Implicit inference of the reionization history with higher-order statistics of the 21-cm signal

Nicolas Cerardi[ORCID],[1]* Sambit K. Giri[ORCID],[2] Michele Bianco[ORCID],[3] Davide Piras[ORCID],[4] Emmanuel de Salis[ORCID],[5] Massimo De Santis[ORCID],[5] Merve Selcuk-Simsek[ORCID],[6] Philipp Denzel[ORCID],[7] Kelley M. Hess[ORCID],[8] M. Carmen Toribio[ORCID],[8] Franz Kirsten[ORCID],[8] Hatem Ghorbel[ORCID][5]

[1]*Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland*
[2]*Department of Astronomy and Oskar Klein Centre, AlbaNova, Stockholm University, SE-10691 Stockholm, Sweden*
[3]*Institute for Particle Physics & Astrophysics (ETHZ), Wolfgang-Pauli-Str 27, 8093 Zurich, Switzerland*
[4]*Département de Physique Théorique, Université de Genève, 24 quai Ernest Ansermet, 1211 Genève 4, Switzerland*
[5]*Haute Ecole Arc Ingénierie, University of Applied Sciences & Arts Western Switzerland (HES-SO), Saint-Imier, Switzerland*
[6]*Institute for Data Science, University of Applied Sciences & Arts Northwestern Switzerland (FHNW), Bahnhofstrasse 6, Windisch, 5210, Switzerland*
[7]*Centre for Artificial Intelligence, Zurich University of Applied Sciences (ZHAW), Technikumstrasse 71, 8400 Winterthur, Switzerland*
[8] *Department of Space, Earth and Environment, Chalmers University of Technology, Onsala Space Observatory, SE-43992 Onsala, Sweden*

ABSTRACT

The Epoch of Reionization (EoR), when the first luminous sources ionised the intergalactic medium, represents a new frontier in cosmology. The Square Kilometre Array Observatory (SKAO) will offer unprecedented insights into this era through observations of the redshifted 21-cm signal, enabling constraints on the Universe's reionization history. We investigate the information content of the average neutral hydrogen fraction ($\bar{x}_{HI}$) in several Gaussian (spherical and cylindrical power spectra) and non-Gaussian (Betti numbers and bispectrum) summary statistics of the 21-cm signal. Mock 21-cm observations are generated using the AA* configuration of SKAO's low-frequency telescope, incorporating noise levels for 100 and 1000 hours. We employ a state-of-the-art implicit inference framework to learn posterior distributions of $\bar{x}_{HI}$ in redshift bins centred at $z = 8.0, 7.2$ and 6.5, for each statistic and noise scenario, validating the posteriors through calibration tests. Using the figure of merit to assess constraining power, we find that Betti numbers alone are on average more informative than the power spectra, while the bispectrum provides limited constraints. However, combining higher-order statistics with the cylindrical power spectrum improves the mean figure of merit by ~ 0.25 dex ($\sim 33\%$ reduction in $\sigma(\bar{x}_{HI})$). The relative contribution of each statistic varies with the stage of reionization. With SKAO observations approaching, our results show that combining power spectra with higher-order statistics can significantly increase the information retrieved from the EoR, maximising the scientific return of future 21-cm observations.

Key words: intergalactic medium – dark ages, reionization, first stars – cosmology: theory - galaxies: formation.

1 INTRODUCTION

The Epoch of Reionization (EoR) marks one of the most significant phase transitions in cosmic history. During this period, the first generation of stars and galaxies formed, and their ultraviolet radiation progressively ionised the neutral hydrogen (HI) that filled the intergalactic medium (IGM). This process, which concluded approximately one billion years after the Big Bang, brought an end to the cosmic "dark ages" and fundamentally shaped the universe we observe today. A primary goal of modern cosmology is to chart the timeline of this process, typically parametrised by the evolution of the global neutral hydrogen fraction $\bar{x}_{HI}$. Observations of the Lyman-α forest in high-redshift quasar spectra have suggested that reionization ended around $z \approx 5.3 - 5.5$ (e.g., Fan et al. 2006; Becker et al. 2015; Kulkarni et al. 2019; Bosman et al. 2022), and several complementing constraints from the CMB optical depth (Planck Collaboration et al. 2020) and galaxy observations (e.g., Mason et al. 2018, 2019; Hoag et al. 2019; Bolan et al. 2022; Jones et al. 2025) indicate that it was halfway complete around $z \approx 7$.

While powerful, these existing observational probes offer an indirect view of the IGM during reionization. Quasar absorption spectra, galaxy damping wings, and Lyman-α emitter statistics rely on interpreting the imprint of the IGM on light originating from or passing near luminous sources (e.g., McGreer et al. 2015; Davies et al. 2018; Greig et al. 2022; Curtis-Lake et al. 2023; Umeda et al. 2024). Inferring the global properties of the IGM from these sightlines requires careful modelling of both the sources and the complex radiative transfer of Lyman-α photons through an inhomogeneous medium (e.g., Huberty et al. 2025; Georgiev et al. 2025). Furthermore, inferring global quantities from a handful of sightlines can be prone to cosmic variance. In contrast, the 21-cm signal, a spectral line from the hyperfine spin-flip transition of neutral hydrogen, offers a more direct window into the IGM during the EoR (see e.g., Pritchard & Loeb 2012, for a review). Originating from the neutral hydrogen dis-

* E-mail: nicolas.cerardi@epfl.ch

tributed throughout the cosmic volume, mapping the 21-cm signal across different redshifts can provide a three-dimensional tomography of the IGM, revealing the evolving morphology and topology of the ionised bubbles (e.g., Friedrich et al. 2011; Giri et al. 2018a; Giri & Mellema 2021). This, in turn, allows us to constrain not only the underlying astrophysical processes but also the reionization history directly (e.g., Giri et al. 2018b; Bianco et al. 2021, 2024). However, detecting this faint signal is an extreme observational challenge. It is buried beneath astrophysical foregrounds that are orders of magnitude brighter, requiring exceptionally precise calibration (e.g., Barry et al. 2016; Mevius et al. 2022; Gan et al. 2023). Current-generation radio interferometers, such as the Low Frequency Array (LOFAR; Mertens et al. 2020, 2025), the Murchison Widefield Array (MWA; Trott et al. 2020; Nunhokee et al. 2025), and the Hydrogen Epoch of Reionization Array (HERA; HERA Collaboration et al. 2022, 2023), have made significant progress, placing the first upper limits on the 21-cm power spectrum. These data have already been used to rule out some extreme scenarios, particularly those involving minimal heating of the IGM as reionization proceeds ('cold reionization') which predict a large 21-cm signal amplitude (e.g., Greig et al. 2021b,a; Ghara et al. 2020, 2021, 2025; HERA Collaboration et al. 2023; Nunhokee et al. 2025).

The forthcoming Square Kilometre Array's Low-frequency component (SKA-Low; Koopmans et al. 2015) is poised to revolutionise the field. With its unprecedented sensitivity, SKA-Low is expected to move beyond upper limits to achieve a robust detection and detailed characterisation of the 21-cm signal across cosmic time. A primary scientific objective is to leverage these data to directly reconstruct the reionization history. The cylindrically averaged power spectrum (PS2D) serves as the primary summary statistic for initial analyses, offering robustness against foreground contamination and being the expected first statistical detection (e.g., Liu & Tegmark 2011; Pober et al. 2014). Indeed, forecast studies have demonstrated the potential of the PS2D to constrain the reionization history (e.g., Greig et al. 2024; Cooper et al. 2025; Pietschke et al. 2025). The importance of this statistic and objective was also highlighted in the recent SKA Science Data Challenge 3b (SDC3b[1]), where international teams competed to infer the reionization history from blinded mock PS2D data. Among the participants, the SEarCH (**S**vea's **E**ndeavour in **A**I and **R**eionization **C**osmology with **H**elvetia) team provided two methods (de Salis et al. 2025, and this work), both securing podium positions in the SDC3b rankings.

The 21-cm signal, however, is highly non-Gaussian, reflecting the complex and evolving morphology of ionised regions (e.g., Iliev et al. 2006; Majumdar et al. 2018; Giri et al. 2018a, 2019). The growth of ionised bubbles, a key tracer of the reionization history, introduces significant non-Gaussian features in the signal's morphology and topology. The PS2D, as a two-point statistic, only captures the complete information for a Gaussian field and is therefore insufficient, potentially missing a wealth of information. This strongly motivates the exploration of non-Gaussian information through higher-order summary statistics, such as the bispectrum (a three-point statistic) and topological descriptors such as Betti numbers (e.g., Elbers & Weygaert 2019; Giri & Mellema 2021; Kapahtia et al. 2021), which are explicitly sensitive to the complex morphology of the ionised regions. In this study, we explore the capabilities of the non-Gaussian information on constraining the reionization history.

Building an inference framework capable of optimally extracting information from the 21-cm signal data faces significant hurdles.

