



## Cosmology with Intensity Mapping via Statistics Beyond the Power Spectrum in the SKAO Era

Suman Majumdar <sup>1</sup>, Debanjan Sarkar <sup>2</sup>, Leon Noble <sup>1</sup>, Manas Mohit Dosibhatla <sup>1</sup>, Zhaoting Chen <sup>3</sup>, Bernhard Vos-Ginés <sup>4</sup>, Vishrut Pandya <sup>1</sup>, Mohd Kamran <sup>5</sup>, Sourav Pal <sup>6</sup>, Chandra Shekhar Murmu <sup>7</sup>, Cora Uhlemann <sup>4</sup>, Pauline Gorbatchev <sup>8,34</sup>, Jean-Luc Starck <sup>9</sup>, José Luis Bernal <sup>10</sup>, Marta Spinelli <sup>11,12</sup>, Caroline S. Heneka <sup>13</sup>, Matteo Viel <sup>14</sup>, Gabriella De Lucia <sup>5</sup>, Fabio Fontanot <sup>5</sup>, Stefano Camera <sup>15,16,17,12</sup>, Steve Cunnington <sup>18</sup>, Benedetta Spina <sup>13</sup>, Dionysios Karagiannis <sup>19</sup>, Samit Kumar Pal <sup>1</sup>, Anshuman Tripathi <sup>1</sup>, Sarah Libanore <sup>20</sup>, Abinash Kumar Shaw <sup>21,22</sup>, Rajesh Mondal <sup>23</sup>, Yashrajsinh Mahida <sup>1</sup>, Yashar Akrami <sup>24</sup>, Philip Bull <sup>18</sup>, Chris Clarkson <sup>25</sup>, Abhirup Datta <sup>1</sup>, José Fonseca <sup>26,27,12</sup>, Yin-Zhe Ma <sup>28</sup>, Roy Maartens <sup>12</sup>, Liantsoa F. Randrianjanahary <sup>29</sup>, Martin Sahlen <sup>30</sup>, Xin Wang <sup>31</sup>, Yougang Wang <sup>32</sup> and Jochen Weller <sup>33</sup>

<sup>1</sup>Department of Astronomy, Astrophysics & Space Engineering, Indian Institute of Technology Indore, Indore 453552, India

<sup>2</sup>Department of Physics and Trotter Space Institute, McGill University, QC H3A 2T8, Canada

<sup>3</sup>Institute for Astronomy, The University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

<sup>4</sup>Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

<sup>5</sup>INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy

<sup>6</sup>Physics and Applied Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

<sup>7</sup>Astrophysics Research Centre of the Open University (ARCO), The Open University of Israel, 1 University Road, PO Box 808, Ra'anana 4353701, Israel

<sup>8</sup>Institutes of Computer Science and Astrophysics, Foundation for Research and Technology - Hellas (FORTH), Greece

<sup>9</sup>Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France

<sup>10</sup>Instituto de Física de Cantabria (IFCA), CSIC-Univ. de Cantabria, Avda. de los Castros s/n, E-39005 Santander, Spain

<sup>11</sup>Observatoire de la Côte d'Azur, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France

<sup>12</sup>Department of Physics & Astronomy, University of the Western Cape, Cape Town 7535, South Africa

<sup>13</sup>Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany

<sup>14</sup>SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste TS, Italy

<sup>15</sup>Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy

<sup>16</sup>INFN – Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy

<sup>17</sup>INAF – Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, Strada Osservatorio 20, 10025 Pino Torinese, Italy

<sup>18</sup>Jodrell Bank Centre for Astrophysics, Department of Physics & Astronomy, The University of Manchester, Manchester M13 9PL, UK

- <sup>19</sup>*Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy*
- <sup>20</sup>*Department of Physics, Ben-Gurion University of the Negev, Be'er Sheva 84105, Israel*
- <sup>21</sup>*Department of Computer Science, University of Nevada Las Vegas, 4505 S. Maryland Pkwy., Las Vegas, NV 89154, USA*
- <sup>22</sup>*Max-Planck-Institut für Astrophysik, Garching D-85748, Germany*
- <sup>23</sup>*Department of Physics, National Institute of Technology Calicut, Calicut, 673601, Kerala, India*
- <sup>24</sup>*Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain*
- <sup>25</sup>*Department of Physics & Astronomy, Queen Mary University of London, London E1 4NS, U.K.*
- <sup>26</sup>*Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto CAUP, 4150-762 Porto, Portugal*
- <sup>27</sup>*Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*
- <sup>28</sup>*Department of Physics, Stellenbosch University, Matieland 7602, South Africa*
- <sup>29</sup>*Astrophysics Research Centre & School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, 4041, South Africa*
- <sup>30</sup>*Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden*
- <sup>31</sup>*School of Physics and Astronomy, Sun Yat-Sen University, No. 2 Daxue Rd., Zhuhai 519082, China*
- <sup>32</sup>*National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*
- <sup>33</sup>*Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany*
- <sup>34</sup>*Department of Physics, University of Crete, Greece*

E-mail: [mid.suman@gmail.com](mailto:mid.suman@gmail.com), [debanjan.sarkar@mcgill.ca](mailto:debanjan.sarkar@mcgill.ca)

The cosmological distribution of neutral hydrogen (HI) during the post-reionization era is highly non-Gaussian due to the underlying non-linear structure formation, complex galaxy biasing, and potential primordial non-Gaussianity. One needs higher-order (beyond two-point) statistics to maximally extract the non-Gaussian information out of the 21-cm intensity maps. This chapter summarizes the potential of several higher-order statistics, including voxel intensity distribution, emission line stacking, probability density functions,  $\ell_1$ -norm, bispectrum, and various marked statistics. Additionally, image-based morphological descriptors, such as the Largest Cluster Statistic, local dimensions, and Minkowski functionals, etc., can potentially characterize the morphology and geometry of the cosmic web encoded in the 21-cm intensity maps. This chapter presents forecasts of the detectability of these higher-order statistics in the context of the future SKAO observations. These forecasts incorporate instrumental noise, observational effects, and, in some cases, foreground removal in their analyses. With its unprecedented sensitivity, the future SKAO 21-cm observations will enable us to measure these higher-order statistics more precisely, possibly helping to break degeneracies between astrophysical and cosmological parameters, and maximizing the science outcome from these surveys.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Different Higher-order Statistics of the observed HI 21-cm field</b>	<b>5</b>
2.1	One-point statistics	5
2.2	Bispectrum	7
2.2.1	Bispectrum multipole moments	7
2.2.2	Modeling the HI bispectrum	7
2.3	Marked Power Spectrum	8
2.4	Starlet $\ell_1$ -norm	9
2.5	Morphological Measures in the Image Domain	9
2.5.1	Largest Cluster Statistic	9
2.5.2	Local Dimension	10
<b>3</b>	<b>Science with Higher-Order Statistics</b>	<b>11</b>
3.1	One-point statistics	11
3.2	21-cm bispectrum	12
3.2.1	Detectability of the 21-cm bispectrum	13
3.2.2	Detectability of redshift-space 21-cm bispectrum multipoles	14
3.3	Marked Power Spectrum	16
3.4	Three-point correlations and beyond	18
3.5	Starlet $\ell_1$ -norm	18
3.6	Morphological Measures in the Image Domain	19
3.6.1	Largest Cluster Statistic	20
3.6.2	Local Dimension	21
<b>4</b>	<b>Discussion</b>	<b>22</b>

## 1 Introduction

The past few decades have established intensity mapping (Bharadwaj et al., 2001; Bharadwaj and Sethi, 2001; Kovetz et al., 2017; Bernal and Kovetz, 2022) as a transformative observational technique for probing the large-scale structure of the Universe across cosmic time. Line Intensity Mapping (LIM) relies on mapping the integrated line emission from unresolved galaxies and diffuse gas on a coarse resolution element. LIM thus circumvents the observational challenges of individual source detection, granting access to populations of faint and unresolved galaxies and enabling wide-area, deep surveys that traditional galaxy surveys cannot achieve. The 21-cm hyperfine transition from neutral hydrogen (HI) (Pritchard and Loeb, 2012), in particular, offers an exceptional tracer of the cosmic matter distribution during the post-reionization epoch ( $z \lesssim 6$ ), with sufficient brightness and abundance to enable high-fidelity mapping across significant cosmic volumes.

The Square Kilometre Array Observatory (SKAO) brings a watershed moment for 21-cm intensity mapping. With its unprecedented sensitivity, collecting area, and redshift coverage spanning  $z \approx 0.2$

to  $z \approx 3$  with the SKA-Mid array, the SKAO will deliver 21-cm intensity maps at resolutions and signal-to-noise levels far exceeding the capabilities of presently operating radio interferometers. This observational leap creates both opportunity and necessity: while the SKAO's raw sensitivity is formidable, maximizing the scientific return from these observations requires moving beyond traditional two-point statistics to extract the full information content encoded in the data.

Traditionally, LIM analyses have relied primarily on two-point statistics, most commonly the power spectrum, to extract information about the underlying matter field and astrophysical processes. However, the power spectrum is a complete statistical descriptor only in the case of Gaussian random fields. In reality, the late-time matter and HI density fields are highly non-Gaussian due to non-linear structure formation, complex galaxy biasing, astrophysics, and, potentially, primordial non-Gaussianities (PNG) sourced during inflation. The Gaussian assumption, therefore, limits the information extracted from the data.

Higher-order statistics (HOS) of 21-cm LIM are thus essential to overcome the limitations of inference via power spectrum alone and to fully exploit the non-Gaussian information encoded in LIM observations. These HOS include statistics defined in both real and Fourier space. For instance, the bispectrum (Sarkar et al., 2019b; Cunnington et al., 2021; Karagiannis et al., 2022) (three-point correlation function in Fourier space) captures information about mode coupling and the shape dependence of non-Gaussian features, while one-point statistics such as the voxel intensity distribution (VID) (Breysse et al., 2017; Bernal, 2024) encode the full temperature PDF of intensity maps, sensitive to the underlying halo mass function and line-luminosity relation. Similarly, marked statistics (Kamran et al., 2025; Massara et al., 2023) and local morphological descriptors— such as Largest Cluster Statistics (LCS; Klypin and Shandarin 1993; Bag et al. 2018), Minkowski functionals (Mecke et al., 1993; Bag et al., 2018; Pathak et al., 2022), local dimension (Sarkar and Bharadwaj, 2009; Dosibhatla et al., 2026), and tidal tensor (Hahn et al., 2007) — probe the geometry and topology of the signal beyond amplitude correlations.

The additional information encoded in HOS helps in tightening constraints on cosmological parameters and breaking degeneracies between them. There are also degeneracies between astrophysical and cosmological parameters, for instance, the mean brightness temperature of a target line and the bias with which it traces the matter distribution (Castorina and White, 2019). This can be partly broken using the VID (Breysse, 2022). The 21-cm bispectrum has been demonstrated to break the degeneracies of the Hubble parameter ( $h$ ) with evolving dark energy and also primordial power spectrum parameters (Randrianjanahary et al., 2024; Pinheiro et al., 2026). The bispectrum is also an effective summary statistic for breaking the degeneracies of the neutrino mass  $M_\nu$  with other cosmological parameters, especially  $\sigma_8$  and  $\Omega_M$  (Hahn et al., 2020). Adding the information of the halo mass function and the void size function to the power spectrum can break multiple cosmological degeneracies, including those of  $\sigma_8$  with  $\Omega_M$ ,  $\Omega_b$ ,  $h$ , and  $n_s$  (Bayer et al., 2021). Similarly, segmenting the cosmic web into its constituent nodes, filaments, walls, and voids with the tidal tensor and using their individual power spectra jointly has the potential of breaking the stiff degeneracy between  $M_\nu$  and  $\sigma_8$  (Sunseri et al., 2025).

In this chapter, we provide a structured overview of the most promising higher-order statistical tools applicable to 21-cm LIM in the post-reionization regime, emphasizing their complementar-

ity with the power spectrum and their capacity to break degeneracies in cosmological inference. Recent advances in simulation-based inference (SBI) approaches, which circumvent explicit definition of the likelihood function, are crucial for inference using HOS, as these statistics are not expected to have simple Gaussian likelihoods (Sun et al., 2026; Modi et al., 2025; Jalali Kanafi and Movahed, 2026; Lemos et al., 2023). Section 2 outlines the theoretical formalisms of various statistics, including image-domain estimators like the VID, LCS, and local dimension, as well as Fourier-space polyspectra such as the bispectrum and trispectrum. In Section 3, we present forecasts regarding the detectability of each of these statistics via SKAO alongside their modeling approaches and noise/systematics considerations. For the forecast, we use the SKA-Mid AA4 array layout (Seethapuram Sridhar et al., 2025) as our reference instrumental configuration. The uncertainty in measuring each statistic from observations can arise from cosmic variance, thermal noise of the radio telescope, residual foregrounds left in the 21-cm maps after foreground cleaning, and various instrumental systematics. Our reference setup considers only the uncertainty due to thermal noise, unless specified otherwise. The required observation time and observation mode for each summary statistic are indicated in the respective sections.

