSPS Grants-in-Aid for Scientific Research (21H01128 and 24H00247), and partial support from the ISM Collaboration Funding “Machine-Learning-Based Cosmogony: From Structure Formation to Galaxy Evolution”. KY is supported by JSPS KAKENHI (JP21H01079, JP25H00625) and by MEXT’s “Program for Promoting Researches on the Supercomputer Fugaku” (Structure and Evolution of the Universe Unraveled by Fusion of Simulation and AI). XZ is supported by the National SKA Program of China (2022SKA0110200, 2022SKA0110203), the National Natural Science Foundation of China (12473091, 12473001, 12533001), and the National 111 Project (B16009). AJN is supported by JSPS Grant-in-Aid for Transformative Research Areas (JP25H01551) and JSPS International Leading Research (22K21349). AKS acknowledges support from the National Science Foundation (Grant No. 2206602).

References

- Z. Abdurashidova et al. *The Astrophysical Journal*, 925(2):221, 2022. doi: 10.3847/1538-4357/ac1c78.
- Z. Abdurashidova et al. *The Astrophysical Journal*, 945(2):124, 2023. doi: 10.3847/1538-4357/acaf50.

- A. Acharya et al. In Advancing Astrophysics with the SKA – II (AASKAII). 2026a. arXiv search: Report number AASKAII/Acharya01.
- A. Acharya et al. In Advancing Astrophysics with the SKA – II (AASKAII). 2026b. arXiv search: Report number AASKAII/Acharya02.
- E. Bañados et al. Astrophys. J., 804(2):118, May 2015. doi: 10.1088/0004-637X/804/2/118.
- E. Bañados et al. Astrophys. J. Lett., 861(2):L14, July 2018. doi: 10.3847/2041-8213/aac511.
- E. Bañados et al. Astrophys. J., 909(1):80, Mar. 2021. doi: 10.3847/1538-4357/abe239.
- E. Bañados et al. Astrophys. J. Suppl. Ser., 265(1):29, Mar. 2023. doi: 10.3847/1538-4365/acb3c7.
- E. Bañados et al. Nature Astronomy, Dec. 2024. doi: 10.1038/s41550-024-02431-4.
- S. Bag et al. In Advancing Astrophysics with the SKA – II (AASKAII). 2026. arXiv search: Report number AASKAII/Bag01.
- R. Barkana. Nature, 555(7694):71–74, 2018. doi: 10.1038/nature25791.
- R. Barkana et al. In Advancing Astrophysics with the SKA – II (AASKAII). 2026. arXiv search: Report number AASKAII/Barkana01.
- A. P. Beardsley et al. The Astrophysical Journal, 833(1):102, 2016. doi: 10.3847/1538-4357/833/1/102.
- G. D. Becker et al. Mon. Not. R. Astron. Soc., 447(4):3402–3419, Mar. 2015. doi: 10.1093/mnras/stu2646.
- G. D. Becker et al. Mon. Not. R. Astron. Soc., 508(2):1853–1869, Dec. 2021. doi: 10.1093/mnras/stab2696.
- G. D. Becker, J. S. Bolton, Y. Zhu, and S. Hashemi. Mon. Not. R. Astron. Soc., 533(2):1525–1540, Sept. 2024. doi: 10.1093/mnras/stae1918.
- S. Belladitta et al. Astron. Astrophys., 635:L7, Mar. 2020. doi: 10.1051/0004-6361/201937395.
- H. T. J. Bevins et al. Nature Astron., 6(12):1473–1483, 2022. doi: 10.1038/s41550-022-01825-6.
- A. Bhagwat, B. Ciardi, E. Zackrisson, and J. Schaye. Mon. Not. R. Astron. Soc., 517(2):2331–2342, Dec. 2022. doi: 10.1093/mnras/stac2663.
- S. E. I. Bosman et al. Mon. Not. R. Astron. Soc., 514(1):55–76, July 2022. doi: 10.1093/mnras/stac1046.
- J. D. Bowman et al. Nature, 555(7694):67–70, 2018. doi: 10.1038/nature25792.