Firstly, modelling the EoR and the resulting 21-cm signal is computationally demanding. Even fast semi-numerical codes such as 21CMFAST (Mesinger et al. 2011), GRIZZLY (Ghara et al. 2018) and BEoRN (Schaeffer et al. 2023) require several CPU hours for a single simulation spanning the relevant cosmic time ($12 \lesssim z \lesssim 5$). Traditional Markov chain Monte Carlo (MCMC) frameworks, which require evaluating millions of models to explore the high-dimensional astrophysical parameter space adequately, become computationally prohibitive (e.g., Greig & Mesinger 2015). Secondly, the likelihood function $p(\mathbf{t}|x_{HI}, \theta)$, which connects $\mathbf{t}$, the summary statistic of the observed data, to the reionization state x_{HI} and underlying parameters θ, is often intractable, especially for higher-order statistics or when complex instrumental effects are considered (Prelogović & Mesinger 2023). Moreover, previous work has demonstrated how combining multiple complementary summary statistics can maximise the astrophysical inference from 21-cm data (Prelogović & Mesinger 2024). Employing fast *emulators* trained on simulation suites have been proposed to accelerate the forward modelling of specific summary statistics within MCMC (e.g., Kern et al. 2017; Schmit & Pritchard 2018; Ghara et al. 2020; Breitman et al. 2024; Meriot & Semelin 2024; Choudhury et al. 2024). While this strategy can address the computational cost after an initial training investment, emulators are typically specific to the statistic they were trained on and require retraining for new observables; crucially, they do not inherently solve the problem of an intractable or unknown likelihood function.

Implicit inference, also known as simulation-based inference (SBI), offer a compelling alternative that circumvents the need for an explicit analytical form for the likelihood $p(\mathbf{t}|\theta)$ (see e.g., Cranmer et al. 2020, for a review). These techniques leverage the fact that complex EoR models can generate simulated data instances which carry an implicit likelihood by sampling the joint distribution $(\theta, \mathbf{t})$. By training neural networks on a sufficiently large set of pre-computed simulations, SBI algorithms learn approximations of e.g. the posterior $p(\theta|\mathbf{t})$, the likelihood $p(\mathbf{t}|\theta)$, or the likelihood ratio $p(\mathbf{t}|\theta)/p(\mathbf{t}|\theta_{ref})$, where θ_{ref} is a chosen fixed reference parameter set. These techniques enable the use of complex simulation-based models within inference tasks, ultimately promising to fully exploit the information contained in observations and tackling systematics. As such, implicit inference has been tested for a wide variety of tasks in cosmology and astrophysics (e.g., Vasist et al. 2023; Lanzieri et al. 2025; Cerardi et al. 2025; Saoulis et al. 2025). Implicit inference provides particular advantages for 21-cm cosmology. First, it naturally handles the intractable likelihoods associated with non-Gaussian summary statistics or complex instrumental effects. Furthermore, the methodology efficiently utilises computationally expensive simulations, as the generation of the training set can be parallelised and generally requires significantly fewer simulation runs than a typical MCMC chain needs to converge (e.g., Schmit & Pritchard 2018; Semelin et al. 2025). Consequently, SBI methods are increasingly being applied to 21-cm signal analysis. Studies have employed various summary statistics, including the power spectrum, PDF moments, and wavelet scattering transforms, within implicit inference frameworks to forecast constraints on astrophysical parameters (e.g., Zhao et al. 2022, 2024; Prelogović & Mesinger 2023; Greig et al. 2024; Meriot et al. 2025). Recently, Semelin et al. (2025) have demonstrated the capability of SBI to combine information from multiple, potentially correlated, summary statistics to derive tighter constraints.

In this paper, we study the non-Gaussian information in the 21-cm signal to improve the constraints on reionization history. This paper is structured as follows. In Sec. 2, we introduce the adopted forward model, including semi-analytical simulations, instrumental effects, and the chosen summary statistics for this study. In Sec. 3, we

---

detail our implicit inference framework to derive posterior estimators from single and combined summary statistics. In Sec. 4, we present our results, showing calibration metrics and comparing summary statistics performance. Lastly, in Sec. 5 we discuss our results and conclusions from this work.

2 DATA MODEL

2.1 Simulations

Our inference framework is trained on a large suite of mock 21-cm signal observations. In this section, we describe the numerical simulations used to generate this training set. We first detail the simulation code and the astrophysical parameterisation adopted (Sec. 2.1.1), and then define the three specific reference models used to test our framework (Sec. 2.1.2).

2.1.1 Method

To generate the training data for our SBI framework, we employ the widely-used semi-numerical simulation code 21cmFAST (Mesinger et al. 2011; Murray et al. 2020). 21cmFAST efficiently models the large-scale structure formation using second-order Lagrangian Perturbation Theory (2LPT; e.g., Scoccimarro 1998) and simulates the process of reionization using an excursion-set formalism based on Furlanetto et al. (2004). This approach allows for the rapid generation of large cosmological volumes necessary for capturing the statistical properties of the 21-cm signal. The code self-consistently calculates the evolution of density, velocity, ionization, and temperature fields, enabling the computation of the 21-cm brightness temperature contrast.

The observable 21-cm signal is quantified by the differential brightness temperature δT_b, given as (e.g., Furlanetto et al. 2006):

$$\delta T_b(\mathbf{x},z) \approx 27 x_{\rm HI}(\mathbf{x},z)[1+\delta_b(\mathbf{x},z)]\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1+z}{10}\right)^{1/2}$$
$$\left(1-\frac{T_\gamma(z)}{T_S(\mathbf{x},z)}\right)\left(\frac{H(z)}{dv_r/dr+H(z)}\right){\rm mK} \quad (1)$$

where $x_{\rm HI}(\mathbf{x},z)$ and $\delta_b(\mathbf{x},z)$ are the neutral hydrogen fraction and baryon overdensity, respectively, at position $\mathbf{x}$ and redshift z. $T_S(\mathbf{x},z)$ is the spin temperature, determined by the coupling of the hyperfine levels to the CMB, the kinetic gas temperature (via collisions), and the Lyman-α background (via Wouthuysen-Field effect). $T_\gamma(z)$ is the CMB temperature at redshift z. Ω_b and Ω_m are the baryon and total matter density contrast parameters, respectively. dv_r/dr is the gradient of the peculiar velocity along the line-of-sight, accounting for redshift-space distortions.

For this work, we use the full spin temperature calculation within 21cmFAST, which accounts for collisional coupling, Wouthuysen-Field coupling from Lyman-α photons, and heating from X-ray photons. This provides a more realistic description than assuming the high spin temperature limit ($T_S \gg T_\gamma$), especially at earlier times or in potentially colder regions of the IGM. Although our primary focus is on the later stages of reionization ($z \lesssim 9$), including the full spin temperature evolution ensures our framework can handle variations in thermal state and is robust even if some regions remain relatively cold, allowing us to test the framework's ability to constrain the reionization history across diverse scenarios.

21cmFAST models the properties of galaxies, which drive reionization and IGM heating, using a set of astrophysical parameters
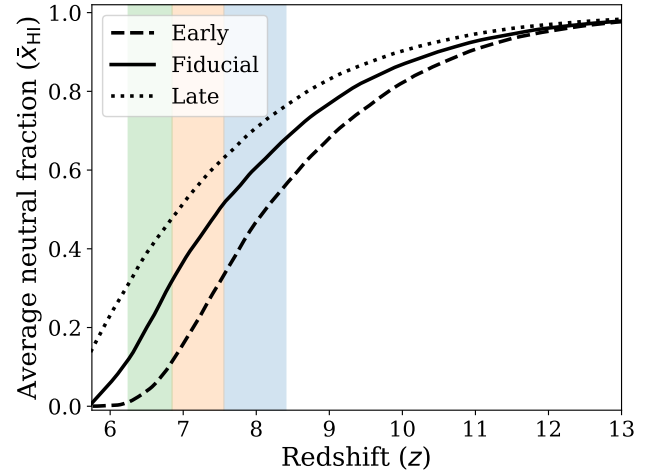
Figure 1. Redshift evolution of sky-averaged neutral fraction $\bar{x}_{\rm HI}$ of three scenarios (early, fiducial and late reionization models). The three coloured bands show the frequency bins considered in this study.

linked to the mass of dark matter halos (M_h), following the parametrisation detailed in Park et al. (2019). The star formation efficiency (SFE), representing the fraction of galactic gas converted into stars, is parameterised by a normalisation $f_{*,10}$ at a halo mass of $10^{10} M_\odot$ and a power-law index α_* describing its scaling with halo mass ($f_* \propto M_h^{\alpha_*}$). Similarly, the escape fraction of ionising UV photons, $f_{\rm esc}$, which dictates how many ionising photons reach the IGM, is modelled with a normalisation $f_{\rm esc,10}$ and a power-law index $\alpha_{\rm esc}$. The ability of low-mass halos to form stars is regulated by $M_{\rm turn}$ below which star formation is exponentially suppressed due to feedback or cooling limitations. The timescale over which galaxies form their stars is parameterised by t_*, expressed as a fraction of the Hubble time. IGM heating is primarily driven by X-rays, modelled by the soft-band (0.5-2 keV) X-ray luminosity per unit star formation rate, $L_{X<2\rm keV}/$SFR, which sets the overall X-ray efficiency. The shape of the X-ray spectrum emergent from galaxies is controlled by E_0, the minimum energy (in keV) required for X-ray photons to escape absorption within the host galaxy, and α_X, the power-law spectral index ($L_X \propto E^{-\alpha_X}$). In principle, varying all nine parameters ($f_{*,10}, \alpha_*, f_{\rm esc,10}, \alpha_{\rm esc}, M_{\rm turn}, t_*, L_{X<2\rm keV}/SFR, E_0, \alpha_X$) allows for a wide exploration of signal characteristics. We describe our prior sampling strategy in the following sub-section.