## 2 Different Higher-order Statistics of the observed HI 21-cm field

The observed 21-cm brightness temperature field can be expressed as a smooth background tracing the cosmic mean HI abundance plus a fluctuating component arising from spatial variations in the HI distribution (Pritchard and Loeb, 2008, 2012; Furlanetto et al., 2009; Furlanetto et al., 2006). In the post-reionization universe, the intergalactic medium is highly ionized, and neutral hydrogen survives only in dense, self-shielded regions such as galaxies and damped Ly $\alpha$  systems. The 21-cm brightness temperature thus can be expressed as

$$T_b(z) = \bar{T}_b(z) [1 + \delta_{\text{HI}}(z)], \quad (1)$$

where  $\delta_{\text{HI}}$  is the HI overdensity and the mean brightness temperature is given by (Battye et al., 2012; Bull et al., 2015; Camera et al., 2015; Padmanabhan et al., 2023)

$$\bar{T}_b(z) = 0.566 h \frac{H_0}{H(z)} \left( \frac{\Omega_{\text{HI}}}{0.003} \right) (1+z)^2 \text{ mK}. \quad (2)$$

Observations and simulations (Padmanabhan et al., 2023; Castorina and Villaescusa-Navarro, 2017; Villaescusa-Navarro et al., 2014) indicate that  $\Omega_{\text{HI}}$  is approximately constant for  $0 \lesssim z \lesssim 6$ , with  $\Omega_{\text{HI}} \simeq 4.7 \times 10^{-4}$ .

### 2.1 One-point statistics

In cosmology, the probability distribution of matter density fluctuations is key to understanding the large-scale structure of the universe. While linear theory predicts Gaussian statistics for early-universe fluctuations, gravitational evolution induces significant non-Gaussianities at late times.

The voxel intensity distribution (VID) is an estimator of the probability distribution function of the measured temperatures in a given voxel (Breyse et al., 2017; Bernal, 2024). Intuitively, such

probability  $\mathcal{P}(T)$  is the marginalization over the conditional probability  $\mathcal{P}(T|N)$  of measuring a temperature  $T$  in a voxel with  $N$  galaxies:

$$\mathcal{P}(T) = \sum_N \mathcal{P}(T|N)\mathcal{P}(N), \quad (3)$$

where  $\mathcal{P}(N)$  is the probability of having  $N$  galaxies contributing within a voxel. Therefore, besides granting access to the non-Gaussian information in line-intensity maps, the VID is very sensitive to the line-luminosity function and hence highly complementary to the power spectrum. While any modeling of the VID conceptually reduces to Eq. (3), there are different approaches (Bernal, 2024) for its computation. If properly combined with the power spectrum (Ihle et al., 2019; Sato-Polito and Bernal, 2022), the VID breaks degeneracies between astrophysical and cosmological uncertainties (see e.g., Breyse 2022), boosting the constraining power of LIM for beyond  $\Lambda$ CDM models (Sabra et al., 2024).

While the VID contains a wealth of information, it is also highly susceptible to observational systematics. There are different approaches involving cross-correlations at one-point that increase the robustness of these measurements. While the deconvolved density estimator (Breyse et al., 2023; Chung et al., 2023) and the conditional VID (Breyse et al., 2019) aim to capture the probability distributions of different tracers, stacking averages the LIM signal around known tracers (Lujan Niemeyer et al., 2022; Chen et al., 2025; Dunne et al., 2025). Given its robustness and higher signal-to-noise ratio, the latter is optimal to characterize the signal in noisier observations and to study instrumental and observational systematics, and therefore can be a preparatory step for more complex studies of one-point statistics. In this subsection, we discuss the emission line stacking of the 21-cm signals. The stacked signal is calculated as

$$I(\Delta\alpha, \Delta\phi, \Delta\nu) = \frac{\sum_i I(\alpha_i + \Delta\alpha, \phi_i + \Delta\phi, \nu_i + \Delta\nu) w_i}{\sum_i w_i}, \quad (4)$$

where  $(\alpha_i, \phi_i)$  are the R.A. and decl. of the voxel in which the  $i^{\text{th}}$  galaxy resides,  $I$  is the flux density, and  $w_i$  is the weight of each voxel,  $\nu_i$  is the  $i^{\text{th}}$  frequency channel and  $\Delta\nu$  is the channel width.

Finally, Large Deviation Theory (LDT) offers a robust framework to model the PDF of densities in spheres by capturing the exponentially rare but physically relevant fluctuations driven by nonlinear gravitational dynamics (Uhlemann et al., 2016). In this context, the PDF is approximated as:

$$\mathcal{P}_R(\rho) \propto \exp[-\Psi_R(\rho)], \quad \Psi_R(\rho) = \frac{1}{2} \frac{\delta_L^2}{\sigma_L^2(R\rho^{1/3})} \frac{\sigma_L^2(R)}{\sigma_{\text{NL}}^2(R)} \quad (5)$$

where  $\rho = 1 + \delta$ ,  $\Psi_R(\rho)$  is the rate function computed at a smoothing radius  $R$ , as a function of the linear variance  $\sigma_L^2$ , the non-linear variance  $\sigma_{\text{NL}}^2$  and the linear overdensity  $\delta_L(\rho)$ , which is usually related to the late-time density  $\rho$  using spherical collapse theory. This formulation enables accurate predictions of the PDF into the non-Gaussian regime, provided that the non-linear variance is well below unity. The PDF of biased tracers such as neutral hydrogen (Leicht et al., 2019) can be modeled considering bias and shot-noise parametrizations in terms of the dark matter overdensity (McCarthy Gould et al., 2025).

## 2.2 Bispectrum

The post-reionization HI 21-cm signal is expected to be significantly non-Gaussian on length scales that have become non-linear. The power spectrum, by definition, measures the amplitude of the signal fluctuations at different scales and only completely characterizes the statistical properties of a signal if it is a Gaussian random field. In order to quantify the non-Gaussianity present in the 21-cm signal, one has to choose a statistic sensitive to that non-Gaussianity. One-point statistics, such as skewness and kurtosis, are such possibilities. However, these one-point statistics can only capture the non-Gaussianity globally and fail to capture the correlations between the non-Gaussianity present at different scales. To comprehensively quantify both, one has to consider higher-order statistics such as the bispectrum. The bispectrum  $B_{21\text{-cm}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$  of the 21-cm differential brightness temperature field  $\delta T_b(\mathbf{x})$  is defined as

$$\langle \Delta_b(\mathbf{k}_1) \Delta_b(\mathbf{k}_2) \Delta_b(\mathbf{k}_3) \rangle = V \delta_{\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3, 0} B_{21\text{-cm}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \quad (6)$$

where the  $\Delta_b(\mathbf{k})$  is the Fourier transform of the differential brightness temperature field  $\delta T_b(\mathbf{x})$  and  $V$  is the total volume of the simulation box or observation volume. The  $\delta_{\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3, 0}$  is the Kronecker delta function, and its numerical value is equal to one when the condition  $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$  is satisfied and zero otherwise. This ensures that only the closed  $k$ -triangles contribute to the bispectrum estimates. The bispectrum can be estimated for different sizes and shapes of the  $k$ -triangles in Fourier space. To identify all of the unique shapes of the triangles in the Fourier space, one can use the bispectrum parametrization introduced in [Bharadwaj et al. \(2020\)](#). According to this parametrization, for a triangle in Fourier space with  $k_1 \leq k_2 \leq k_3$ , its size is determined by  $k_1$ , and the shape is determined by  $t = k_2/k_1$  and  $\mu$ , the cosine of the angle ( $\cos \theta$ ) between  $k_1$  and  $k_2$ .

### 2.2.1 Bispectrum multipole moments

The anisotropies due to the redshift space distortions (RSD) in the 21-cm bispectrum contain a wealth of cosmological information, which is completely quantified using HI bispectrum multipole moments. As RSD introduces angular dependence in the bispectrum with respect to the line of sight, the redshift-space bispectrum is expanded in spherical harmonics as ([Bharadwaj et al., 2020](#))

$$B_{\text{HI}}^s(k_1, \mu, t, \hat{\mathbf{p}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} B_{\ell}^m(k_1, \mu, t) Y_{\ell}^m(\hat{\mathbf{p}}), \quad (7)$$

where  $B_{\ell}^m$  are the bispectrum multipoles and  $\hat{\mathbf{p}}$  characterizes the orientation of the triangle relative to the line of sight. To highlight anisotropic effects, we define the reduced multipoles

$$Q_{\ell}^m(k_1, \mu, t, z) \equiv \frac{B_{\ell}^m(k_1, \mu, t, z)}{B_{\text{HI}}^r(k_1, \mu, t, z)}, \quad (8)$$

which remove the overall amplitude  $\bar{T}_b^3$  and emphasize the RSD- and Finger-of-God (FoG)-induced anisotropies.

### 2.2.2 Modeling the HI bispectrum

On large scales, the HI field can be modeled as a locally biased tracer of the matter overdensity:

$$\delta_{\text{HI}} = b_1 \delta_m + \frac{b_2}{2} \delta_m^2, \quad (9)$$

where  $b_1$  and  $b_2$  are the linear and quadratic bias parameters, respectively. We adopt redshift-dependent fits calibrated from simulated real-space HI bispectrum measurements (Sarkar et al., 2016, 2019a). At tree level in perturbation theory, the real-space HI bispectrum for a closed Fourier triangle  $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$  can be written schematically as

$$B_{\text{HI}}(k_1, k_2, k_3) = \frac{1}{b_1} \left[ 2F_2 P_{\text{HI}}(k_1) P_{\text{HI}}(k_2) + \text{cyc.} \right] + \frac{b_2}{b_1^2} \left[ P_{\text{HI}}(k_1) P_{\text{HI}}(k_2) + \text{cyc.} \right], \quad (10)$$

where  $F_2$  is the standard symmetrized second-order kernel. The first term describes non-linear gravitational mode coupling (shape-dependent), while the second captures isotropic contributions from non-linear HI bias. As these two terms have distinct geometric dependences, measurements of the bispectrum can independently constrain  $b_1$  and  $b_2$ , improving upon power-spectrum-only analyses.

### 2.3 Marked Power Spectrum

Traditional two-point statistics, such as the power spectrum, are optimal for Gaussian random fields; however, the late-time large-scale structure of the Universe is highly non-Gaussian due to gravitational evolution. As structures collapse and form filaments, clusters, and voids, higher-order correlations emerge that the standard power spectrum cannot capture. Marked statistics, as introduced in White (2016), address this limitation by introducing a non-linear transformation—a ‘mark’—applied to the density field before computing its two-point function. This mark is typically a function of local environmental properties such as smoothed brightness temperature, density, halo mass, or velocity dispersion, and it effectively re-weights different regions of space to enhance sensitivity to specific physical regimes. This approach provides an independent and computationally cheaper alternative to the higher-order statistics to describe the non-Gaussianity in the field.

The Marked Power spectrum  $\mathbf{P}_{\mathbf{M}}(\mathbf{k})$  of a field  $A(\mathbf{x})$  can be defined as

$$\langle m_A(\mathbf{k}) m_A^*(\mathbf{k}') \rangle = V \delta_{\mathbf{k}+\mathbf{k}'} \mathbf{P}_{\mathbf{M}}(\mathbf{k}), \quad (11)$$

where  $V$  is the simulation volume and  $m_A(\mathbf{k})$  is the Fourier transform of the mark  $m_A(\mathbf{x})$ —a function of  $A(\mathbf{x})$ , acting as a transformation.

Recent works by Kamran et al. (2025) and Pandya et al. (2026) have demonstrated that applying marked statistics to simulated 21-cm fields can enhance sensitivity to the evolving spatial distribution of HI across redshift. For instance, weighting the HI field by inverse powers of the local smoothed density suppresses contributions from overdense, highly biased regions, thereby amplifying the signal from under-dense environments that are otherwise washed out in unmarked analyses. This reweighting helps to probe the environmental dependence of HI clustering, trace changes in the cosmic HI reservoir, and break degeneracies in cosmological parameter inference. As future 21-cm surveys such as SKA-Mid target the post-reionization era, marked statistics will be an essential tool to access higher-order information encoded in the large-scale distribution of neutral hydrogen.