- J. Burba et al. In Advancing Astrophysics with the SKA – II (AASKAII). 2026. arXiv search: Report number AASKAII/Burba01.
- C. Cain, A. D’Aloisio, N. Gangolli, and G. D. Becker. Astrophys. J. Lett., 917(2):L37, Aug. 2021. doi: 10.3847/2041-8213/ac1ace.
- M. Caleb et al. In Advancing Astrophysics with the SKA – II (AASKAII). 2026. arXiv search: Report number AASKAII/Caleb02.
- J. Cang, Y. Gao, and Y.-Z. Ma. JCAP, 03(03):012, 2022. doi: 10.1088/1475-7516/2022/03/012.
- J. Cang, Y. Gao, and Y.-Z. Ma. Phys. Rev. D, 112(10):103048, Nov. 2025. doi: 10.1103/69jk-m5xg.
- J. Cang et al. Astron. Astrophys., 698:A152, 2025. doi: 10.1051/0004-6361/202452982.
- C. L. Carilli, N. Y. Gnedin, and F. Owen. Astrophys. J., 577(1):22–30, Sept. 2002. doi: 10.1086/342179.
- T. K. Chan et al. Mon. Not. R. Astron. Soc., 528(2):1296–1326, Feb. 2024. doi: 10.1093/mnras/stae114.
- J. Chardin, E. Puchwein, and M. G. Haehnelt. Mon. Not. R. Astron. Soc., 465(3):3429–3445, Mar.

2017. doi: 10.1093/mnras/stw2943.
- T. R. Choudhury, A. Paranjape, and S. E. I. Bosman. *Mon. Not. R. Astron. Soc.*, 501(4):5782–5796, Mar. 2021. doi: 10.1093/mnras/stab045.
- H. M. Christenson et al. *Astrophys. J.*, 923(1):87, Dec. 2021. doi: 10.3847/1538-4357/ac2a34.
- B. Ciardi et al. *Mon. Not. R. Astron. Soc.*, 366(2):689–696, Feb. 2006. doi: 10.1111/j.1365-2966.2005.09908.x.
- B. Ciardi, J. S. Bolton, A. Maselli, and L. Graziani. *Mon. Not. R. Astron. Soc.*, 423(1):558–574, June 2012. doi: 10.1111/j.1365-2966.2012.20902.x.
- B. Ciardi et al. *Mon. Not. R. Astron. Soc.*, 428(2):1755–1765, Jan. 2013. doi: 10.1093/mnras/sts156.
- B. Ciardi et al. *Mon. Not. R. Astron. Soc.*, 453(1):101–105, Oct. 2015. doi: 10.1093/mnras/stv1640.
- A. P. Curtin et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Curtin01.
- A. D’Aloisio, M. McQuinn, and H. Trac. *Astrophys. J. Lett.*, 813(2):L38, Nov. 2015. doi: 10.1088/2041-8205/813/2/L38.
- F. B. Davies and S. R. Furlanetto. *Mon. Not. R. Astron. Soc.*, 460(2):1328–1339, Aug. 2016. doi: 10.1093/mnras/stw931.
- E. de Lera Acedo et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/deLeraAcedo01.
- G. Drouart et al. *Publ. Astron. Soc. Aust.*, 37:e026, July 2020. doi: 10.1017/pasa.2020.6.
- R. Endsley et al. *Mon. Not. R. Astron. Soc.*, 520(3):4609–4620, Apr. 2023. doi: 10.1093/mnras/stad266.
- Euclid Collaboration et al. *Astron. Astrophys.*, 662:A112, June 2022. doi: 10.1051/0004-6361/202141938.
- Euclid Collaboration et al. *Astron. Astrophys.*, 691:A1, Nov. 2024. doi: 10.1051/0004-6361/202450446.
- A. Ewall-Wice, J. S. Dillon, A. Mesinger, and J. Hewitt. *Mon. Not. R. Astron. Soc.*, 441(3):2476–2496, July 2014. doi: 10.1093/mnras/stu666.