2.1.2 Reference cases

To test the performance of our inference framework, we define three reference scenarios corresponding to an 'Early', 'Fiducial', and 'Late' reionization history. The astrophysical parameters defining these three models are taken from our test set and are detailed in Table 1. The resulting evolution of the sky-averaged neutral fraction $\bar{x}_{\rm HI}$ for each case is shown in Fig. 1. The 'Fiducial' model's reionization history begins at $z \approx 13$ and ends by $z \approx 5.5$, which is consistent with current constraints on reionization (see e.g., Giri et al. 2024; Qin et al. 2025, for discussion about these constraints). The other two models represent more extreme scenarios. While all models begin at a similar epoch, reionization in the 'Early' model proceeds very rapidly and is nearly complete by $z \approx 6.5$ with a neutral fraction of $\bar{x}_{\rm HI} \approx 0.02$. Conversely, the 'Late' model shows a more gradual trend, remaining significantly neutral with an average neutral frac-

Table 1. The astrophysical parameter values of the three reference scenarios from our testing set, along with the corresponding sky-averaged neutral fractions within ~15 MHz frequency band at the three redshifts considered in this study.

| Model	$\log_{10}f_{\star,10}$	$\alpha_\star$	$\log_{10}f_{esc,10}$	α_{esc}	$\log_{10}M_{turn}$	$t_\star$	$\bar{x}_{HI}(z=8.0)$	$\bar{x}_{HI}(z=7.2)$	$\bar{x}_{HI}(z=6.5)$
Early	-1.05	-0.01	-1.01	-0.91	9.60	0.60	0.46	0.21	0.02
Fiducial	-0.38	-0.21	-2.11	0.31	8.63	0.72	0.59	0.43	0.20
Late	-0.43	0.21	-2.25	0.05	8.22	0.19	0.70	0.56	0.39

tion $\bar{x}_{HI} \gtrsim 0.1$ at $z \approx 6$. The specific values of the neutral fraction for each model at our three redshifts of interest ($z = 8.0, 7.2,$ and 6.5) are listed in Table 1.

To properly model the signal for such evolutionary histories, we generate *lightcone* data volumes from our 21CMFAST models. A *lightcone* data represents the 21-cm signal as it evolves along the line-of-sight, since different observed frequencies, ν_{obs}, correspond to different redshifts (and thus different cosmic times). For our analysis, we focus on measurements within three distinct redshift bins centred at $z = 8.0, 7.2,$ and 6.5 ($\nu_{obs} \simeq 158, 173$ and 189 MHz), each with a bandwidth of ~15 MHz. These bins are highlighted by the coloured bands in Fig. 1. While the signal evolution within this bandwidth introduces the lightcone effect (e.g., Datta et al. 2012; Giri et al. 2018a), our explicit construction of lightcones for all models in the training set ensures that this effect is consistently included and does not introduce bias into our inference framework.

2.2 Sampling prior

For this study, we vary the six key astrophysical parameters θ controlling the galaxy UV properties and star formation activity: $f_{\star,10}, \alpha_\star, f_{esc,10}, \alpha_{esc}, M_{turn}, t_\star$. The remaining three parameters describing the X-ray properties of galaxies ($L_{X<2keV}/SFR$, E_0, α_X) were kept fixed to the fiducial values recommended in Park et al. (2019). We initially sampled the six varied parameters from broad, uniform prior distributions: $\log_{10}(f_{\star,10}) \in [-3,1]$, $\alpha_\star \in [-0.5,1]$, $\log_{10}(f_{esc,10}) \in [-3,0]$, $\alpha_{esc} \in [-1,0.5]$, $\log_{10}(M_{turn}/M_\odot) \in [8,10]$, and $t_\star \in (0,1]$. The initial sampling distributions $p_0(\theta)$ for these astrophysical parameters are shown as dashed grey lines in the top two rows of Fig. 2. However, this uniform sampling in the astrophysical parameter space leads to a highly non-uniform, bimodal distribution of the resulting reionization histories, $\bar{x}_{HI}(z)$, as illustrated by the dashed grey lines in the bottom row of Fig. 2 for $z = 8.0, 7.2,$ and 6.5. This bimodality, with peaks corresponding to very early/rapid or very late/slow reionization scenarios, presents challenges for training neural networks in SBI frameworks, potentially leading to inefficient learning or biased posteriors.

To address this, we implemented a custom sampling strategy designed to yield a more balanced distribution of reionization histories. We leveraged the publicly available 21CMEMU emulator (Breitman et al. 2024), which provides a rapid mapping from the 21CMFAST astrophysical parameters to the evolution of the sky-averaged neutral fraction, $\bar{x}_{HI}$. We drew parameter sets θ from the initial uniform priors $p_0(\theta)$ and used 21CMEMU to predict the corresponding $\bar{x}_{HI}$. We then applied rejection sampling, discarding models where reionization either completed too early (defined as $\bar{x}_{HI}(z = 8.4) \le 0.01$) or had not significantly progressed by the end of our redshift range of interest (defined as $\bar{x}_{HI}(z = 6.25) \ge 0.95$). Every few hundred iterations, we stop the parameter sampling and modify the upper and lower limits to target specific reionization scenarios, aiming to obtain a more balanced distribution of $\bar{x}_{HI}$ across the three redshift bins. Nevertheless, we can observe that the highest (lowest) redshift has a preference for late (early) scenarios. We continued this process until we collected ~15,900 valid parameter sets. These sets were then
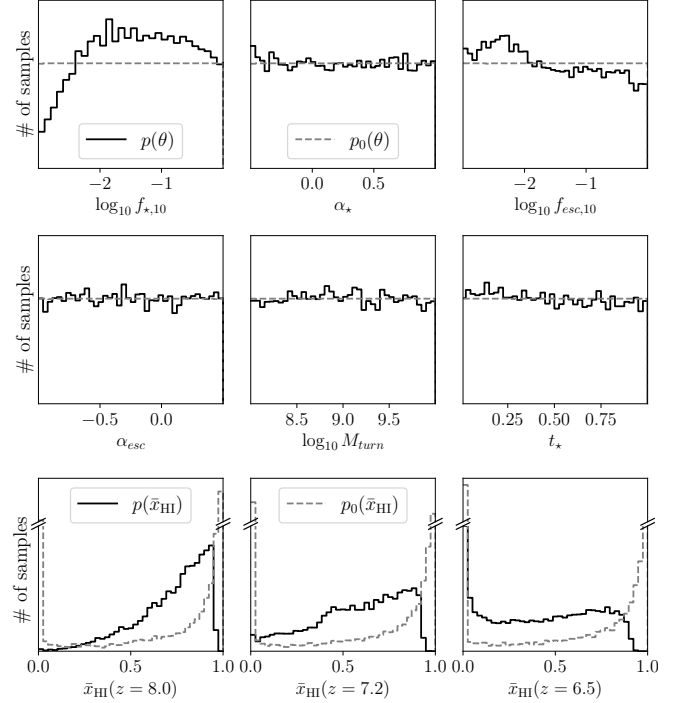
Figure 2. Marginal distributions on the simulation parameters θ (top and middle row) and on the resulting $\bar{x}_{HI}$ for three different redshifts (bottom row). y-axes show the number of sample in the dataset. A uniform sampling $p_0(\theta)$ of the astrophysical parameters (dashed histograms) give a highly bimodal distribution $p_0(\bar{x}_{HI})$ of $\bar{x}_{HI}$ in the three frequency bins. With our custom sampling strategy $p(\theta)$ (see Sec. 2.2), we get a relatively more balanced distribution $p(\bar{x}_{HI})$ of these bins (solid histograms).

used to run full 21CMFAST simulations to generate our final training dataset. We vary the cosmological initial conditions for each sample to consistently include the cosmic variance in this dataset.

The effective sampling priors $p(\theta)$ resulting from this procedure are shown as solid dark histograms in the top two rows of Fig. 2. While no longer strictly uniform, they still cover the initial parameter ranges broadly. Crucially, the distribution of neutral fractions $p(\bar{x}_{HI})$ resulting from this custom sample, shown as solid dark histograms in the bottom row of Fig. 2, is significantly more balanced and less bimodal compared to the uniform astrophysical prior case. Although not perfectly uniform in $\bar{x}_{HI}$, this improved distribution provides a much better basis for training our inference framework.

2.3 Forward-modelling instrumental effects

Implicit inference methods critically rely on the fidelity of the training set in mimicking the observation. Realistic forward-modelling of observational contaminants and instrumental effects is crucial for

the neural networks to learn the correct mapping between the data summaries to underlying parameters, especially when the likelihood function is intractable. In this work, we incorporate the impact of instrumental noise and finite resolution characteristic of SKA-Low observations.

We consider two mock observational setups corresponding to the AA* antenna configuration of SKA-Low, which includes 307 antenna stations. The initial science viable data will be produced with this configuration. The first setup assumes a total integration time of 100 hours, while the second assumes a deeper observation of 1000 hours. For both setups, we assume 6 hours of observation per day, centred on the field transiting zenith at a declination of -30 degrees. Instrumental noise is a primary contaminant and we calculate the expected noise standard deviation per visibility ($\sigma_{\rm noise}$) as following (e.g., Ghara et al. 2017):

$$\sigma_{\rm noise} = \frac{\rm SEFD}{\sqrt{2\Delta\nu\Delta t}} \qquad (2)$$

where $\Delta\nu$ is the channel width and Δt is the integration time per visibility. The SEFD is the system equivalent flux density that depends on the system temperature $T_{\rm sys}$ and the effective area $A_{\rm eff}$ of each station (SEFD $= 2k_B T_{\rm sys}/A_{\rm eff}$, where k_B is the Boltzmann constant). We adopt the frequency-dependent $T_{\rm sys}$ and $A_{\rm eff}$ models relevant for SKA-Low, as defined in Giri et al. (2018b). Following the procedure in Giri et al. (2018b), we generate Gaussian random noise realisations on a grid matching the 21cmFAST simulation resolution, consistent with this $\sigma_{\rm noise}$.