## 2.4 Starlet $\ell_1$ -norm

The  $\ell_1$ -norm of isotropic undecimated wavelet coefficients offers a compact, multiscale summary of HI intensity maps, sensitive to sparse, non-Gaussian structures not captured by traditional two-point statistics.

Let  $M(\hat{n})$  denote the HI temperature fluctuation map on the sphere, where  $\hat{n}$  is the sky direction. The spherical Starlet transform decomposes this map into a set of wavelet coefficients  $w_j(\hat{n})$  at each scale  $j = 1, \dots, J$ , and a coarse residual map  $c_J(\hat{n})$ , as follows:

$$M(\hat{n}) = \sum_{j=1}^J w_j(\hat{n}) + c_J(\hat{n}). \quad (12)$$

Each wavelet scale  $w_j(\hat{n})$  captures angular features localized in a specific band of multipoles. The coefficients are normalized by a scale-dependent factor  $\mathcal{N}_j$  to account for varying amplitudes across scales. The normalized coefficients are then binned in signal-to-noise ratio (SNR) space.

We define the  $\ell_1$ -norm of the normalized coefficients at scale  $j$  within SNR bin  $k$  as:

$$\ell_1^{(j,k)} = \sum_{\hat{n} \in \mathcal{B}_k} \left| \frac{w_j(\hat{n})}{\mathcal{N}_j} \right|, \quad (13)$$

where  $\mathcal{B}_k$  is the set of pixels whose normalized coefficient falls in the  $k$ -th bin:

$$\mathcal{B}_k = \left\{ \hat{n} \left| T_k \leq \frac{w_j(\hat{n})}{\mathcal{N}_j} < T_{k+1} \right. \right\}, \quad (14)$$

and  $T_k$  are the SNR bin edges. The full summary statistic is  $\ell_1^{(j,k)} \in \mathbb{R}^{J \times K}$ , where  $K$  is the number of bins and  $J$  is the number of wavelet scales.

## 2.5 Morphological Measures in the Image Domain

Even though marked statistics and polyspectra retain the non-Gaussian information in intensity maps, they cannot preserve the complete phase information in the Fourier space, resulting in information loss of morphological features of the LSS. This motivates analyzing intensity maps in the image domain to characterize their topology and morphology.

### 2.5.1 Largest Cluster Statistic

In a binary field that can be segmented into clusters and regions not belonging to any cluster, the largest cluster statistic (LCS) is defined as the fraction of cluster volume that resides in the largest cluster (Klypin and Shandarin, 1993; Bharadwaj et al., 2000; Bag et al., 2018; Pathak et al., 2022; Dasgupta et al., 2023; Pal et al., 2025b):

$$\text{LCS} = \frac{\text{Volume of the largest cluster}}{\text{Total cluster volume}}. \quad (15)$$

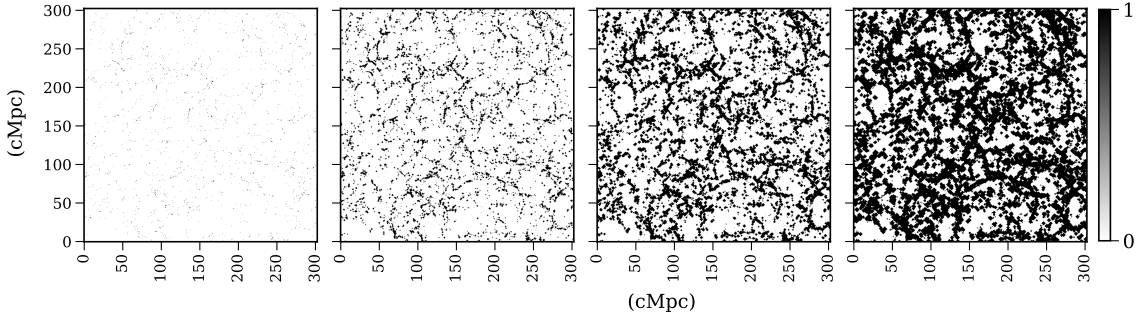
In line intensity maps targeting emission lines from galaxies, a cluster can be defined as a connected set of bright spots in the line emission.

Percolation transition refers to the onset of large-scale connectivity in the cosmic web, where the matter distribution connects to form one contiguous large cluster spanning the whole universe. The LCS is useful for studying the percolation transition of an intensity map (Shandarin, 1983; Klypin and Shandarin, 1993; Bharadwaj et al., 2000; Pathak et al., 2022; Dasgupta et al., 2023; Regős et al., 2024; Pal et al., 2025b), which is quantified by the evolution of the LCS with the filling factor (FF) of the map, defined as

$$\text{FF} = \frac{\text{Total cluster volume}}{\text{Total survey volume}}. \quad (16)$$

As the bright spots grow due to the growth of structure, the cluster volume, and hence the filling factor, grow steadily. However, at the stage where large clusters start to merge, the volume of the largest cluster grows sharply and becomes comparable to the total cluster volume. This makes the LCS approach 1 abruptly and marks the percolation transition.

In practice, the resolution of intensity maps may not be adequate to reveal the connectivity of the matter distribution, as intensity maps trace line emission from biased tracers (galaxies and diffuse gas). Therefore, we employ an iterative coarse-graining scheme as shown in Figure 1, where cells adjacent to bright cells are marked as bright (Bharadwaj et al., 2000; Dosibhatla et al., 2026). This is done repeatedly, and the FF and LCS are calculated after every iteration. The evolution of the LCS with FF gives the percolation curve. The percolation transition is expected to be attained at different filling factors for different intensity maps, and therefore, the LCS is a useful higher-order summary statistic.



**Figure 1:** Illustration of the iterative coarse-graining scheme used for percolation analysis. The four panels show the same slice of a field without coarse-graining (first from left) and after 2, 4, and 6 iterations in order. On each iteration, the cells sharing a face with a bright cell are marked as bright. The fields are binary, with the dark cells marked by zeros and the bright cells by ones.

### 2.5.2 Local Dimension

The LSS of the universe consists of filaments, sheets, nodes, and voids, which form a cellular structure known as the “cosmic web” (Gregory and Thompson, 1978; Einasto et al., 1984; Zeldovich et al., 1982; Bond et al., 1996; Bharadwaj et al., 2000; Aragón-Calvo et al., 2010; Libeskind et al., 2017; Tojeiro and Kraljic, 2025). The contribution of the different building blocks of the cosmic web to the line emission in an intensity map contains a wealth of non-Gaussian and morphological information. The local dimension (Sarkar and Bharadwaj, 2009; Sarkar et al., 2012; Sarkar and Pandey, 2019; Pandey and Sarkar, 2020) is a simple yet effective way of determining whether a

cell in an intensity map belongs to a filament, sheet, or a volume-filling environment (node or void) (Dosibhatla et al., 2026). The local dimension ( $D$ ) of a cell in an intensity map determines how the number of bright cells,  $N(R)$ , in a sphere centered at the given bright cell scales with the radius of the sphere.

$$N(R) = AR^D \quad (17)$$

where  $A$  is a normalization constant. For straight filaments and plane sheets passing through the center of the sphere, and volume-filling environments,  $D$  takes values  $\simeq 1, 2,$  and  $3$  respectively. However, fractional values are allowed for environments intermediate between any two kinds.

To compute the local dimension of any bright cell in an intensity map, we count  $N(R)$  by incrementing  $R$  by a multiple of the grid resolution  $\Delta R$  between fixed values  $R_{\min}$  and  $R_{\max}$ . We then fit the variation of  $N(R)$  with  $R$  using equation 17. If the fit converges with  $\chi^2_{\nu} \leq \chi^2_{\nu, \max}$ , the cell is classifiable, and it belongs to the environment corresponding to  $D$  in the length scale  $R_{\min}$  to  $R_{\max}$ . The value of  $\chi^2_{\nu, \max}$  is chosen such that a good fit is obtained and a large number of cells is classifiable. The fraction of cells in different environments characterizes the morphology of an intensity map effectively.

### 3 Science with Higher-Order Statistics

#### 3.1 One-point statistics

Figure 2a presents the Probability Density Function (PDF) of HI temperature fluctuations,  $\mathcal{P}(\Delta T_{\text{HI}})$ , for three smoothing radii  $R = 15.5, 24.8, 31.0$  Mpc/h of simulated HI Intensity Maps where foreground-removal and telescope beam observational systematics have been taken into account. Our foreground removal technique approximately emulates a FastICA method with  $N_{\text{IC}} = 4$  independent components removed. Dashed and solid lines show the signal without and with including thermal noise, respectively, which slightly increases the variance. We also show the  $1\sigma$  Jackknife errors as shaded regions. We consider a standard SKA-Mid AA4 configuration, although thermal noise is computed by modifying the pixel area and the channel width such that it corresponds to an Intensity Map with  $L_{\text{cell}} = 1$  Mpc/h. Figure 2b presents a  $(\Omega_M, \sigma_8)$  Fisher forecast comparing the constraining power of the Power spectrum ( $P(k)$ ),  $(\sigma_{\Omega_M}, \sigma_{\sigma_8}) = (6.7, 9.5) \cdot 10^{-3}$ , and its combination with the PDF at a single scale ( $P(k) + \mathcal{P}(\Delta T_{\text{HI}})$ ) at  $R = 31.0$  Mpc/h,  $(\sigma_{\Omega_M}, \sigma_{\sigma_8}) = (2.8, 5.2) \cdot 10^{-3}$ , for an SKA-Mid survey (with a redshift range  $1.1 < z < 1.5$  and a sky fraction  $f_{\text{sky}} = 0.3$  corresponding to a survey volume of  $V_{\text{survey}} = 16.3$  [Gpc/h] $^3$ ). Including the PDF alleviates the degeneracy between  $\Omega_M$  and  $\sigma_8$  when compared with the power spectrum. The correlation factors for both parameter covariance matrices are  $r_{P(k)+\mathcal{P}(\Delta T_{\text{HI}})} = 0.867$  and  $r_{P(k)} = 0.964$ . Finally, we show that the cosmology inferred from the predicted statistics applied to the UNIT simulations is well below  $1\sigma$  (black and red points) (Vos-Ginés et al. in prep.), although higher-volume simulations are required as the survey volume considered in this setup exceeds the volume of our simulations.

We use `HMcode-2020` (Mead et al., 2021) integrated in `CAMB` together with a bias and shot-noise prescriptions and observational systematic kernels to predict the HI power spectrum. We use `CosMomentum` (Friedrich et al., 2020) to predict the HI PDF, considering bias and shot-noise prescriptions and introducing systematics as a variance rescaling. Those theoretical statistics are

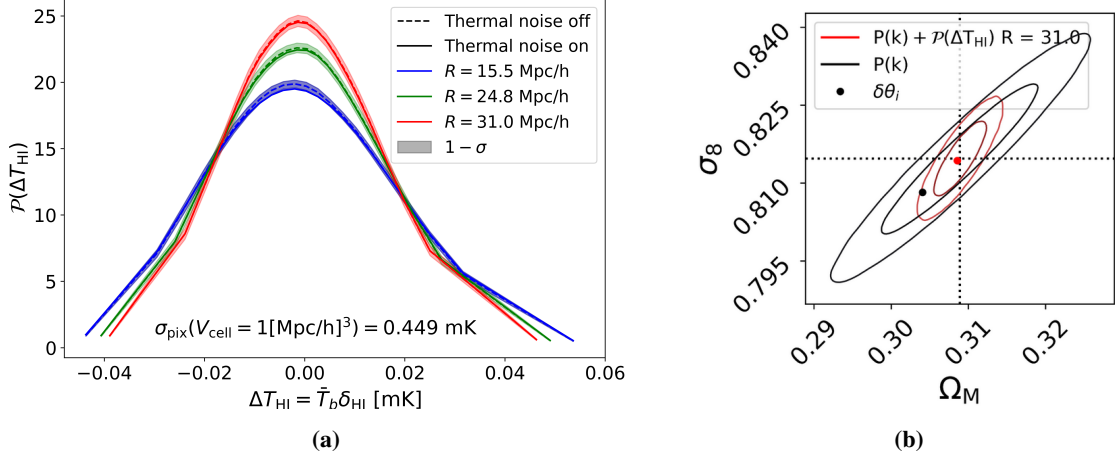
validated against statistics measured in high-resolution simulations at the fiducial cosmology (see Table 4 of [Planck Collaboration \(2016\)](#)). We use all four available realizations of the high-resolution dark matter simulation UNIT ([Chuang et al., 2019](#)) snapshots at  $z = 1.321$ . For each realization, a semi-analytical model of galaxy evolution (SAGE) is applied to derive the cold gas mass distribution ([Croton et al., 2016](#)). An HI prescription converts the cold gas mass to the HI mass, and telescope beam and foreground removal systematics are added to the HI catalog following [Avila et al. \(2021\)](#). An Intensity Map is constructed in real space considering the assignment of a Nearest Grid Point (NGP) to cubic cells of side length  $L_{\text{cell}} = 1 \text{ Mpc}/h$  ([Avila et al., 2021](#)). Finally, we add thermal noise to each cubic cell,  $\sigma_{\text{pix}}$  ([Battye et al., 2013](#)). We compute the PDF of the HI field smoothed over spheres with three radii  $R = 15.5, 24.8, 31.0 \text{ Mpc}/h$ , considering 22 quantiles for each scale. We disregard the first and last PDF quantiles to limit the impact of nonlinearities and non-Gaussian distribution of the data. We consider the Power spectrum only up to  $k_{\text{max}} = 0.091 \text{ Mpc}/h$ , since we expect smaller scales to be dominated by thermal noise. The data covariance matrix is computed for both  $P(k)$  and its combination with the PDF at fixed cosmology, considering Jackknife resampling with  $n = 125$  realizations, and its inverse is corrected using the Hartlap factor ([Kaufman, 1967](#); [Hartlap, J. et al., 2007](#)).