- A. Ewall-Wice et al. *Monthly Notices of the Royal Astronomical Society*, 460(4):4320–4347, 2016. doi: 10.1093/mnras/stw1022.
- G. Facchinetti, L. Lopez-Honorez, Y. Qin, and A. Mesinger. *JCAP*, 01:005, 2024. doi: 10.1088/1475-7516/2024/01/005.
- X. Fan et al. *Astron. J.*, 122(6):2833–2849, Dec. 2001. doi: 10.1086/324111.
- C. Feng and G. Holder. *Astrophys. J. Lett.*, 858(2):L17, May 2018. doi: 10.3847/2041-8213/aac0fe.
- A. Fialkov and R. Barkana. *Mon. Not. R. Astron. Soc.*, 486(2):1763–1773, June 2019. doi: 10.1093/mnras/stz873.
- A. Fialkov and A. Loeb. *J. Cosmol. Astropart. Phys.*, 2016(5):004, May 2016. doi: 10.1088/1475-7516/2016/05/004.
- A. Fialkov and A. Loeb. *Astrophys. J. Lett.*, 846(2):L27, Sept. 2017. doi: 10.3847/2041-8213/aa8905.
- G. B. Field. *Proceedings of the IRE*, 46:240–250, Jan. 1958. doi: 10.1109/JRPROC.1958.286741.
- D. J. Fixsen et al. *Astrophys. J.*, 734(1):5, June 2011. doi: 10.1088/0004-637X/734/1/5.

- J. Flitter and E. D. Kovetz. *Phys. Rev. D*, 109(4):043512, 2024. doi: 10.1103/PhysRevD.109.043512.
- S. R. Furlanetto. *Mon. Not. R. Astron. Soc.*, 370(4):1867, Jun 2006. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.10603.x.
- S. R. Furlanetto. *Mon. Not. R. Astron. Soc.*, 371(2):867–878, Sept. 2006. doi: 10.1111/j.1365-2966.2006.10725.x.
- S. R. Furlanetto and A. Loeb. *Astrophys. J.*, 579:1–9, Nov. 2002. doi: 10.1086/342757.
- S. R. Furlanetto, S. P. Oh, and F. H. Briggs. *Phys. Rep.*, 433(4-6):181–301, Oct. 2006. doi: 10.1016/j.physrep.2006.08.002.
- P. Gaikwad et al. *Mon. Not. R. Astron. Soc.*, 494(4):5091–5109, June 2020. doi: 10.1093/mnras/staa907.
- P. Gaikwad et al. *Mon. Not. R. Astron. Soc.*, 525(3):4093–4120, Nov. 2023. doi: 10.1093/mnras/stad2566.
- A. J. Gloude-mans et al. *Astron. Astrophys.*, 668:A27, Dec. 2022. doi: 10.1051/0004-6361/202244763.
- A. J. Gloude-mans et al. *Astron. Astrophys.*, 678:A128, Oct. 2023. doi: 10.1051/0004-6361/202347582.
- A. J. Gloude-mans et al. *Astrophys. J.*, 986(2):130, June 2025. doi: 10.3847/1538-4357/adddb9.
- A. D. Goulding et al. *Astrophys. J. Lett.*, 955(1):L24, Sept. 2023. doi: 10.3847/2041-8213/acf7c5.
- B. Greig et al. *Mon. Not. R. Astron. Soc.*, 530(3):3208–3227, May 2024. doi: 10.1093/mnras/stae1080.
- O. Gupta, P. Beniamini, P. Kumar, and S. L. Finkelstein. *Astrophys. J.*, 986(1):100, June 2025. doi: 10.3847/1538-4357/add14c.
- Y. Harikane et al. *Astrophys. J.*, 959(1):39, Dec. 2023. doi: 10.3847/1538-4357/ad029e.
- T. Hashimoto et al. *Mon. Not. R. Astron. Soc.*, 497(4):4107–4116, Oct. 2020. doi: 10.1093/mnras/staa2238.
- S. Heimersheim, N. S. Sartorio, A. Fialkov, and D. R. Lorimer. *Astrophys. J.*, 933(1):57, July 2022. doi: 10.3847/1538-4357/ac70c9.