The interferometer does not sample all spatial frequencies uniformly; its sampling pattern in the uv-plane is determined by the station layout and Earth's rotation. Using the publicly available package Tools21cm (Giri et al. 2020), we simulate the daily uv-coverage tracks based on the SKA-Low AA* station positions. The simulated noise grids are then convolved with the weights derived from these uv-tracks. We assume a natural weighting scheme, which optimally weights the visibilities based on sampling density to minimise the noise variance in the final map, thereby maximising sensitivity.

Furthermore, the finite extent of the interferometer limits the angular resolution. We mimic this by filtering the simulated signal and noise fields in Fourier space, retaining only the modes corresponding to baselines shorter than the maximum baseline of the SKA-Low core (~1 km diameter). This effectively reduces the angular resolution of the data. To ensure isotropic resolution, we also match the resolution along the frequency axis by smoothing the data cubes to have approximately the same physical scale per voxel as in the angular directions after filtering.

A crucial aspect of interferometry is the absence of a zero-spacing baseline, meaning the sky-averaged signal within each frequency channel cannot be measured. Therefore, before computing any summary statistics, we subtract the spatial mean from each frequency slice of our simulated data cubes (signal plus noise) to mimic this observational characteristic. This procedure of generating noise on a grid and then convolving with the uv-weights is computationally efficient compared to direct visibility simulation, allowing us to rapidly produce multiple noise realisations. To capture this stochasticity in our training set, we generate approximately 25 independent noise realisations for each 21cmFAST model lightcone. These realisations aid the implicit inference framework learn to marginalise over the noise properties during inference.
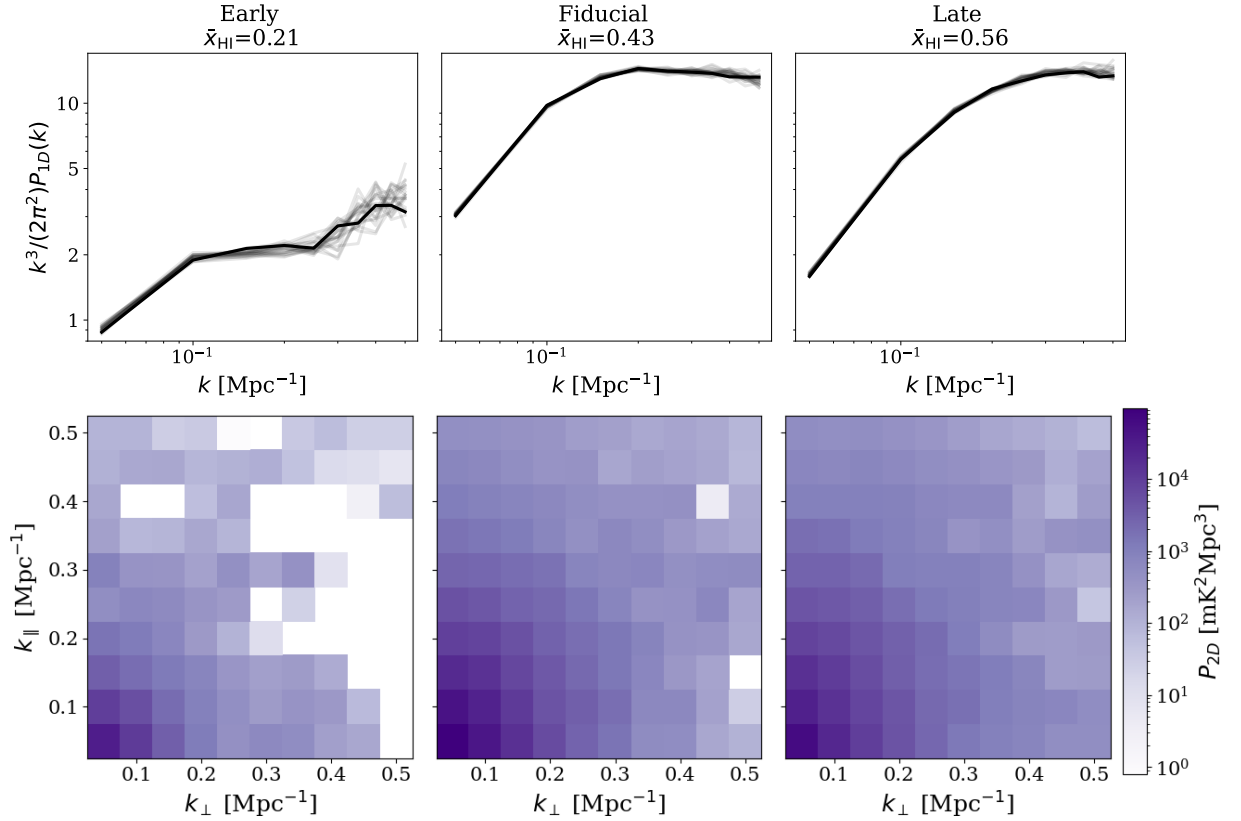
2.4 Summary statistics

A direct (or field-level) comparison of the simulated 21-cm data with observations presents a high-dimensional inference problem that is computationally prohibitive. Furthermore, initial data from interferometers such as SKA-Low are expected to be noise-dominated, making such a direct analysis sub-optimal. We therefore compress this information by estimating several summary statistics from the cubes at different redshift bins, incorporating the instrumental effects described above. These statistics serve as the input data for our inference framework.

2.4.1 2-point statistics

The primary statistic in 21-cm cosmology to extract information is the power spectrum (PS), which quantifies the variance of the signal fluctuations as a function of spatial scale. As a two-point statistic, it fully describes a Gaussian random field and is proportional to $\bar{x}_{\rm HI}^2$ (e.g., Furlanetto et al. 2004; Georgiev et al. 2022). We consider two versions. The first is the spherically averaged power spectrum (PS1D), $P_{\rm 1D}(k)$, which averages the signal power within spherical shells in Fourier space, providing information on the overall amplitude of fluctuations at different scales k. The second is the cylindrically averaged power spectrum (PS2D), $P_{\rm 2D}(k_\perp, k_\parallel)$, which averages power in cylindrical annuli, separating scales perpendicular ($k_\perp$) and parallel ($k_\parallel$) to the line-of-sight. This retains information about line-of-sight effects such as redshift-space distortions (e.g., Jensen et al. 2013). Both statistics are computed using modules available within Tools21cm (Giri et al. 2020).

Fig. 3 illustrates these Gaussian statistics at $z = 7.2$ for our three reference models. The top row shows the PS1D for the three models. The different light solid lines in each panel show the PS1D of the signal with several different noise realisations for a 1000-hour SKA-Low observation, highlighting the variations introduced by instrumental effects. Instrumental noise adds a positive bias to the power spectrum, which can make the variations between realisations difficult to see. To better visualise these variations, we have removed this noise bias from the curves. One realisation is highlighted with a darker coloured curve. We can see that the shape and amplitude of the PS1D are distinct for each reference model, reflecting their different average neutral fractions at this redshift.

The bottom row of Fig. 3 shows the PS2D for the same single noise realisation as the dark curve in the top panel, but with the noise bias subtracted (similar to the PS1D plots). This bias subtraction reveals the underlying signal structure but also highlights the impact of noise. For instance, in the 'Early' model, which has the lowest signal amplitude, many pixels are white. This is because these pixels were noise-dominated, and subtracting the large noise bias resulted in negative values, which are not displayed on the colour map. This illustrates the different noise properties of the two statistics. The PS1D benefits from more compression (averaging within spherical shells) and thus has a lower relative noise error. The PS2D, by preserving the $k_\perp$ and $k_\parallel$ separation, is more prone to noise but contains more astrophysical information from the signal's anisotropy. We will test and compare the physical information that can be extracted from both metrics in the results (Sec. 4). It is important to note that our inference framework does not subtract the noise bias; it is trained on and compares the statistics (signal + noise) directly. This avoids any issues related to bias subtraction, such as the negative values seen in the figure.

Figure 3. Gaussian summary statistics at $z = 7.2$ for the three reference models (left to right: Early, Fiducial, Late), assuming 1000h SKA-Low noise. *Top row:* The spherically averaged power spectrum, $P_{1D}(k)$. Lighter lines show multiple noise realisations, and the darker line highlights a random one. The instrumental noise bias has been removed from all curves for clarity. *Bottom row:* The cylindrically averaged power spectrum, $P_{2D}(k_\perp, k_\parallel)$, for the single realisation highlighted in the top row, again with the noise bias removed. White pixels (e.g., in the 'Early' model) indicate noise-dominated modes where bias subtraction resulted in negative values.