Using the simulation of the dark matter halos and their HI content, we generate random realizations of the HI field to study the detectability of the stacking signal. Following the mock pipeline described in [Chen et al. \(2025\)](#), we generate 100 realizations of the stacked 21-cm signal following the same noise statistics mentioned above. Assuming that halos with  $M_{\text{HI}} > 10^{10.5} M_{\odot}$  will be selected for stacking, the stacked signal can be calculated using [Equation 4](#). The realizations are then used to calculate the covariance of the stacked signal. The corresponding signal-to-noise ratio is shown in [Figure 3](#).

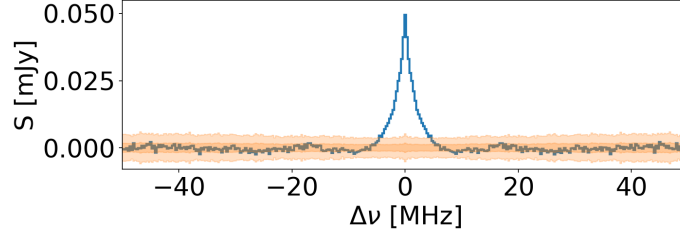
### 3.2 21-cm bispectrum

21-cm bispectrum is one of the promising higher-order statistics, which can be used to extract information contained in the non-Gaussian components of the 21-cm signal. Measurements of the 21-cm bispectrum can independently constrain both linear and quadratic HI bias, thereby improving upon power-spectrum-only analyses. In a previous work, [Sarkar et al. \(2019b\)](#) studied the HI bispectrum using a set of semi-numerical simulations of the HI distribution. They determine the scale and redshift range where the HI bispectrum can be adequately modeled using the predictions of second-order perturbation theory, and they used this to predict the redshift evolution of the linear and quadratic HI bias. In the follow-up work, [Chhabra and Bharadwaj \(2025\)](#) showed that the HI mass-halo mass relation can be constrained by combining the 21-cm power spectrum and 21-cm bispectrum for all unique  $k$  – triangles. [Cunnington et al. \(2021\)](#) presented a model of the redshift space 21-cm bispectrum based on second-order perturbation theory. Additionally, they studied the effect of the radio telescope beam and 21-cm foreground contamination.

The redshift space distortions can enhance the amplitude of the 21-cm bispectrum. [Figure 4](#) shows that at  $z = 2$  and  $k_1 = 0.1 h \text{ Mpc}^{-1}$ , the monopole  $Q_0^0$  has a strong overall enhancement in redshift space, with the largest amplification near the equilateral configuration and significant boosts toward stretched and squeezed triangles. Its peak location shifts with redshift, favoring stretched shapes at lower  $z$ . Among the  $\ell = 2$  quadrupoles,  $Q_2^0$  is strongest for linear and stretched triangles ( $\mu \rightarrow 1$ ,



**Figure 2:** **Left:** Normalized HI PDF including telescope beam and foreground removal systematics, calculated for smoothing scales  $R = 15.5, 24.8, 31.0$  Mpc/h. Dashed lines do not include thermal noise, and solid lines include thermal noise.  $1 - \sigma$  errors are represented as shaded contours. **Right:**  $(\Omega_M, \sigma_8)$  Fisher forecast for SKA-Mid HI IM survey considering  $z \sim 1.1 - 1.5$  and  $f_{\text{sky}} = 0.3$ . We show  $1\sigma$  and  $2\sigma$  contours of the  $P(k)$  ( $P(k) + \mathcal{P}(\Delta T_{\text{HI}})$ ) at  $R = 31.0$  Mpc/h as black (red) ellipses. Black and red points represent the cosmologies inferred from the predicted  $P(k)$  and PDF applied to the UNIT simulations, respectively.

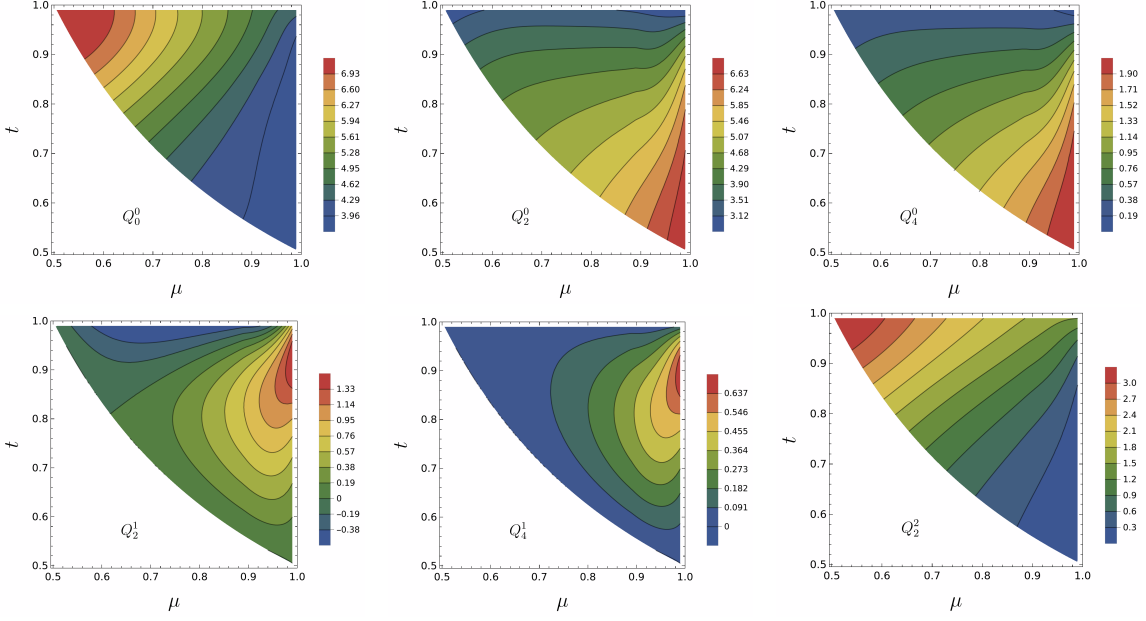


**Figure 3:** The forecast for the emission line stacking using single dish HI intensity mapping with SKA-Mid at  $z \sim 1.1 - 1.5$ . The blue solid line shows the expected HI signal. The outer orange shaded region shows the expected  $3\sigma$  noise level, whereas the inner region shows the  $1\sigma$  noise level.

$t \lesssim 0.8$ ) and becomes suppressed near the equilateral limit. The  $Q_2^1$  mode alternates between enhancement and suppression across the  $(\mu, t)$  plane, peaking near squeezed configurations and turning negative near the L-isosceles limit. The  $Q_2^2$  mode mirrors the monopole behavior, with its largest amplitude near the equilateral shape. Higher-order multipoles ( $\ell \geq 4$ ) exhibit more complex but weaker angular structure; FoG damping suppresses them more effectively, leaving only modest enhancement in  $Q_4^0$  for stretched configurations and mild signatures in  $Q_4^1$  near squeezed triangles.

### 3.2.1 Detectability of the 21-cm bispectrum

In the Figure 5, we show the forecast for the detectability of spherically averaged 21-cm bispectrum for all unique  $k -$  triangles with SKA-Mid observations. The right panel of the Figure 5 show the 21-cm bispectrum for all unique  $k -$  triangles for  $k_1 = 0.61 \text{Mpc}^{-1}$  at redshift  $z = 0.99$ . The 21-cm intensity maps used for this forecast were generated by postprocessing the galaxy catalogs from the

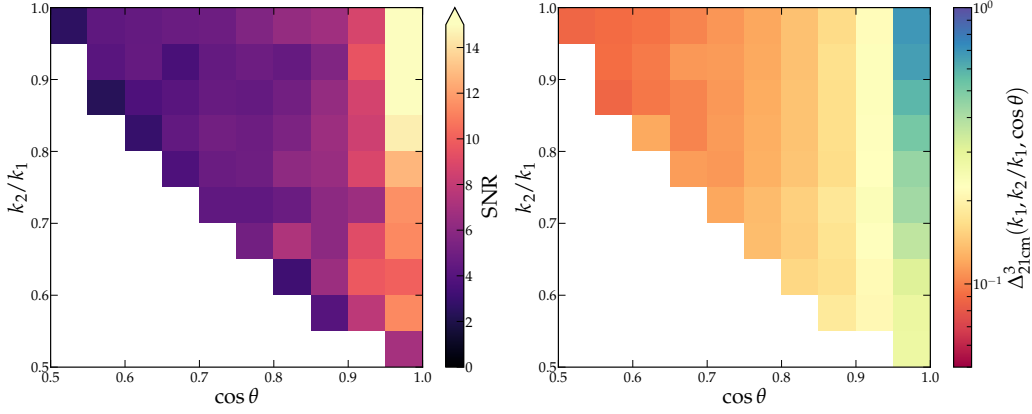


**Figure 4:** Heatmaps of the reduced bispectrum multipoles of the HI density contrast in the  $(\mu, t)$ -plane. The dimensionless reduced bispectrum multipoles are defined as  $Q_\ell^m = B_\ell^m / B_{\text{HI}}^r$ , where  $B_\ell^m$  denotes the redshift-space bispectrum multipoles including the FoG effect, and  $B_{\text{HI}}^r$  represents the real-space HI bispectrum. The results are shown for  $k_1 = 0.1 \text{ hMpc}^{-1}$  at redshift  $z = 2$ . The redshift-space bispectrum multipoles  $Q_\ell^0$  with  $\ell = 0, 2, 4$  are amplified in the stretched configurations, while  $Q_2^1$  and  $Q_4^1$  exhibit enhancement along the linear triangles, peaking near the squeezed limit. In contrast, the  $Q_2^2$  multipole shows the strongest enhancement near the equilateral configuration.

GAEA semi-analytic galaxy formation model (De Lucia et al., 2014; Hirschmann et al., 2016; De Lucia et al., 2024; Fontanot et al., 2025, and references therein). For the signal-to-noise ratio (SNR) estimation, we only include the uncertainty due to thermal noise and neglect any contribution from the cosmic variance (Flöss et al., 2023) and foreground residuals (Cunnington et al., 2021). To estimate the SNR, we simulate 50 statistically independent realizations of the system noise map within the volume of the 21-cm signal, and add the system noise and 21-cm signal. To estimate the system noise map, we considered 300 hours of SKA-Mid AA4 observations for each pointing in the interferometric mode. The right panel of Figure 5 shows the SNR prediction of the 21-cm bispectrum for all unique  $k$  – triangles for  $k_1 = 0.61 \text{ Mpc}^{-1}$  at redshift  $z = 0.99$ . Except for a few unique  $k$  – triangles, all others have SNR higher than  $3\sigma$  and  $k$  – triangles in the vicinity of the squeezed-limit ( $k_2/k_1 = 1$  and  $\cos \theta \rightarrow 1$ ) show SNR higher than  $10\sigma$  (Noble et al., 2026).