- R. Hills, G. Kulkarni, P. D. Meerburg, and E. Puchwein. *Nature*, 564(7736):E32–E34, 2018. doi: 10.1038/s41586-018-0796-5.
- L. Ighina et al. *Astron. Astrophys.*, 647:L11, Mar. 2021. doi: 10.1051/0004-6361/202140362.
- L. Ighina et al. *Mon. Not. R. Astron. Soc.*, 519(2):2060–2068, Feb. 2023. doi: 10.1093/mnras/stac3668.
- L. Ighina et al. *Astron. Astrophys.*, 692:A241, Dec. 2024. doi: 10.1051/0004-6361/202451376.
- L. Ighina et al. *Astron. Astrophys.*, 698:A158, June 2025. doi: 10.1051/0004-6361/202453650.
- K. Ioka and P. Mészáros. *Astrophys. J.*, 619(2):684–696, Feb. 2005. doi: 10.1086/426785.
- J. Jeon, V. Bromm, B. Liu, and S. L. Finkelstein. *Astrophys. J.*, 979(2):127, Feb. 2025. doi: 10.3847/1538-4357/ad9f3a.
- L. Jiang et al. *Astron. J.*, 138(1):305–311, July 2009. doi: 10.1088/0004-6256/138/1/305.
- K. Kadota et al. *J. Cosmol. Astropart. Phys.*, 2023(3):017, Mar. 2023. doi: 10.1088/1475-7516/2023/03/017.
- K. Kakiichi et al. *arXiv e-prints*, art. arXiv:2503.07074, Mar. 2025. doi: 10.48550/arXiv.2503.07074.
- D. Kashino et al. *Astrophys. J.*, 888(1):6, Jan. 2020. doi: 10.3847/1538-4357/ab5a7d.

- M. Kawasaki, W. Nakano, H. Nakatsuka, and E. Sonomoto. *J. Cosmol. Astropart. Phys.*, 2021(4):019, Apr. 2021. doi: 10.1088/1475-7516/2021/04/019.
- L. C. Keating et al. *Mon. Not. R. Astron. Soc.*, 491(2):1736–1745, Jan. 2020. doi: 10.1093/mnras/stz3083.
- G. Kulkarni et al. *Mon. Not. R. Astron. Soc.*, 485(1):L24–L28, May 2019. doi: 10.1093/mnras/slz025.
- A. Lidz et al. *Astrophys. J.*, 670(1):39–59, Nov. 2007. doi: 10.1086/521974.
- Y. Liu et al. *Astrophys. J.*, 908(2):124, Feb. 2021. doi: 10.3847/1538-4357/abd3a8.
- K. J. Mack and J. S. B. Wyithe. *Mon. Not. R. Astron. Soc.*, 425(4):2988–3001, Oct. 2012. doi: 10.1111/j.1365-2966.2012.21561.x.
- R. Maiolino et al. *Nature*, 627(8002):59–63, Mar. 2024. doi: 10.1038/s41586-024-07052-5.
- R. Maiolino et al. *Mon. Not. R. Astron. Soc.*, 538(3):1921–1943, Apr. 2025. doi: 10.1093/mnras/staf359.
- B. Maity. *Astron. Astrophys.*, 689:A340, Sept. 2024. doi: 10.1051/0004-6361/202451160.
- G. Mazzolari et al. *arXiv e-prints*, art. arXiv:2412.04224, Dec. 2024. doi: 10.48550/arXiv.2412.04224.
- I. D. McGreer, R. H. Becker, D. J. Helfand, and R. L. White. *Astrophys. J.*, 652(1):157–162, Nov. 2006. doi: 10.1086/507767.
- M. McQuinn. *Annual Review of Astronomy and Astrophysics*, 54:313–362, 2016. doi: 10.1146/annurev-astro-082214-122355.
- A. Meiksin. *Mon. Not. R. Astron. Soc.*, 417(2):1480–1509, Oct. 2011. doi: 10.1111/j.1365-2966.2011.19362.x.