2.4.2 Statistics sensitive to non-Gaussian information

The 21-cm signal is expected to be highly non-Gaussian, reflecting the complex morphology of ionised regions (e.g., Iliev et al. 2006; Giri et al. 2018a, 2019). To capture this information, we compute the bispectrum, the three-point correlation function in Fourier space, $B(k_1, k_2, k_3)$, which is explicitly sensitive to non-Gaussianity. We use the publicly available PYLIANS3 package (Villaescusa-Navarro 2018), which implements a Fast Fourier Transform (FFT) based algorithm to estimate it (e.g., Sefusatti et al. 2016; Watkinson et al. 2017). This algorithm requires the input data to be on a cubic grid. Since our lightcone data volumes are generally not cubic, we crop the data along the angular dimensions to create cubic sub-volumes before calculating the bispectrum.

The bispectrum depends on the configuration of the three wavevectors $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$ (where $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$). We focus on two specific configurations known to be sensitive probes of non-Gaussianity (see Lewis et al. 2011 for a detailed study of shapes probed by different configurations):

- *Equilateral*: $k_1 = k_2 = k_3 = k$. This configuration is sensitive to the typical shapes of the ionised regions and their clustering (e.g., Watkinson et al. 2019; Majumdar et al. 2018).
- *Squeezed-limit*: $k_1 \approx k_2 \gg k_3$. This configuration primarily probes the coupling between large-scale modulation and small-scale signal fluctuations (e.g., Giri et al. 2019).

We assume a k-bin width of 0.1 Mpc^{-1} for these calculations. Our

analysis specifically uses the reduced bispectrum, which is defined as (e.g., Scoccimarro et al. 1998):

$$Q(k_1, k_2, k_3) = \frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2) + P(k_2)P(k_3) + P(k_3)P(k_1)} . \quad (3)$$

This normalisation isolates the purely non-Gaussian part of the signal. The top two rows of Fig. 4 show this reduced bispectrum for the 'Equilateral' and 'Squeezed-limit' configurations, respectively. As in Fig. 3, the columns correspond to the three reference models, and the plots show multiple noise realisations (faint lines) with one highlighted (dark curve). These plots show that the shape and amplitude of the reduced bispectrum also vary significantly between the models. Later in our forecast study, we infer constraints from both configurations jointly and refer to this combination as 'Bispec'.

In contrast to the Fourier-space statistics mentioned above, we also compute Betti numbers. These are topological invariants calculated directly from the 21-cm signal images, which are an expected data product from SKA-Low (e.g., Mellema et al. 2015; Giri & Mellema 2021). Specifically, β_0 counts the number of connected components (e.g., isolated ionised regions), β_1 counts the number of tunnels or loops, and β_2 counts the number of enclosed voids. We calculate these as a function of a threshold value v, defined relative to the standard deviation of the data ($v = \delta T_b / \sigma_{\delta T_b}$), using 21 threshold
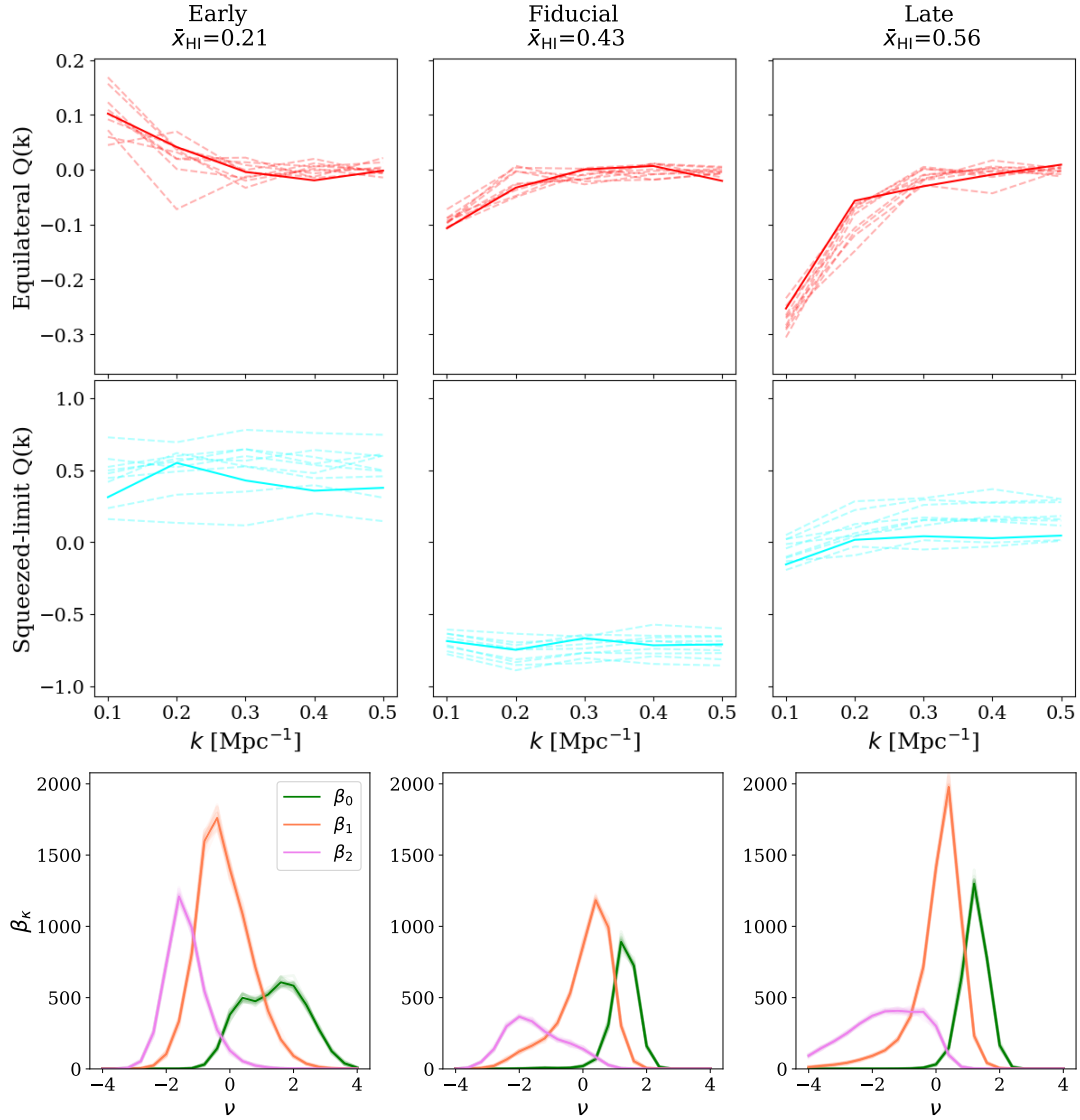
Figure 4. Non-Gaussian summary statistics at $z = 7.2$ for the three reference models (Early, Fiducial, Late), assuming 1000h SKA-Low noise. *Top row:* The reduced equilateral bispectrum, $Q(k)$. *Middle row:* The reduced squeezed-limit bispectrum, $Q(k)$. *Bottom row:* The Betti numbers $(\beta_0, \beta_1, \beta_2)$ as a function of the threshold ν. In all panels, light lines show multiple noise realisations, and the dark lines highlight a single, common realisation.

bins between $\nu = -4$ and 4. This provides information about the topology at different intensity levels[2].

The instrumental noise in the images is approximately Gaussian in nature (e.g., Giri et al. 2018b). For a reference Gaussian random field, the Betti number curves are known to be symmetric, with β_0 and β_2 as mirror images of each other (e.g., Pranav et al. 2019; Giri & Mellema 2021). The bottom row of Fig. 4, showing the Betti curves for our reference models, clearly deviates from this symmetry. The distinct, asymmetric shapes of these curves, particularly the peaks and their locations, demonstrate that they capture the unique non-Gaussian topological signatures of each reionization scenario, which are invisible to the power spectrum.

3 IMPLICIT LIKELIHOOD AND PRIOR INFERENCE

We perform the joint inference of $\bar{\mathbf{x}}_{HI} = [\bar{x}_{HI}(z = 8.0), \bar{x}_{HI}(z = 7.2), \bar{x}_{HI}(z = 6.5)]$ given $\mathbf{t}$, a summary statistic measured in the three redshift (or frequency) bins of the study (see §2.4). Our data model provides samples $(\bar{\mathbf{x}}_{HI}, \mathbf{t})$ drawn from the simulator, which can be seen as an implicit likelihood distribution $p(\bar{\mathbf{x}}_{HI}, \mathbf{t}|\theta)$. We aim to estimate the posterior distribution $p(\bar{\mathbf{x}}_{HI}|\mathbf{t})$, which following Bayes' theorem can be evaluated as:

$$p(\bar{\mathbf{x}}_{HI}|\mathbf{t}) \propto p(\mathbf{t}|\bar{\mathbf{x}}_{HI}) \, p(\bar{\mathbf{x}}_{HI}). \tag{4}$$

However, unlike in a classical forward model where the inferred quantity is an input to the model, here the mean neutral fraction $\bar{\mathbf{x}}_{HI}$ is one of its output. Hence, the simulator carries an implicit prior

---

[2] This thresholding strategy or *filtration* differs from that used in Giri & Mellema (2021) where ionised regions were first explicitly identified using a pattern recognition framework across all redshifts, and Betti numbers were then calculated as a function of redshift. Our approach focuses on topology as a function of signal intensity within the three specific redshift bins considered in this study.