### 3.2.2 Detectability of redshift-space 21-cm bispectrum multipoles

To assess detectability, we model the thermal noise for SKA-Mid AA4 with  $N_{\text{ant}} = 197$  antennas of diameter  $D_{\text{dish}} = 15 \text{ m}$  and aperture efficiency  $\eta = 0.7$ . The effective area per dish is  $A_e = \eta \pi D_{\text{dish}}^2 / 4$ . Each Fourier mode corresponds to a baseline  $b = (\lambda k_\perp r) / (2\pi)$ , where  $\lambda = 0.21(1+z) \text{ m}$  and  $r(z)$  is the comoving distance. Assuming a baseline distribution  $\rho(b) \propto b^{-2}$ , an observing time  $T_{\text{obs}} = 400 \text{ hr}$  per pointing, bandwidth  $B = 32 \text{ MHz}$ , and system temperature  $T_{\text{sys}} = 60 \text{ K}$  at  $z = 2$ ,



**Figure 5: Right:** 21-cm bispectrum for all unique  $k$  – triangles for  $k_1 = 0.61\text{Mpc}^{-1}$  at  $z = 0.99$ . **Left:** The signal-to-ratio of 21-cm bispectrum for all unique  $k$  – triangles for 300 hours of observation per pointing with SKA-Mid.

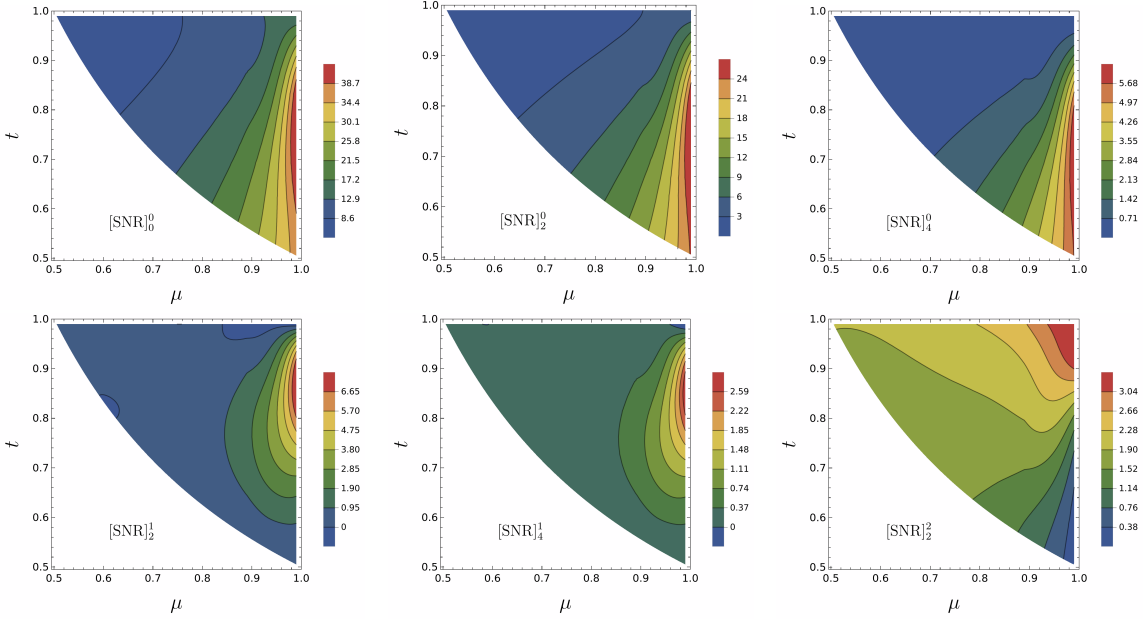
the thermal-noise power is

$$P_N(k, \mu) = \frac{A_e T_{\text{sys}}^2 r^2(z)}{t_k} \frac{dr}{dv}, \quad (18)$$

where  $t_k$  is the effective integration time per mode. The bispectrum multipoles are then estimated as weighted sums over closed triangle triplets within each bin (Mazumdar et al., 2020; Pal et al., 2025a; Pal et al., 2025). The covariance matrix  $C_{\ell\ell'}^{mm'}$  ( $k_1, \mu, t$ ) includes both cosmic variance and thermal noise contributions. The signal-to-noise ratio (SNR) for each multipole is

$$[\text{SNR}]_{\ell}^m(k_1, \mu, t) = \frac{B_{\ell}^m(k_1, \mu, t)}{\sqrt{C_{\ell\ell}^{mm}(k_1, \mu, t)}}. \quad (19)$$

Using the noise and covariance model above, we compute the SNR maps for SKA-Mid at  $z = 2$  and show the results in Figure 6. The SNR peaks in the limit of linear-triangles ( $\mu \rightarrow 1, 0.6 \lesssim t \lesssim 0.85$ ), where FoG suppression is weakest, and declines toward equilateral configurations. For the monopole ( $\ell = 0, m = 0$ ), the maximum SNR reaches  $\sim 40$  near linear triangles and  $\sim 8$  near equilateral. The quadrupole ( $\ell = 2, m = 0$ ) attains  $\sim 25$  in the same region, decreasing to  $\sim 3$  near equilateral. The hexadecapole ( $\ell = 4, m = 0$ ) peaks around  $\sim 6$  near linear configurations but is heavily damped at equilateral. Among the azimuthal modes,  $(\ell, m) = (2, 1)$  peaks near the linear limit with  $\text{SNR} \lesssim 10$ ,  $(4, 1)$  remains small (maximum  $\sim 2.5$ ), and  $(2, 2)$  shows a broader pattern peaking toward the squeezed limit with  $\text{SNR} \sim 3$ . Overall, the monopole dominates the detectable signal, followed by the  $\ell = 2$  modes, with higher  $\ell$  and  $m > 0$  multipoles measurable but subdominant. Configurations close to linear triangles offer the highest detectability, whereas equilateral shapes are most strongly affected by FoG damping. These results highlight the power of the multipole framework to isolate anisotropic signatures of RSD and to pinpoint the most informative regions of configuration space for future 21-cm surveys with SKA-Mid (Cosmic Visions 21 cm Collaboration et al., 2018; Dash and Sarkar, 2025; Yadav and Sarkar, 2025; Dash et al., 2023; Mazumdar et al., 2020; Pal et al., 2025a; Pal et al., 2025).



**Figure 6:** Signal-to-noise ratio (SNR) maps of the bispectrum multipoles up to  $(\ell, m) = (4, 1)$  in the  $(\mu, t)$ -plane. The SNR is computed for  $k_1 = 0.1 h \text{ Mpc}^{-1}$  at redshift  $z = 2$ . For the  $B_\ell^0$  multipoles with  $\ell = 0, 2, 4$ , the SNR for SKA-Mid peaks at  $\mu \rightarrow 1$  and  $0.6 \leq t \leq 0.85$ . The  $B_2^1$  and  $B_4^1$  multipoles exhibit maximum SNR near the linear triangle configurations (i.e.,  $\mu \rightarrow 1, 0.75 \leq t \leq 0.95$ ), while the  $B_2^2$  multipole attains its highest SNR near the squeezed limit ( $\mu \rightarrow 1, t \rightarrow 1$ ).

### 3.3 Marked Power Spectrum

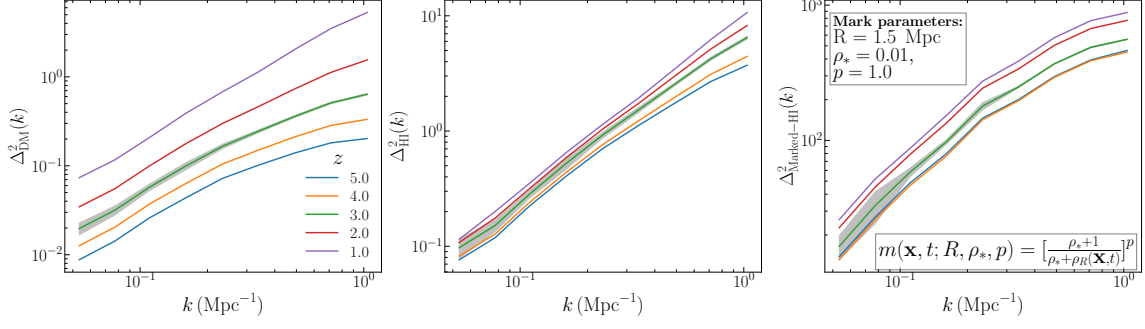
In this section, we take a step further and explore how marked statistics can be leveraged in the context of the upcoming 21-cm intensity mapping surveys with SKA-Mid. Given its sensitivity and redshift coverage spanning  $z \sim 0.2$  to  $z \sim 3$ , SKA-Mid will enable precise measurements of large-scale HI clustering during the post-reionization era. This makes it an ideal testbed for applying marked power spectrum techniques to observationally motivated scenarios. Our goal here is to demonstrate the application of density-dependent weighting schemes to the HI field, but now within realistic SKA-Mid forecast simulations that include instrumental effects such as beam smoothing and thermal noise.

To perform this exercise, we use a simple functional form of the mark first used in [White \(2016\)](#), as

$$m(\mathbf{x}, t; R, \rho_\star, p) = \left[ \frac{\rho_\star + 1}{\rho_\star + \rho_R(\mathbf{x}, t)} \right]^p \quad (20)$$

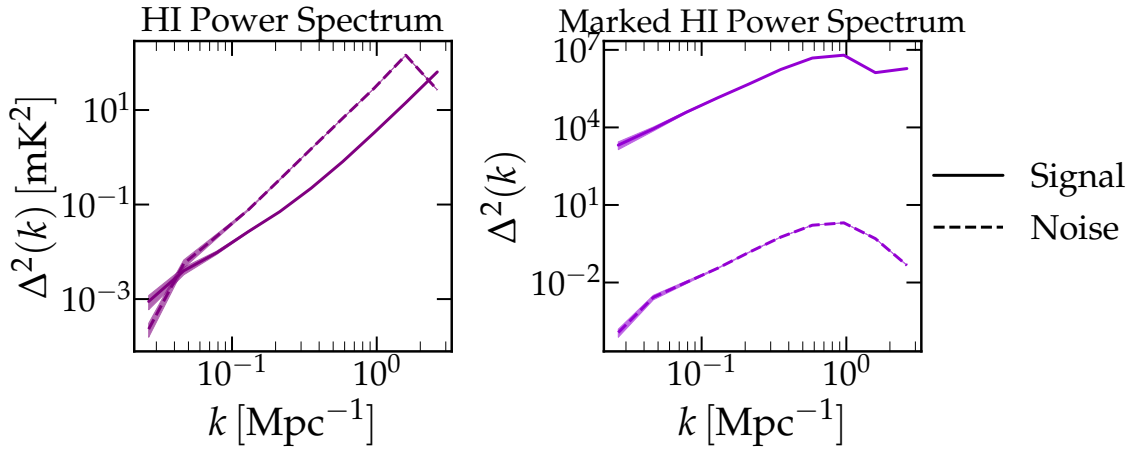
where the free parameters  $R$ ,  $\rho_\star$ , and  $p$  are referred to as the mark parameters. Both  $\rho_R$  and  $\rho_\star$  are expressed in units of the mean density  $\bar{\rho}$ . The parameter  $\rho_R$  quantifies the sensitivity of the mark to the local density around each point. The exponent  $p$  introduces an additional contrast between points residing in different environments by weighting them differently.

As mentioned earlier, suitable marks can selectively highlight distinct environments in the map. It



**Figure 7:** The dimensionless power spectrum of the matter (left panel), the H I (middle panel), and the H I with density values  $\rho/\bar{\rho} < 3.0$  (right panel). The distribution is smoothed for a fixed radius of  $R = 280$  kpc. The power spectra are demonstrated as a function of  $k$ -mode at five different redshifts in the range  $z \sim 1$ –5. The shaded regions in each panel represent  $\pm 1\sigma$  spread around the mean value at  $z = 3$ .

is possible to construct a functional form of the mark that boosts the signal strength. To validate this, we simulate mock LIM noise maps of [HI] 21-cm emission lines at  $z = 1$ , for SKA-Mid AA4 configuration with 5000 hours of integration time. Modeling of the [HI] 21-cm signal is done by postprocessing the IllustrisTNG simulation subhalo catalog as done in (Dosibhatla et al., 2026) using an analogous functional form of the mark as given in Figure 7 for the HI 21-cm signal. To boost the signal strength, we fix the mark parameters to up-weight the high density points in the maps at a relatively small smoothing length scale of  $R = 3$  Mpc. In Figure 8, we compare the strength of signal fluctuation as a function using the marked power spectrum with their corresponding noise power spectrum. The signal and the noise are marked and up-weighted in the same way. As clearly observed, a suitable mark is able to boost the signal strength significantly, which is evident from the marked power spectrum as compared to the standard power spectrum.



**Figure 8:** Comparing the signal strength, at different  $k$ -modes, against the corresponding noise levels using marked statistics designed to boost the signal strength at redshift  $z = 1$  with the standard signal power spectrum. *Left panel:* The dimensionless HI power spectrum and mark power spectrum of the signal fluctuation and noise. *Right panel:* The dimensionless HI power spectrum and marked field power spectrum of the signal and noise. The shaded region corresponds to  $\pm 1\sigma$  sigma spread around the mean.