- A. Meiksin. *Mon. Not. R. Astron. Soc.*, 491(4):4884–4893, Feb. 2020. doi: 10.1093/mnras/stz3395.
- F. G. Mertens et al. *Monthly Notices of the Royal Astronomical Society*, 493(2):1662–1685, 2020. doi: 10.1093/mnras/staa327.
- A. Mesinger. *Mon. Not. R. Astron. Soc.*, 407(2):1328–1337, Sept. 2010. doi: 10.1111/j.1365-2966.2010.16995.x.
- A. Mesinger, S. Furlanetto, and R. Cen. *Mon. Not. R. Astron. Soc.*, 411(2):955–972, Feb. 2011. doi: 10.1111/j.1365-2966.2010.17731.x.
- J. Mirocha and S. R. Furlanetto. *Mon. Not. R. Astron. Soc.*, 483(2):1980–1992, Feb. 2019. doi: 10.1093/mnras/sty3260.
- J. B. Muñoz and A. Loeb. *Nature*, 557(7707):684, 2018. doi: 10.1038/s41586-018-0151-x.
- J. B. Muñoz et al. *Mon. Not. Roy. Astron. Soc.*, 511(3):3657–3681, 2022. doi: 10.1093/mnras/stac185.
- S. Murray et al. *The Journal of Open Source Software*, 5(54):2582, Oct. 2020. doi: 10.21105/joss.02582.
- M. Nakane et al. *Astrophys. J.*, 967(1):28, May 2024. doi: 10.3847/1538-4357/ad38c2.
- R. Nakatani, A. Fialkov, and N. Yoshida. *Astrophys. J.*, 905(2):151, Dec. 2020. doi: 10.3847/1538-4357/abc5b4.
- G. Naruse et al. *J. Cosmol. Astropart. Phys.*, 2024(10):091, Oct. 2024. doi: 10.1088/1475-7516/2024/10/091.
- F. Nasir and A. D’Aloisio. *Mon. Not. R. Astron. Soc.*, 494(3):3080–3094, Apr. 2020. doi: 10.1093/

- mnras/staa894.
- A. J. Nishizawa, P. K. Natwariya, and K. Kadota. *Phys. Rev. D*, 111(8):083546, Apr. 2025. doi: 10.1103/PhysRevD.111.083546.
- Q. Niu et al. *Astrophys. J.*, 978(2):145, Jan. 2025. doi: 10.3847/1538-4357/ad9b97.
- G. Paciga et al. *Monthly Notices of the Royal Astronomical Society*, 433(1):639–647, 2013. doi: 10.1093/mnras/stt753.
- L. Pagano et al. *Astron. Astrophys.*, 635:A99, Mar. 2020. doi: 10.1051/0004-6361/201936630.
- M. Pagano and H. Fronenberg. *Mon. Not. R. Astron. Soc.*, 505(2):2195–2206, Aug. 2021. doi: 10.1093/mnras/stab1438.
- H. Park et al. *Astrophys. J.*, 831(1):86, Nov. 2016. doi: 10.3847/0004-637X/831/1/86.
- A. H. Patil et al. *The Astrophysical Journal*, 838(1):65, 2017. doi: 10.3847/1538-4357/aa63e7.
- S. K. Patil, T. Šoltinský, S. Maitra, and G. Kulkarni. *Mon. Not. R. Astron. Soc.*, Feb. 2026. doi: 10.1093/mnras/stag236.
- A. H. Pawlik et al. *Mon. Not. R. Astron. Soc.*, 466(1):960–973, Apr. 2017. doi: 10.1093/mnras/stw2869.
- D. Prelogović and A. Mesinger. *Mon. Not. Roy. Astron. Soc.*, 524(3):4239–4255, 2023. doi: 10.1093/mnras/stad2027.
- E. Puchwein et al. *Mon. Not. R. Astron. Soc.*, 519(4):6162–6183, Mar. 2023. doi: 10.1093/mnras/stac3761.
- Y. Qin, A. Mesinger, S. E. I. Bosman, and M. Viel. *Mon. Not. R. Astron. Soc.*, 506(2):2390–2407, Sept. 2021. doi: 10.1093/mnras/stab1833.