$p(\bar{\mathbf{x}}_{HI})$, which can be written as a function of the simulator's implicit likelihood and the simulator's prior $p(\theta)$ (see §2.2):

$$p(\bar{\mathbf{x}}_{HI}) = \iint p(\bar{\mathbf{x}}_{HI}, \mathbf{t}|\theta) p(\theta) d\theta d\mathbf{t}. \quad (5)$$

Since we do not have explicit expressions for both $p(\bar{\mathbf{x}}_{HI}, \mathbf{t}|\theta)$ and $p(\theta)$, it is clear that the prior $p(\bar{\mathbf{x}}_{HI})$ is implicit. Likewise, the likelihood $p(\mathbf{t}|\bar{\mathbf{x}}_{HI})$ in eq. 4 can be written as a function of the simulator's implicit likelihood and prior:

$$p(\mathbf{t}|\bar{\mathbf{x}}_{HI}) = \int \frac{p(\bar{\mathbf{x}}_{HI}, \mathbf{t}|\theta)}{p(\bar{\mathbf{x}}_{HI})} p(\theta) d\theta. \quad (6)$$

As all terms are intractable in this expression, it demonstrates that also the likelihood term of in Eq. (4) is implicit. While implicit inference is classically required when the likelihood only is intractable, it is in our case also justified by the intractable prior on $\bar{\mathbf{x}}_{HI}$. This problem requires us to use Neural Posterior Estimation (NPE, Papamakarios & Murray 2018), as it is the only class of implicit inference algorithm that learns the prior together with the likelihood.

Here, we follow a state-of-the-art inference strategy to model $p(\bar{\mathbf{x}}_{HI}|\mathbf{t})$. We use the NPE method, e.g. training a conditional distribution $q_{\varphi}(\bar{\mathbf{x}}_{HI}|\cdot)$ with weights φ to learn both the likelihood and prior. Instead of conditioning q_{φ} directly on the summaries $\mathbf{t}$, we perform variational mutual information maximisation (VMIM, Jeffrey et al. 2020) to optimally extract the relevant information from $\mathbf{t}$. To do so, we train a neural network $\mathbf{y}_{\psi}(\mathbf{t})$ with weights ψ. Its output yields the same shape as $\bar{\mathbf{x}}_{HI}$. In practice, $\mathbf{y}_{\psi}(\mathbf{t})$ is obtained by training $\mathbf{y}_{\psi}$ simultaneously with q_{φ}. Hence, we minimise the combined VMIM loss:

$$L_{VMIM}(\varphi, \psi) = -\mathbb{E}_{p(\mathbf{x}_{HI}, \mathbf{t})} \log\left(q_{\varphi}(\mathbf{x}_{HI}|\mathbf{y}_{\psi}(\mathbf{t}))\right). \quad (7)$$

For $\mathbf{y}_{\psi}(\mathbf{t})$, we opt for a simple network architecture that can be indifferently applied to all data types. This allows for a fair comparison of the posterior estimators. We used a fully connected architecture with 3 dense layers of 256 neurons and Rectified Linear Unit activations. In de Salis et al. (2025) we explored complex deep neural architectures to infer $\bar{\mathbf{x}}_{HI}$ from PS2D; we devote a similar study for higher-order statistics to a future work. For q_{φ}, we use the regular NPE implentation in the SBI package (Tejero-Cantero et al. 2020). For an optimal compression, $p(\bar{\mathbf{x}}_{HI}|\mathbf{y})$ should be identical to $p(\bar{\mathbf{x}}_{HI}|\mathbf{t})$. As such, we refer similarly to both distributions as 'the posterior'.

When combining different kind of statistics $(\mathbf{t}_1, \ldots, \mathbf{t}_n)$, we separately train the VMIM compressors and concatenate their respective outputs $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ into $\mathbf{y}_{1 \times \cdots \times n}$. Then, the NPE training is performed on $\mathbf{y}_{1 \times \cdots \times n}$. In the following, we will focus on combinations of gaussian and higher-order statistics, in particular PS2D+Betti, PS2D+Bispec, and PS2D+Betti+Bispec.

Through the simulation of $\sim 15,900$ independent mock lightcones and the various noise realisations applied to each (see §2.2 and §2.3), we end up with a sample of $\sim 330,000$ $(\bar{\mathbf{x}}_{HI}, \mathbf{t})$ pairs, of which 170,000 (30,000) are used for the VMIM training (validation), 60,000 (20,000) for the NPE training (validation), and the remaining for testing. To check the robustness of our approach, we train 30 independent compressors and posteriors for each level of noise and kind of summary $\mathbf{t}$.

4 RESULTS

4.1 Inference framework calibration

Assessing the calibration of the learnt posterior is critical for implicit inference methods. We here use the test of accuracy with random
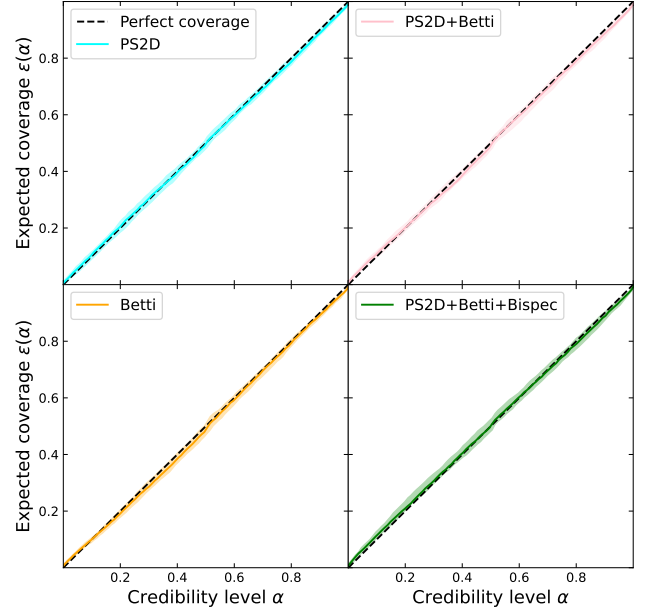
Figure 5. Coverage test on the best-calibrated models for PS2D (cyan), Betti numbers (gold), PS2D+Betti (pink) and PS2D+Betti+Bispec (green). For each model we ran 20 TARP realisations and show medians as solid lines and 95% of the samples as shaded regions.

points (TARP, Lemos et al. 2023), an efficient method to detect miscalibrated multidimensional posteriors. It provides us with the expected coverage probability as a function of the credibility level, $\varepsilon(\alpha)$. We computed the empirical frequency $\hat{p}_k$ of $\varepsilon(\alpha)$ over $K = 100$ credibility level bins by bootstraping the TARP 40 times. Similarly to Saoulis et al. (2025), we then computed the calibration error as the mean squared error between $\hat{p}_k$ and the ideal p_k:

$$C = \frac{1}{K} \sum_k \left(\frac{\hat{p}_k - p_k}{p_k}\right)^2. \quad (8)$$

For an ideally calibrated posterior, $\varepsilon(\alpha) = \alpha$ and hence $\hat{p}_k = p_k = 1/K$. We computed the calibration error for all learnt posterior, and selected for each summary (including combinations of summaries) the posterior model that achieves the lower C. These best-calibrated posteriors typically yield $C \sim 0.002$. In Fig. 5 we show the TARP outputs $\varepsilon(\alpha)$ for PS2D (cyan), Betti numbers (gold), PS2D+Betti (pink) and PS2D+Betti+Bispec (green). Solid lines indicate the median from 20 TARP realisations and shaded regions encompass 95% of the samples. From the figure we do not note any significant deviation from the 1:1 line, meaning that these posteriors are unbiased and well calibrated. In the remaining of the paper, when referring to a posterior estimator of a specific kind, we in fact refer to the best-calibrated model for the corresponding summary.

To estimate the amount of information provided by a posterior, we employ the figure of merit (FoM). We empirically estimate the posterior covariance $\text{Cov}[\theta|\mathbf{t}_0]$ for a test $\mathbf{t}_0$ and obtain the FoM:

$$\text{FoM} = |\text{Cov}[\theta|\mathbf{t}_0]|^{-1/N}, \quad (9)$$

with N the number of dimension of the posterior, 3 in our case, and $|\cdot|$ the matrix determinant. Since the determinant scales as the N-th power of the covariance matrix elements, in our definition the FoM roughly scales as the inverse square of the standard deviation of the posterior samples, or σ^{-2}. For instance, an increase of the FoM of
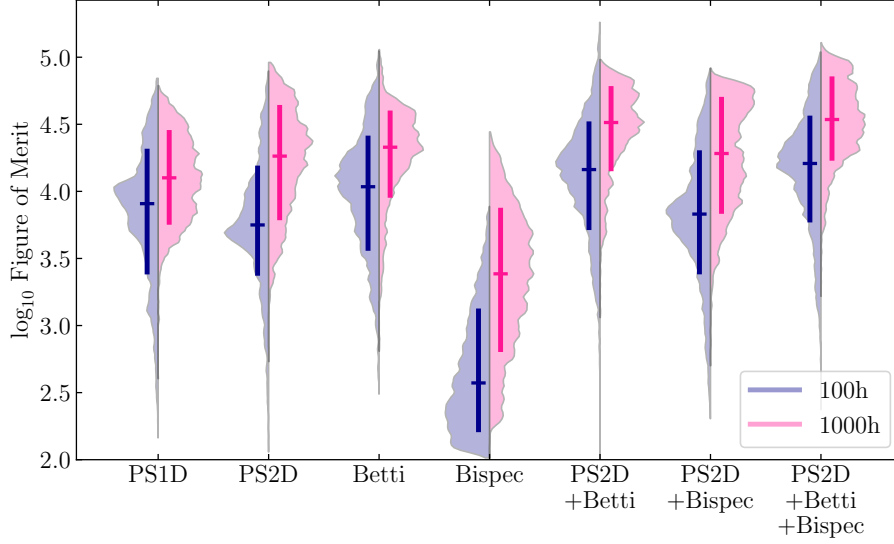
Figure 6. Figure of merit distributions for all tested summary statistics, from left to right: spherically averaged power spectrum (PS1D), cylindrically averaged power spectrum (PS2D), Betti numbers (Betti), reduced bispectrum (Bispec), and their combinations PS2D+Betti, PS2D+Bispec, PS2D+Betti+Bispec. Horizontal bars denote the median of the distribution while vertical bars range from the 14th to the 84th percentiles of the samples. We show the results for both 100h (blue) and 1000h (pink) SKA-Low AA* noise level.