### 3.4 Three-point correlations and beyond

The Alcock–Paczynski (AP) effect imprints anisotropic geometric distortions on higher-order statistics by differentially rescaling transverse and radial Fourier components when an incorrect distance–redshift mapping is assumed. This anisotropic mapping reshapes triangle (bispectrum) and tetrahedral (trispectrum) configurations in  $k$ -space, altering both the amplitudes and angular dependences of measured bispectra and trispectra and producing multipole mixing that must be modeled to avoid biased cosmological inferences (Karagiannis et al., 2022; Philcox et al., 2022; Randrianjanahary et al., 2024). For the bispectrum, the AP distortion rescales the three triangle sides and mixes bispectrum multipoles; robust inference therefore requires joint modeling of AP, redshift-space distortions, bias, and survey effects (beam, window, and foreground mode loss) so that geometric information [e.g.  $D_A(z)$ ,  $H(z)$  or the AP scaling parameters  $\alpha_{\parallel}\alpha_{\perp}$ ] can be disentangled from systematic contributions. For the trispectrum, the situation is yet more complex—the richer set of shape degrees of freedom (quadrilaterals) means AP both rescales and can change shape classification, and explicit trispectrum with AP treatments remain comparatively sparse in the 21-cm intensity-mapping literature. The 21-cm experiments, such as SKA-Mid, therefore need to include AP transformations into their higher-order statistics analysis pipeline to extract geometric constraints while marginalizing over astrophysical and instrumental nuisance parameters.

In this context, both primordial (PNG) and induced sources of non-Gaussianity are important. From the primordial side, non-zero bispectra and trispectra can arise from inflationary physics, and the 21-cm signal at large scales can retain imprints of such early-Universe signatures. For example, the local-type PNG generates a characteristic scale-dependent bias in the bispectrum that can be isolated with appropriate triangle configurations (Karagiannis et al., 2018). On the other hand, induced non-Gaussianity arises from late-time structure formation, contributes significantly to the bispectrum and trispectrum at intermediate and small scales. SKA-Mid is expected to shed light on both of them.

Beyond higher-order statistics such as bispectrum and trispectra, modern machine learning offers the opportunity to learn statistics that capture non-Gaussian features, and to do so directly from 21-cm maps. Here, a compressor learns a summary that is optimal in the sense of providing unbiased and precise posteriors for a given inference task. Optionally, the summary can be Gaussianised to more closely resemble power spectra as summary statistics. The performance of the learned statistics can then be judged, for example, via simulation-based calibration of posteriors, or measurements of mutual information (Makinen et al., 2024). This approach has been shown, for example, to infer properties on inflation and dark matter from 21-cm maps, and provide more robust results than fixed summaries when faced with noise and modeling uncertainties (Ore et al., 2025; Schosser et al., 2025; Andrianomena and Hassan, 2025). Such methods can certainly be explored with SKA-Mid-produced 21-cm maps.

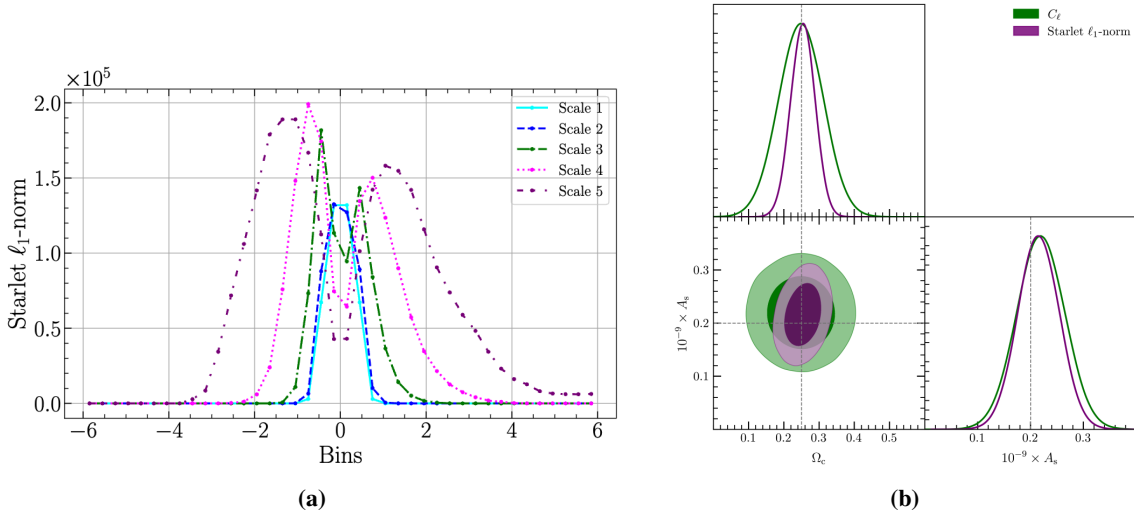
### 3.5 Starlet $\ell_1$ -norm

Figure 9 shows the  $\ell_1$ -norm statistic of the starlet coefficients as a function of the bin threshold for five decomposition scales of the HI intensity map (left), alongside the posterior distributions for  $\Omega_c$  and  $10^{-9}A_s$  inferred via simulation-based inference (right). The green and purple contours

correspond to the angular power spectrum  $C_\ell$  and the starlet  $\ell_1$ -norm statistic, respectively, with shaded regions indicating the 68% and 95% confidence intervals. Inference is performed on the three parameters  $(\Omega_c, 10^{-9}A_s, h)$ , of which only the first two are shown, as  $h$  is weakly constrained (Gorbachev et al., 2026).

The starlet  $\ell_1$ -norm statistic produces significantly tighter posterior contours than the angular power spectrum, highlighting its enhanced sensitivity to cosmological information. This improvement arises from its multiscale, higher-order nature, which captures non-Gaussian features of the intensity field that are inaccessible to two-point statistics alone.

The results are based on lognormal simulations generated with GLASS (Tessore et al., 2023) and CAMB (<https://camb.info>), analyzed using a likelihood-free SBI pipeline implemented in JAXILI (<https://jaxili.readthedocs.io/en/latest/>), which efficiently handles the intractability of the likelihood of the starlet  $\ell_1$ -norm statistic. The maps correspond to full-sky simulations with a sky fraction of  $f_{\text{sky}} = 0.19$ , defined by the applied mask, at a resolution of  $N_{\text{side}} = 1024$  and at redshift  $z = 0.425$ . They include instrumental noise and are smoothed with a 30 arcmin Gaussian beam, and do not include any foreground contaminants. The starlet decomposition is performed over five scales with effective sizes of [3.7, 5.6, 11, 22, 60] arcmin.



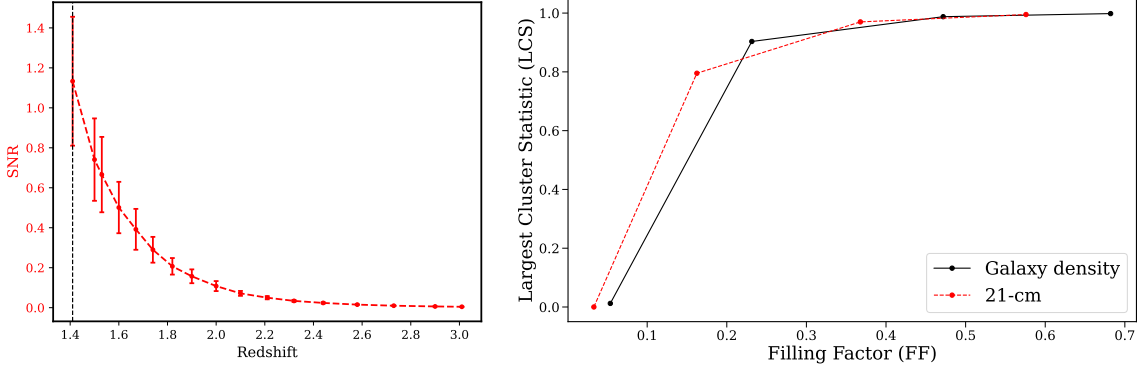
**Figure 9:** **Left:**  $\ell_1$ -norm of the starlet coefficients as a function of the bin threshold for five decomposition scales of the HI intensity map. **Right:** Posterior distributions for  $\Omega_c$  and  $10^{-9} A_s$  inferred with simulation-based inference (SBI). The green contours and curves correspond to the power spectrum  $C_\ell$ , and the purple ones to the Starlet  $\ell_1$ -norm. Contours indicate the 68% and 95% confidence regions. The inference was performed on three parameters  $(\Omega_c, 10^{-9}A_s, h)$ , of which only two are displayed, as  $h$  is weakly constrained (Gorbachev et al., 2026).

### 3.6 Morphological Measures in the Image Domain

Imaging line intensity maps at sub-Mpc resolutions can reveal the intricate cosmic web structure of the universe. Summary statistics in the image domain can be computed on these maps to extract the non-Gaussian and morphological information from these fields. These summary statistics can,

in principle, tighten constraints and break degeneracies of astrophysical as well as cosmological parameters due to their higher information content.

However, obtaining maps at such a high spatial and hence angular resolution is a significant challenge due to foreground contamination coupled with instrumental systematics. Instrumental noise also washes out the information of the LSS in intensity maps. Imaging at high resolution requires information from long baselines, which are sparsely distributed. Besides, a high line-of-sight resolution requires a high spectral resolution, which in turn boosts the thermal noise contamination. In this section, we refer to the SKA-Mid AA4 setup with frequency coverage from 0.58 GHz to 15 GHz as SKA-Mid, and the same array with frequency coverage extended to 50 GHz as SKA Phase 2. The left panel of Figure 10 shows the SNR estimates for mock 21-cm maps observed by SKA-Mid in the redshift range where SKA Phase 2 can observe the CO(1-0) line (Dosibhatla et al., 2026; Sarkar et al., 2026) in the presence of thermal noise only. In this section, the signal maps are modeled by postprocessing IllustrisTNG simulation subhalo catalogs as per Dosibhatla et al. (2026).



**Figure 10: Left:** Signal-to-noise ratios (SNR) of the 21-cm brightness temperature fluctuations at spatial resolution  $\delta x \approx 1$  Mpc in the overlapping redshift range where SKA-Mid and SKA Phase 2 can observe the 21-cm and CO(1-0) LIM signals, respectively. The noise estimates are made assuming Gaussian random thermal noise for 5000 hours per pointing of SKA-Mid observations. The black dashed line corresponds to  $z = 1.41$ , the redshift where the 21-cm maps just become signal-dominant. The  $1\sigma$  error bars denote standard deviation in SNR across  $8 (150 \text{ Mpc})^3$  subcubes each corresponding to a  $4 \text{ deg}^2$  SKA-Mid survey, due to cosmic variance in signal fluctuations. **Right:** Percolation curves for the galaxy mass density (solid black) and 21-cm brightness temperature (dashed red) maps at a spatial resolution of 0.5 Mpc.

It is seen that even with 5000 hours of observation time at a channel-averaged spectral resolution of eight SKA-Mid channels, the 21-cm maps are highly noise-dominated. In this chapter, image analysis results are shown at  $z = 1.41$ , where the 21-cm SNR approaches 1.

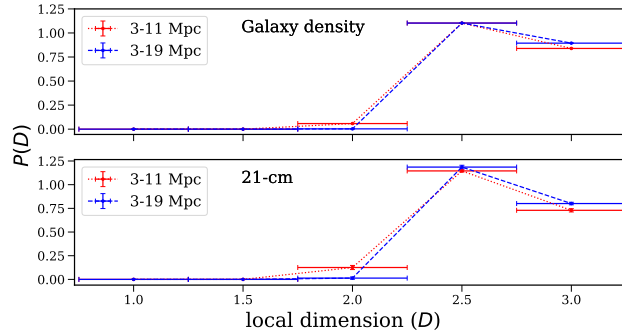
### 3.6.1 Largest Cluster Statistic

The value of FF at which percolation occurs is expected to differ for different maps. The analysis of Bharadwaj et al. (2000) implies that fields with a higher degree of overall filamentarity percolate at lower filling factors. Thus, percolation analysis can distinguish between astrophysical or cosmological models that induce different degrees of filamentarity in an intensity map.