- I. Reis, A. Fialkov, and R. Barkana. *Mon. Not. R. Astron. Soc.*, 499(4):5993–6008, Dec. 2020. doi: 10.1093/mnras/staa3091.
- F. Sawyer et al. *Mon. Not. R. Astron. Soc.*, 540(3):2238–2252, July 2025. doi: 10.1093/mnras/staf858.
- A. Saxena et al. *Mon. Not. Roy. Astron. Soc.*, 525(4):6097–6111, 2023. doi: 10.1093/mnras/stad2659.
- B. Semelin. *Mon. Not. R. Astron. Soc.*, 455(1):962–973, Jan. 2016. doi: 10.1093/mnras/stv2312.
- Y. Shao et al. *Nature Astronomy*, 7:1116–1126, Sept. 2023. doi: 10.1038/s41550-023-02024-7.
- Y. Shao, T.-Y. Sun, M.-L. Zhao, and X. Zhang. *Phys. Rev. D*, 112(6):063513, Sept. 2025. doi: 10.1103/vpd1-1kyj.
- A. K. Shaw et al. *arXiv e-prints*, art. arXiv:2409.03255, Sept. 2024. doi: 10.48550/arXiv.2409.03255.
- H. Shimabukuro. *Phys. Rev. D*, 113(8):083525, Apr. 2026. doi: 10.1103/bl2w-crry.
- H. Shimabukuro, K. Ichiki, S. Inoue, and S. Yokoyama. *Phys. Rev. D*, 90(8):083003, Oct. 2014. doi: 10.1103/PhysRevD.90.083003.
- H. Shimabukuro, K. Ichiki, and K. Kadota. *Phys. Rev. D*, 101(4):043516, Feb. 2020a. doi: 10.1103/PhysRevD.101.043516.
- H. Shimabukuro, K. Ichiki, and K. Kadota. *Phys. Rev. D*, 102(2):023522, July 2020b. doi: 10.1103/PhysRevD.102.023522.
- H. Shimabukuro, K. Ichiki, and K. Kadota. *Phys. Rev. D*, 107(12):123520, June 2023. doi: 10.1103/PhysRevD.107.123520.
- H. Shimabukuro, Y. Xu, and Y. Shao. *Phys. Rev. D*, 112(6):063557, Sept. 2025. doi: 10.1103/

- nrx4-14gb.
- S. Sikder et al. *arXiv e-prints*, art. arXiv:2509.11175, Sept. 2025. doi: 10.48550/arXiv.2509.11175.
- P. H. Sims and J. C. Pober. *Mon. Not. Roy. Astron. Soc.*, 492(1):22–38, 2020. doi: 10.1093/mnras/stz3388.
- D. Spergel et al. *arXiv e-prints*, art. arXiv:1503.03757, Mar. 2015. doi: 10.48550/arXiv.1503.03757.
- B. Spina et al. *Astron. Astrophys.*, 688:L26, Aug. 2024. doi: 10.1051/0004-6361/202450798.
- T.-Y. Sun et al. *Communications Physics*, 8(1):220, May 2025. doi: 10.1038/s42005-025-02139-5.
- Y. Sun et al. *Phys. Rev. D*, 111(4):043015, 2025. doi: 10.1103/PhysRevD.111.043015.
- T. T. Takeuchi et al. *arXiv e-prints*, art. arXiv:1603.01938, Mar. 2016. doi: 10.48550/arXiv.1603.01938.
- T. T. Takeuchi et al. *Astrophys. J. Suppl. Ser.*, 271(2):44, Apr. 2024. doi: 10.3847/1538-4365/ad2517.
- N. Thyagarajan. *Astrophys. J.*, 899(1):16, Aug. 2020. doi: 10.3847/1538-4357/ab9e6d.
- E. O. Vasiliev and Y. A. Shchekinov. *Astrophys. J.*, 777:8, 2013. doi: 10.1088/0004-637X/777/1/8.
- D. Ďurovčiková et al. *Astrophys. J.*, 969(2):162, July 2024. doi: 10.3847/1538-4357/ad4888.