Table 2. Figure of merit comparison. The central part of the table shows $\log_{10}$ FoM for sets of summary statistics (single and combined). We quote for 100h (top section) and 1000h (central section) SKA-Low AA* noise level, the 16th, 50th 84th percentiles of the FoM distributions sampled over 4000 test mocks. The right-hand side of the table shows the relative improvement of these quantities going from PS2D alone to PS2D+Betti and PS2D+Betti+Bispec. The bottom section of the table shows the relative change of the 50th percentile going from 100h to 1000h noise level.

noise level	percen-tiles	$\log_{10}$ FoM							Relative improvement (dex)	
		PS1D	PS2D	Betti	Bispec	PS2D +Betti	PS2D +Bispec	PS2D +Betti +Bispec	(PS2D+Betti) /PS2D	(PS2D+Betti +Bispec)/PS2D
100h	16th	3.40	3.39	3.58	2.22	3.73	3.40	3.79	0.34	0.40
	50th	3.91	3.75	4.03	2.57	4.16	3.83	4.21	0.41	0.46
	84th	4.30	4.17	4.40	3.11	4.50	4.29	4.55	0.33	0.37
1000h	16th	3.77	3.80	3.97	2.82	4.17	3.85	4.25	0.37	0.44
	50th	4.10	4.26	4.33	3.39	4.51	4.28	4.54	0.25	0.27
	84th	4.44	4.63	4.58	3.86	4.77	4.69	4.84	0.14	0.21
1000h / 100h (dex)	50th	0.19	0.51	0.29	0.81	0.35	0.45	0.33		

0.3 dex represents an decrease of $\sim 40\%$ on the uncertainty on $\bar{x}_{\mathrm{HI}}$. For each model, we selected 4000 random $\mathbf{t_0}$ samples from the test set and measured the FoM of their associated posterior.

### 4.2 Comparison of constraining power of the statistics

We start by comparing the FoM distributions for all best-calibrated models. In Fig. 6 we show from left to right the FoM distributions from PS1D, PS2D, Betti, Bispec, PS2D+Betti, PS2D+Bispec and PS2D+Betti+Bispec. The distributions are represented in both cases of noise level, 100h (blue) and 1000h (pink). To help comparing the distributions, we show the median (horizontal bar) and the [16th-84th] percentiles interval (vertical bar). We also provide the values of the median, 16th and 84th percentiles for all distributions in table 2. As expected, all distributions systematically shift upwards when the observational time is ten times longer. The PS1D shows the mildest improvement (0.19 dex) from 100h to 1000h, while PS2D

yields a 0.51 dex increase in FoM. The Bispectrum has the biggest increase (0.81 dex), but remains significantly less informative than other statistics.

For gaussian statistics alone, we observe that the PS1D and the PS2D have roughly equivalent FoMs. At 100h, the PS2D is slightly below the PS1D, but we observe an opposite trend at 1000h. This shows that the PS2D is more informative than the PS1D, provided that there is sufficient signal-to-noise.

Then, looking at the higher-order statistics alone, the Betti numbers appears to be more informative than the PS2D. However, this is not true for the Bispec, which FoMs are about 1.5 dex below the other summaries alone. Several reasons may explain this. First, the bispectrum computation required an angular scale cut, excluding the largest scales that are the most informative on $\bar{x}_{\mathrm{HI}}$, as explained in §2.4.2. Then, we computed the reduced bispectrum, in which 2pt-information are cancelled out.

Lastly, we look at the FoMs for combined statistics. The
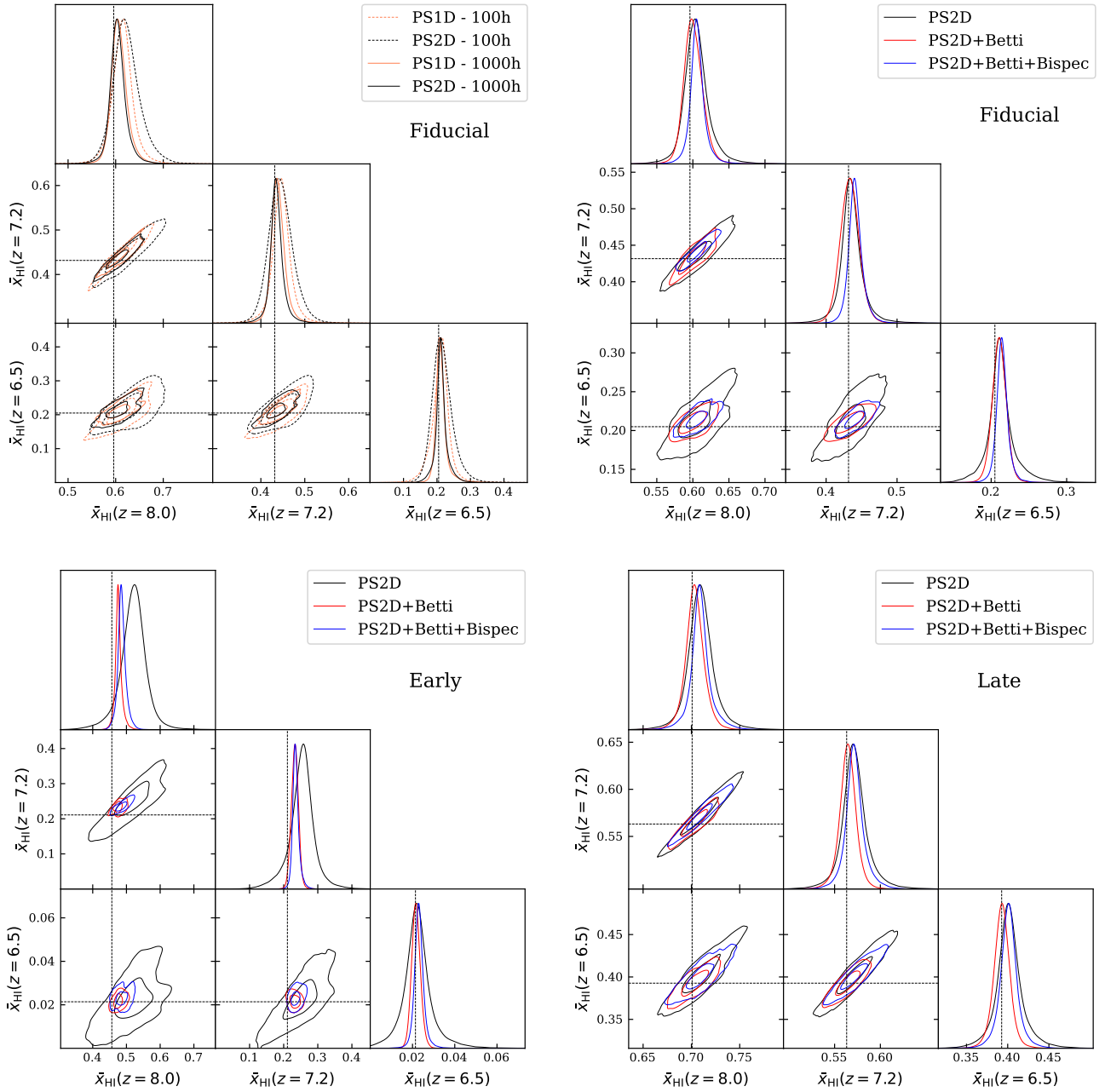
Figure 7. Posterior constraints on the 3D reionization history ($\bar{x}_{HI}(z=8.0)$, $\bar{x}_{HI}(z=7.2)$, $\bar{x}_{HI}(z=6.5)$) for our three reference mocks. *Top-left*: Comparison of posteriors from PS1D and PS2D for the Fiducial mock, shown for both 100h (dashed lines) and 1000h (solid lines) SKA-Low noise. *Other panels*: Comparison of posteriors from PS2D (black), PS2D+Betti (red), and PS2D+Betti+Bispec (blue), all assuming 1000h noise. The mock is the Fiducial model (top-right), the Early model (bottom-left), and the Late model (bottom-right). The true values for each mock are indicated by the dashed grey lines.

PS2D+Betti median is improved by 0.25 (0.18) dex with respect to PS2D (Betti) alone. These summaries hence contain complementary information on $\bar{x}_{HI}$. The PS2D+Bispec also slightly improves on PS2D alone (0.02 dex for the median), even if the Bispec alone is much less constraining $\bar{x}_{HI}$. This is in fact not unexpected as the reduced bispectrum captures only non-gaussian information, while the PS2D is restriced to gaussian information. There combination can only improve the quantity of information available on $\bar{x}_{HI}$. Finally, the combination PS2D+Betti+Bispec appears to be the most informative one. The median FoM is improved by 0.27 dex with re-

spect to the PS2D alone. Interestingly, the combination of statistics benefits more to the lower end of the distribution (the 16th percentile is increased by 0.37 and 0.44 dex respectively for PS2D+Betti and PS2D+Betti+Bispec).