The right panel of Figure 10 shows the percolation curves for the galaxy density field and 21-cm intensity maps at  $z = 1.41$ , simulated from the same underlying galaxy distribution. The FF and LCS values are computed after every iteration of the coarse-graining scheme mentioned in section 2.5.1. As the two maps follow different percolation curves despite being tracers of the same galaxy distribution, the LCS is sensitive to the complex astrophysics of line emissions. For more details, see Dosibhatla et al. (2026). Due to its sensitivity to factors that affect the connectivity and filamentarity of the LSS, LCS can be an effective higher-order summary statistic to infer both astrophysics and cosmology.

### 3.6.2 Local Dimension

Once the local dimensions of a subset of cells in an intensity map are determined as in 2.5.2, the fraction of cells lying in filaments, sheets, and volume-filling environments (nodes and voids) can be computed. Figure 11 shows the distribution of local dimensions of the galaxy density field and 21-cm intensity maps divided into bins of width 0.5. The maps have a spatial resolution of  $\delta s = 107.52$  kHz, corresponding to a spectral resolution of  $\delta \nu = 107.52$  kHz (channel-averaged spectral resolution of eight SKA-Mid channels). The analysis presented here only accounts for cosmic variance. The detectability of the local dimensions is studied in the presence of thermal noise in Dosibhatla et al. (2026). Filamentary features in intensity maps are captured by large  $k_{\parallel}$  modes, which are less foreground-dominated than small  $k_{\parallel}$  modes lying in the foreground wedge. Therefore, the detectability of the local dimensions is not expected to be significantly impacted by residuals of continuum foregrounds.



**Figure 11:** Distribution of local dimensions of classifiable cells out of  $10^5$  randomly sampled bright cells in the galaxy density (top) and  $T_{21\text{cm}}$  (bottom) maps simulated with galaxy positions in redshift space. The  $D$  values are binned into intervals of size 0.5. The different curves correspond to different length scales specified by the  $R_{\min}$  and  $R_{\max}$  values. The vertical  $1\sigma$  error bars denote standard deviation in  $P(D)$  across  $8 (150\text{Mpc})^3$  21-cm subcubes, each corresponding to a  $4\text{ deg}^2$  SKA-Mid survey for  $10^4$  randomly sampled bright cells.

As gravitational clustering makes the late-time matter field highly non-Gaussian, and it also leads to the formation of filaments, sheets, nodes, and voids, identifying these building blocks can yield information on the non-Gaussianity in the matter field. The local dimension distribution can also distinguish between different astrophysical scenarios. Therefore, local dimension can be employed as a higher-order summary statistic to constrain astrophysics and also cosmology, as different cosmological models lead to different gravitational clustering patterns.

## 4 Discussion

In this chapter, we explored and demonstrated the detectability of various higher-order statistics of the 21-cm LIM, beyond the traditional two-point correlation function and power spectrum, using the future SKA-Mid. We observe that different higher-order statistics exhibit complementary sensitivities to cosmological parameters, physical processes, and systematic effects, as they can quantify the inherent non-Gaussianity in the signal, which is mostly driven by the non-linearity in the matter distribution. While some one-point statistics can effectively break degeneracies in cosmological parameters and constrain the HI luminosity function, they are susceptible to calibration uncertainties. On the other hand, the bispectrum can potentially provide shape-dependent constraints on HI-dark matter bias parameters and redshift-space distortions, whereas image-domain morphological statistics can, in principle, directly capture the topology of the cosmic web, although this is highly dependent on the resolution of the 21-cm maps. This effectively suggests that joint constraints by multiple higher-order statistics might enhance the reliability and robustness of these constraints. Such joint constraints might also help validate that the inferred parameters are not artifacts of particular systematic effects.

While higher-order statistics offer significantly enhanced constraining power, they also present substantial challenges: theoretical and numerical modeling becomes increasingly complex and often requires forward-modeling approaches or emulator-based pipelines calibrated on large-volume simulations. Covariance estimation is computationally expensive and often requires either a large ensemble of mock observations or analytical approximations with perturbative corrections. Moreover, systematics such as beam effects, foreground residuals, and survey geometry must be carefully accounted for to avoid bias in parameter inference exercises using 21-cm maps.

Despite these challenges, the information gained from including higher-order statistics in the inference pipeline is compelling. Once next-generation experiments, such as SKA-Mid, are operational, these techniques will become not just useful but necessary for achieving the full scientific potential of 21-cm LIM. This chapter consolidates current efforts within the SKA-LIM working group to advance these methodologies and provides a roadmap for future studies aiming to incorporate higher-order statistics into the LIM cosmology pipeline.

### Author List Ordering

Authors for this chapter are ordered according to their overall level of contribution, in line with that expected for a small author list publication.

### Acknowledgment

SM, MMD, LN, and AD acknowledge financial support through the Core Research Grant titled “Observing the Cosmic Dawn in Multicolour using Next Generation Telescopes” from the Science and Engineering Research Board (SERB) and the Department of Science and Technology (DST), Government of India. SM, MV, AD, LN, SKP and MMD acknowledge financial support through the project titled “Illuminating the Dark Sector of the Cosmos in the SKA Era” (Project No. P3497)

funded under the ‘‘Scheme for Promotion of Academic and Research Collaboration (SPARC)’’ from the Ministry of Education, India. LN acknowledges support from the Abdus Salam International Centre for Theoretical Physics (ICTP) under the ‘ICTP Sandwich Training Educational Programme (STEP)’ SMR.3991 and SMR.4129. SKP, LN, and YM acknowledge the financial support by the Department of Science and Technology, Government of India, through the INSPIRE Fellowship. DS acknowledges the support of the Canada 150 Chairs program, the Fonds de recherche du Québec Nature et Technologies (FRQNT) and the Natural Sciences and Engineering Research Council of Canada (NSERC) joint NOVA grant, and the Trottier Space Institute Postdoctoral Fellowship program. BVG and CU were supported by the European Union (ERC StG, LSS\_BeyondAverage, 101075919). JLB acknowledges funding from the grant UC-LIME (PID2022-140670NA-I00), financed by MCIN/AEI/ 10.13039/501100011033/FEDER, UE. AKS acknowledges the support from the National Science Foundation (grant No. 2206602). CH’s work is funded by the Volkswagen Foundation, and supported through Germany’s Excellence Strategy through EXC~2181/1 – 390900948 (the Heidelberg STRUCTURES Excellence Cluster) and with support by the Federal Ministry of Education and Research (BMBF) and the Ministry of Science, Research and the Arts of Baden-Württemberg. PG’s work was supported by the TITAN ERA Chair project (contract no. 101086741) within the Horizon Europe Framework Program of the European Commission and the Agence Nationale de la Recherche (ANR-22-CE31-0014-01 TOSCA).

## References

- S. Andrianomena and S. Hassan. *Astrophys. Space Sci.*, 370(2):14, Feb. 2025. doi: 10.1007/s10509-025-04405-y.
- M. A. Aragón-Calvo, R. van de Weygaert, and B. J. T. Jones. *Monthly Notices of the Royal Astronomical Society*, 408(4):2163–2187, 10 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.17263.x. URL <https://doi.org/10.1111/j.1365-2966.2010.17263.x>.
- S. Avila et al. *Monthly Notices of the Royal Astronomical Society*, 510(1):292–308, 11 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab3406. URL <https://doi.org/10.1093/mnras/stab3406>.
- S. Bag et al. *Monthly Notices of the Royal Astronomical Society*, 477(2):1984–1992, 2018. doi: 10.1093/mnras/sty714. URL <http://dx.doi.org/10.1093/mnras/sty714>.
- R. Battye et al. arXiv preprint arXiv:1209.1041, 2012.
- R. A. Battye et al. *MNRAS*, 434(2):1239–1256, 07 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt1082. URL <https://doi.org/10.1093/mnras/stt1082>.
- A. E. Bayer et al. *Astrophys. J.*, 919(1):24, Sept. 2021. doi: 10.3847/1538-4357/ac0e91.
- J. L. Bernal. *Phys. Rev. D*, 109(4):043517, 2024. doi: 10.1103/PhysRevD.109.043517.
- J. L. Bernal and E. D. Kovetz. *Astron. Astrophys. Rev.*, 30(1):5, Dec. 2022. doi: 10.1007/s00159-022-00143-0.
- S. Bharadwaj and S. K. Sethi. *Journal of Astrophysics and Astronomy*, 22(4):293–307, Dec. 2001. doi: 10.1007/BF02702273.
- S. Bharadwaj et al. *The Astrophysical Journal*, 528(1):21, jan 2000. doi: 10.1086/308163. URL <https://dx.doi.org/10.1086/308163>.

- S. Bharadwaj, B. B. Nath, and S. K. Sethi. *Journal of Astrophysics and Astronomy*, 22(1):21–34, Mar. 2001. doi: 10.1007/BF02933588.
- S. Bharadwaj, A. Mazumdar, and D. Sarkar. *Mon. Not. Roy. Astron. Soc.*, 493(1):594–602, 2020. doi: 10.1093/mnras/staa279.
- J. R. Bond, L. Kofman, and D. Pogosyan. *Nature*, 380(6575):603–606, Apr. 1996. doi: 10.1038/380603a0.
- P. C. Breyse. *arXiv e-prints*, art. arXiv:2209.01223, Sept. 2022. doi: 10.48550/arXiv.2209.01223.
- P. C. Breyse et al. *Mon. Not. Roy. Astron. Soc.*, 467(3):2996–3010, 2017. doi: 10.1093/mnras/stx203.
- P. C. Breyse, C. J. Anderson, and P. Berger. *Phys. Rev. Lett.*, 123(23):231105, 2019. doi: 10.1103/PhysRevLett.123.231105.
- P. C. Breyse, D. T. Chung, and H. T. Ihle. *Mon. Not. Roy. Astron. Soc.*, 525(2):1824–1838, 2023. doi: 10.1093/mnras/stad2350.
- P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos. *Astrophys. J.*, 803(1):21, 2015. doi: 10.1088/0004-637X/803/1/21.
- S. Camera et al. *PoS, AASKA14:025*, 2015. doi: 10.22323/1.215.0025.
- E. Castorina and F. Villaescusa-Navarro. *Mon. Not. Roy. Astron. Soc.*, 471(2):1788–1796, 2017. doi: 10.1093/mnras/stx1599.
- E. Castorina and M. White. *J. Cosmol. Astropart. Phys.*, 2019(6):025, June 2019. doi: 10.1088/1475-7516/2019/06/025.
- Z. Chen et al. *Astrophys. J. Suppl. Ser.*, 279(1):19, July 2025. doi: 10.3847/1538-4365/add897.
- M. Chhabra and S. Bharadwaj. *arXiv e-prints*, art. arXiv:2508.19126, Aug. 2025. doi: 10.48550/arXiv.2508.19126.
- C.-H. Chuang et al. *Monthly Notices of the Royal Astronomical Society*, 487(1):48–59, 05 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1233. URL <https://doi.org/10.1093/mnras/stz1233>.
- D. T. Chung et al. *Mon. Not. Roy. Astron. Soc.*, 520(4):5305–5316, 2023. doi: 10.1093/mnras/stad359.
- Cosmic Visions 21 cm Collaboration et al. *arXiv e-prints*, art. arXiv:1810.09572, Oct. 2018. doi: 10.48550/arXiv.1810.09572.
- D. J. Croton et al. *The Astrophysical Journal Supplement Series*, 222(2):22, feb 2016. doi: 10.3847/0067-0049/222/2/22. URL <https://dx.doi.org/10.3847/0067-0049/222/2/22>.
- S. Cunnington, C. Watkinson, and A. Pourtsidou. *Mon. Not. R. Astron. Soc.*, 507(2):1623–1639, Oct. 2021. doi: 10.1093/mnras/stab2200.
- S. Dasgupta et al. *Journal of Cosmology and Astroparticle Physics*, 2023(05):014, may 2023. doi: 10.1088/1475-7516/2023/05/014. URL <https://dx.doi.org/10.1088/1475-7516/2023/05/014>.
- C. B. V. Dash and T. G. Sarkar. *JCAP*, 01:070, 2025. doi: 10.1088/1475-7516/2025/01/070.
- C. B. V. Dash, T. G. Sarkar, and A. A. Sen. *Mon. Not. Roy. Astron. Soc.*, 527(4):11694–11706, 2023. doi: 10.1093/mnras/stad3946.
- G. De Lucia et al. *Mon. Not. R. Astron. Soc.*, 445(1):970–987, Nov. 2014. doi: 10.1093/mnras/stu1752.
- G. De Lucia, F. Fontanot, L. Xie, and M. Hirschmann. *Astron. Astrophys.*, 687:A68, July 2024.