- P. Villanueva-Domingo and K. Ichiki. *Publ. Astron. Soc. Jpn.*, 75:S33–S49, Feb. 2023. doi: 10.1093/pasj/psab119.
- T. Šoltinský et al. *Mon. Not. R. Astron. Soc.*, 506(4):5818–5835, Oct. 2021. doi: 10.1093/mnras/stab1830.
- T. Šoltinský et al. *Mon. Not. R. Astron. Soc.*, 519(2):3027–3045, Feb. 2023. doi: 10.1093/mnras/stac3710.
- T. Šoltinský, G. Kulkarni, S. P. Tendulkar, and J. S. Bolton. *Mon. Not. R. Astron. Soc.*, 537(1):364–378, Feb. 2025. doi: 10.1093/mnras/staf026.
- F. Wang et al. *Astrophys. J.*, 884(1):30, Oct. 2019. doi: 10.3847/1538-4357/ab2be5.
- D. H. Weinberg. In *The Emergence of Cosmic Structure*, volume 666 of *AIP Conference Proceedings*, pages 157–169, 2003. doi: 10.1063/1.1581786.
- L. H. Weinberger, M. G. Haehnelt, and G. Kulkarni. *Mon. Not. R. Astron. Soc.*, 485(1):1350–1366, May 2019. doi: 10.1093/mnras/stz481.
- C. J. Willott et al. *Astron. J.*, 140(2):546–560, Aug. 2010. doi: 10.1088/0004-6256/140/2/546.
- J. Wolf et al. *Astron. Astrophys.*, 691:A30, Nov. 2024. doi: 10.1051/0004-6361/202451035.
- S. A. Wouthuysen. *AJ*, 57:31–32, Jan. 1952. doi: 10.1086/106661.
- Y. Xu et al. *Astrophys. J.*, 704(2):1396–1404, Oct. 2009. doi: 10.1088/0004-637X/704/2/1396.
- Y. Xu, A. Ferrara, and X. Chen. *Mon. Not. R. Astron. Soc.*, 410(3):2025–2042, Jan. 2011. doi: 10.1111/j.1365-2966.2010.17579.x.
- J. Yang et al. *Astrophys. J. Suppl. Ser.*, 269(1):27, Nov. 2023. doi: 10.3847/1538-4365/acf99b.
- S. Yoshiura and K. Takahashi. *Mon. Not. R. Astron. Soc.*, 473(2):1570–1575, Jan. 2018. doi: 10.1093/mnras/stx2462.
- B. Yue, B. Ciardi, E. Scannapieco, and X. Chen. *Mon. Not. R. Astron. Soc.*, 398(4):2122–2133, Oct. 2009. doi: 10.1111/j.1365-2966.2009.15261.x.
- G. R. Zeimann et al. *Astrophys. J.*, 736(1):57, July 2011. doi: 10.1088/0004-637X/736/1/57.
- Z.-X. Zhang, J. Cang, Y. Gao, and H. Li. *JCAP*, 07:027, 2025. doi: 10.1088/1475-7516/2025/07/027.
- Y. Zheng, P. Montero-Camacho, Z. Cai, and Y. Mao. *Mon. Not. R. Astron. Soc.*, 545(4):staf2014,

- Feb. 2026. doi: 10.1093/mnras/staf2014.
- Y. Zhu et al. *Astrophys. J.*, 923(2):223, Dec. 2021. doi: 10.3847/1538-4357/ac26c2.
- Y. Zhu et al. *Astrophys. J.*, 932(2):76, June 2022. doi: 10.3847/1538-4357/ac6e60.
- Y. Zhu et al. *Astrophys. J.*, 955(2):115, Oct. 2023. doi: 10.3847/1538-4357/aceef4.
- Y. Zhu et al. *Mon. Not. R. Astron. Soc.*, 533(1):L49–L56, Sept. 2024. doi: 10.1093/mnrasl/slae061.
- J. J. Ziegler et al. *arXiv e-prints*, art. arXiv:2411.02699, Nov. 2024. doi: 10.48550/arXiv.2411.02699.