- doi: 10.1051/0004-6361/202349045.
- M. M. Dosibhatla et al. *J. Cosmol. Astropart. Phys.*, 2026(6):014, June 2026. doi: 10.1088/1475-7516/2026/06/014.
- D. A. Dunne et al. *Astron. Astrophys.*, 702:A247, Oct. 2025. doi: 10.1051/0004-6361/202554780.
- J. Einasto, A. A. Klypin, E. Saar, and S. F. Shandarin. *Monthly Notices of the Royal Astronomical Society*, 206(3):529–558, 02 1984. ISSN 0035-8711. doi: 10.1093/mnras/206.3.529. URL <https://doi.org/10.1093/mnras/206.3.529>.
- T. Flöss, M. Biagetti, and P. D. Meerburg. *Phys. Rev. D*, 107(2):023528, Jan. 2023. doi: 10.1103/PhysRevD.107.023528.
- F. Fontanot et al. *Astron. Astrophys.*, 699:A108, 2025. doi: 10.1051/0004-6361/202452029.
- O. Friedrich et al. *Mon. Not. R. Astron. Soc.*, 498(1):464–483, Oct. 2020. doi: 10.1093/mnras/staa2160.
- S. Furlanetto, S. P. Oh, and F. Briggs. *Phys. Rept.*, 433:181–301, 2006. doi: 10.1016/j.physrep.2006.08.002.
- S. R. Furlanetto et al. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010, page 82, Jan. 2009. doi: 10.48550/arXiv.0902.3259.
- P. Gorbachev, J.-L. Starck, S. Camera, and M. Spinelli. 5 2026.
- S. A. Gregory and L. A. Thompson. *Astrophys. J.*, 222:784–799, June 1978. doi: 10.1086/156198.
- C. Hahn, F. Villaescusa-Navarro, E. Castorina, and R. Scoccimarro. *J. Cosmol. Astropart. Phys.*, 2020(3):040, Mar. 2020. doi: 10.1088/1475-7516/2020/03/040.
- O. Hahn, C. Porciani, C. M. Carollo, and A. Dekel. *Monthly Notices of the Royal Astronomical Society*, 375(2):489–499, Feb. 2007. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.11318.x. URL <https://doi.org/10.1111/j.1365-2966.2006.11318.x>.
- Hartlap, J., Simon, P., and Schneider, P. *A&A*, 464(1):399–404, 2007. doi: 10.1051/0004-6361:20066170. URL <https://doi.org/10.1051/0004-6361:20066170>.
- M. Hirschmann, G. De Lucia, and F. Fontanot. *Mon. Not. R. Astron. Soc.*, 461(2):1760–1785, Sept. 2016. doi: 10.1093/mnras/stw1318.
- H. T. Ihle et al. *Astrophys. J.*, 871(1):75, 2019. doi: 10.3847/1538-4357/aaf4bc.
- M. H. Jalali Kanafi and S. M. S. Movahed. *Phys. Rev. D*, 113(6):063543, Mar. 2026. doi: 10.1103/yw4d-vssh.
- M. Kamran, M. Sahlén, D. Sarkar, and S. Majumdar. *J. Cosmol. Astropart. Phys.*, 2025(7):054, July 2025. doi: 10.1088/1475-7516/2025/07/054.
- D. Karagiannis et al. *Mon. Not. R. Astron. Soc.*, 478(1):1341–1376, July 2018. doi: 10.1093/mnras/sty1029.
- D. Karagiannis, R. Maartens, and L. F. Randrianjanahary. *JCAP*, 11:003, 2022. doi: 10.1088/1475-7516/2022/11/003.
- G. M. Kaufman. Report no. 6710. Technical report, Center for Operations Research and Econometrics, Catholic University of Louvain, Heverlee, Belgium, 1967.
- A. Klypin and S. F. Shandarin. *Astrophys. J.*, 413:48, Aug. 1993. doi: 10.1086/172975.
- E. D. Kovetz et al. *arXiv e-prints*, art. arXiv:1709.09066, Sept. 2017. doi: 10.48550/arXiv.1709.09066.
- O. Leicht et al. *Mon. Not. R. Astron. Soc.*, 484(1):269–281, Mar. 2019. doi: 10.1093/mnras/sty3469.

- P. Lemos et al. In Machine Learning for Astrophysics, page 18, July 2023. doi: 10.48550/arXiv.2310.15256.
- N. I. Libeskind et al. Monthly Notices of the Royal Astronomical Society, 473(1):1195–1217, 08 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx1976. URL <https://doi.org/10.1093/mnras/stx1976>.
- M. Lujan Niemeyer et al. Astrophys. J., 929(1):90, 2022. doi: 10.3847/1538-4357/ac5cb8.
- T. L. Makinen et al. arXiv e-prints, art. arXiv:2410.07548, Oct. 2024. doi: 10.48550/arXiv.2410.07548.
- E. Massara et al. The Astrophysical Journal, 951(1):70, July 2023. ISSN 0004-637X. doi: 10.3847/1538-4357/acd44d. URL <https://dx.doi.org/10.3847/1538-4357/acd44d>. Publisher: The American Astronomical Society.
- A. Mazumdar, S. Bharadwaj, and D. Sarkar. Mon. Not. Roy. Astron. Soc., 498(3):3975–3984, 2020. doi: 10.1093/mnras/staa2548.
- B. McCarthy Gould, L. Castiblanco, C. Uhlemann, and O. Friedrich. The Open Journal of Astrophysics, 8:1, Jan. 2025. doi: 10.33232/001c.127800.
- A. J. Mead, S. Brieden, T. Tröster, and C. Heymans. Monthly Notices of the Royal Astronomical Society, 502(1):1401–1422, 01 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab082. URL <https://doi.org/10.1093/mnras/stab082>.
- K. R. Mecke, T. Buchert, and H. Wagner. Robust Morphological Measures for Large-Scale Structure in the Universe, Dec. 1993. URL <http://arxiv.org/abs/astro-ph/9312028>. arXiv:astro-ph/9312028.
- C. Modi et al. Mon. Not. R. Astron. Soc., 536(1):254–265, Jan. 2025. doi: 10.1093/mnras/stae2473.
- L. Noble et al. arXiv e-prints, 4 2026.
- A. Ore, C. Heneka, and T. Plehn. SciPost Physics, 18(5):155, May 2025. doi: 10.21468/SciPostPhys.18.5.155.
- H. Padmanabhan, R. Maartens, O. Umeh, and S. Camera. arXiv e-prints, art. arXiv:2305.09720, May 2023. doi: 10.48550/arXiv.2305.09720.
- S. Pal, D. Sarkar, and A. Aviles. arXiv e-prints, art. arXiv:2508.19106, Aug. 2025. doi: 10.48550/arXiv.2508.19106.
- S. Pal, D. Sarkar, R. Samanta, and S. Pal. Mon. Not. Roy. Astron. Soc., 542:1, 2025a. doi: 10.1093/mnras/staf1218.
- S. K. Pal et al. Journal of Cosmology and Astroparticle Physics, 2025(10):096, oct 2025b. doi: 10.1088/1475-7516/2025/10/096. URL <https://doi.org/10.1088/1475-7516/2025/10/096>.
- B. Pandey and S. Sarkar. Monthly Notices of the Royal Astronomical Society, 498(4):6069–6082, 09 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa2772. URL <https://doi.org/10.1093/mnras/staa2772>.
- V. Pandya et al. 5 2026.
- A. Pathak et al. Journal of Cosmology and Astroparticle Physics, 2022(11):027, nov 2022. doi: 10.1088/1475-7516/2022/11/027. URL <https://dx.doi.org/10.1088/1475-7516/2022/11/027>.
- O. H. E. Philcox et al. Phys. Rev. D, 106(4):043530, 2022. doi: 10.1103/PhysRevD.106.043530.

- R. F. Pinheiro, A. A. Costa, and Y. Sang. *arXiv e-prints*, art. arXiv:2603.20160, Mar. 2026. doi: 10.48550/arXiv.2603.20160.
- Planck Collaboration. *Astron. Astrophys.*, 594:A13, Sept. 2016. doi: 10.1051/0004-6361/201525830.
- J. R. Pritchard and A. Loeb. *Phys. Rev. D*, 78:103511, 2008. doi: 10.1103/PhysRevD.78.103511.
- J. R. Pritchard and A. Loeb. *Reports on Progress in Physics*, 75(8):086901, Aug. 2012. doi: 10.1088/0034-4885/75/8/086901.
- J. R. Pritchard and A. Loeb. *Reports on Progress in Physics*, 75(8):086901, 2012.
- L. F. Randrianjanahary, D. Karagiannis, and R. Maartens. *Phys. Dark Univ.*, 45:101530, 2024. doi: 10.1016/j.dark.2024.101530.
- E. Regós et al. *The Astrophysical Journal*, 974(1):126, oct 2024. doi: 10.3847/1538-4357/ad7138. URL <https://dx.doi.org/10.3847/1538-4357/ad7138>.
- V. I. Sabla, J. L. Bernal, G. Sato-Polito, and M. Kamionkowski. *Phys. Rev. D*, 110(2):023507, 2024. doi: 10.1103/PhysRevD.110.023507.
- D. Sarkar, S. Bharadwaj, and S. Anathpindika. *Mon. Not. Roy. Astron. Soc.*, 460(4):4310–4319, 2016. doi: 10.1093/mnras/stw1111.
- D. Sarkar, S. Majumdar, and S. Bharadwaj. *Mon. Not. Roy. Astron. Soc.*, 490(2):2880–2889, 2019a. doi: 10.1093/mnras/stz2799.
- D. Sarkar, S. Majumdar, and S. Bharadwaj. *Mon. Not. Roy. Astron. Soc.*, 490(2):2880–2889, 2019b. doi: 10.1093/mnras/stz2799.
- D. Sarkar et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Sarkar01.
- P. Sarkar and S. Bharadwaj. *Monthly Notices of the Royal Astronomical Society: Letters*, 394(1): L66–L69, 03 2009. ISSN 1745-3925. doi: 10.1111/j.1745-3933.2008.00612.x. URL <https://doi.org/10.1111/j.1745-3933.2008.00612.x>.
- P. Sarkar, B. Pandey, and S. Bharadwaj. *Monthly Notices of the Royal Astronomical Society*, 423(1):955–961, 05 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.20932.x. URL <https://doi.org/10.1111/j.1365-2966.2012.20932.x>.
- S. Sarkar and B. Pandey. *Monthly Notices of the Royal Astronomical Society*, 485(4):4743–4753, 03 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz499. URL <https://doi.org/10.1093/mnras/stz499>.
- G. Sato-Polito and J. L. Bernal. *Phys. Rev. D*, 106(10):103534, 2022. doi: 10.1103/PhysRevD.106.103534.
- B. Schosser, C. Heneka, and B. M. Schäfer. *arXiv e-prints*, art. arXiv:2508.10094, Aug. 2025. doi: 10.48550/arXiv.2508.10094.
- S. Seethapuram Sridhar et al. Ska low and mid subarray templates, June 2025. URL <https://doi.org/10.5281/zenodo.16951088>.
- S. F. Shandarin. *Soviet Astronomy Letters*, 9:104–106, Feb. 1983.
- G. Sun et al. *J. Cosmol. Astropart. Phys.*, 2026(2):008, Feb. 2026. doi: 10.1088/1475-7516/2026/02/008.
- J. Sunseri, A. E. Bayer, and J. Liu. *Phys. Rev. D*, 112(6):063516, Sept. 2025. doi: 10.1103/grx3-hj7w.
- N. Tessore et al. *The Open Journal of Astrophysics*, 6:11, Mar. 2023. doi: 10.21105/astro.2302.

01942.

- R. Tojeiro and K. Kraljic. arXiv e-prints, art. arXiv:2503.21759, Mar. 2025. doi: 10.48550/arXiv.2503.21759.
- C. Uhlemann et al. Mon. Not. R. Astron. Soc., 460(2):1529–1541, Aug. 2016. doi: 10.1093/mnras/stw1074.
- F. Villaescusa-Navarro, M. Viel, K. K. Datta, and T. R. Choudhury. JCAP, 09:050, 2014. doi: 10.1088/1475-7516/2014/09/050.
- M. White. Journal of Cosmology and Astroparticle Physics, 2016(11):057–057, Nov. 2016. ISSN 1475-7516. doi: 10.1088/1475-7516/2016/11/057. URL <http://dx.doi.org/10.1088/1475-7516/2016/11/057>.
- M. Yadav and T. G. Sarkar. European Physical Journal C, 85(11):1337, Nov. 2025. doi: 10.1140/epjc/s10052-025-15055-3.
- I. B. Zeldovich, J. Einasto, and S. F. Shandarin. Nature, 300(5891):407–413, Dec. 1982. doi: 10.1038/300407a0.