4.3 Posterior constraints for reference scenarios

Having compared the overall FoM distributions, we now examine the detailed 2D and 1D posterior constraints for our reference mocks, presented in Fig. 7. The top-left panel focuses on the Fiducial mock,
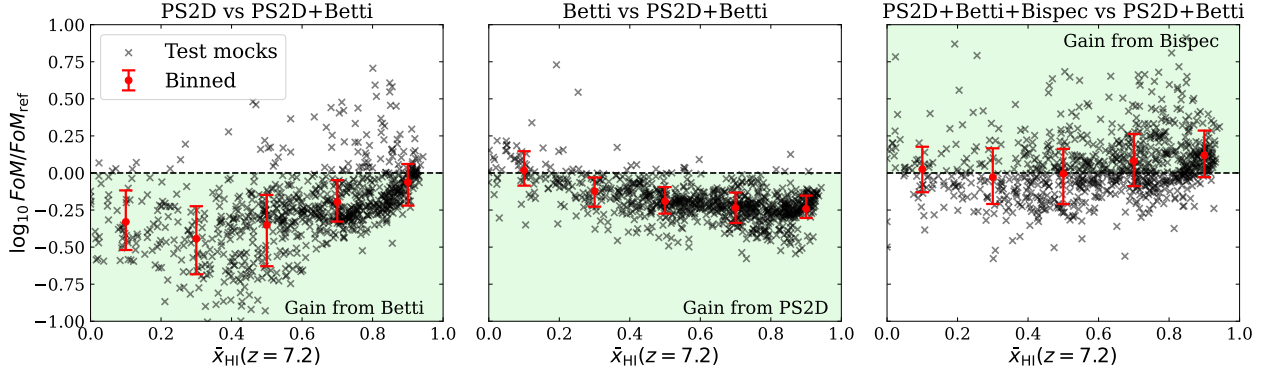
Figure 8. Relative improvement of the figure of merit as a function of the x_{HI} value in the central frequency range, for the 1000h noise level case. The reference posterior is PS2D+Betti across all panels: this makes all point of comparisons only one step away from the reference (i.e., one statistic is either removed or added). From left to right, we show the relative FoM of the posteriors from PS2D, Betti, and PS2D+Betti+Bispec. The ratio FoM/FoM$_{ref}$ for each test mock is represented as a function of $x_{HI}(z = 7.2)$ (black crosses). We divide the data points into five $\bar{x}_{HI}$ bins and show the median (horizontal red bar) and the (16th, 84th) percentiles interval (vertical red bar). In green we show the region where the additional summary leads to a gain in constraining power.

comparing the posteriors obtained from the PS1D and PS2D statistics at both 100h (dashed lines) and 1000h (solid lines) noise levels. As expected, the deeper 1000h observation yields significantly tighter constraints in all cases. We find that at 1000h, the constraining power of PS1D and PS2D is comparable. However, for the shallower 100h observation, the PS1D provides slightly better constraints than the PS2D, which is consistent with the FoM distributions for lower signal-to-noise ratio data we saw in Fig. 6.

The remaining three panels investigate the gain from adding non-Gaussian statistics, assuming the deep 1000h SKA-Low observation. Since PS1D and PS2D are comparable at this noise level, we use PS2D as the fiducial Gaussian baseline (black). The top-right panel shows these results for the Fiducial mock, the bottom-left panel for the Early mock, and the bottom-right panel for the Late mock. Each of these plots compares the constraints from PS2D (black) against those from PS2D+Betti (red) and the full combination of PS2D+Betti+Bispec (blue). In general, adding non-Gaussian information improves the constraints. For both the Fiducial and Early models, the posteriors shrink significantly when Betti numbers are included; in the case of the Early model they shrink further with the addition of the bispectrum. This demonstrates the complementary nature of these statistics.

However, we observe an interesting exception in the Late model case (bottom-right). While adding Betti numbers (red) still provides a substantial improvement over PS2D alone, the full combination including the bispectrum (blue) results in a weaker constraint, with a posterior volume larger than that from PS2D+Betti alone. This suggests that the constraining power of a given statistic may not be uniform, but rather depends on the underlying reionization history (i.e., the neutral fraction). We explore this dependence in detail in the next section.

4.4 Dependence of constraining power on reionization state

The results from the Late model in Fig. 7, where adding the bispectrum to the PS2D+Betti combination unexpectedly degraded the posterior, suggest that the constraining power of different statistics is not uniform. This motivates a more detailed investigation into how the reionization state, i.e., the average neutral fraction $\bar{x}_{HI}$, impacts the relative performance of our summary statistics. To explore this, we use our full test set of mocks. Fig. 8 plots the relative FoM for the

1000h noise case as a function of the neutral fraction at $z = 7.2$. We use the PS2D+Betti combination as a reference (FoM$_{ref}$) and plot the log-ratio of the FoM for other statistics relative to it. The faint crosses show individual mocks, while the red bars show the binned median and [16th-84th] percentile range.

The panels in Fig. 8 reveal the complementary nature of the statistics and their state-dependence. The left panel shows the performance of PS2D relative to PS2D+Betti. The ratio is on average well below zero ($\log_{10}(FoM/FoM_{ref}) < 0$), confirming that Betti numbers provide significant, complementary information at all stages of reionization, as $FoM_{PS2D} < FoM_{PS2D+Betti}$. This improvement is most pronounced (around 0.4 dex) at low-to-mid neutral fractions ($\bar{x}_{HI} < 0.6$). Similarly, the middle panel shows the performance of Betti numbers alone. As expected, this ratio is also consistently below zero, indicating that the PS2D adds crucial information not captured by Betti numbers, especially at high neutral fractions ($\bar{x}_{HI} > 0.6$) where the red line dips lowest. Moreover, it is worth noting that for $\bar{x}_{HI} < 0.2$, Betti numbers alone yield similar FoM than PS2D+Betti. The right panel is the most revealing, showing the effect of adding the bispectrum (PS2D+Betti+Bispec) relative to our reference. In agreement with Fig. 6 and Table 2, the PS2D+Betti+Bispec only mildly improves on PS2D+Betti. However, this improvement depends on the neutral fraction. At low-to-mid neutral fractions ($\bar{x}_{HI} \lesssim 0.6$), the ratio is centred around zero, indicating that on average the bispectrum adds no relevant information. The result observed for our Late model ($\bar{x}_{HI} \approx 0.70, 0.56, 0.39$) corresponds to one of the dark crosses below 0 in the central $\bar{x}_{HI}$ bin. However, at high neutral fractions ($\bar{x}_{HI} \gtrsim 0.6$), the binned median trends above zero. This indicates that for highly neutral scenarios, adding the bispectrum in general brings some new information. Still, dark crosses lying below 0 in the right panel shows that adding the bispectrum can be detrimental for some particular cases. This unlucky negative contribution from higher-order statistics was also found in Semelin et al. (2025), using the pixel density function to improve constraints the astrophysical parameters.

5 SUMMARY AND CONCLUSIONS

In this work, we have used an implicit inference framework to forecast constraints on the reionization history, parameterised by the average

neutral fraction $\bar{x}_{HI}$ at three distinct redshifts ($z = 8.0, 7.2$ and 6.5). We specifically focused on the added value of higher-order summary statistics—the bispectrum (Bispec) and Betti numbers—when combined with the power spectrum, in the context of upcoming SKA-Low observations. The bispectrum, for which we focused on the equilateral and squeezed-limit configurations, probes the three-point correlations and non-Gaussianity of the field, while Betti numbers capture its topological structure, such as the connectivity of ionised or neutral regions in the 21-cm signal.

Our primary finding is that higher-order statistics contain significant, complementary information to power spectrum. To quantify this robustly—a necessary strategy given that the true reionization history is still loosely constrained—we evaluated the figure of merit (FoM) over our entire test set rather than relying on only a few example mocks. This metric, which is inversely related to the posterior volume, showed that the FoM for constraints from the 2D power spectrum (PS2D) alone was consistently and significantly improved by adding Betti numbers (PS2D+Betti). The bispectrum (PS2D+Bispec) also improved constraints over the PS2D alone. This statistical improvement was also visually confirmed by our analysis of the full posterior constraints for our reference mocks (Fig. 7), where the PS2D+Betti combination yielded much tighter posteriors than the PS2D alone, for both 100h and 1000h observation times. This highlights that the topological and three-point information is largely independent of the two-point information in the power spectrum.

However, we also find that the constraining power of a given statistic is not universal, but depends on the reionization state itself. This was most evident with the bispectrum. Our analysis of the full test set (Fig. 8) showed that while PS2D+Bispec improves on PS2D, its benefit varies. The full PS2D+Betti+Bispec combination, which we tested on our reference mocks, provided the tightest constraints for the Early and Fiducial models, but it degraded the constraints for the Late mock (Fig. 7). Fig. 8 confirmed this trend, showing that the bispectrum adds information in low-to-mid neutral fraction scenarios ($\bar{x}_{HI} \lesssim 0.6$) but can degrade the FoM in highly neutral regimes. This is a critical finding, as it implies that the optimal choice of summary statistics for 21-cm cosmology is state-dependent and must be carefully considered for the specific epoch being probed.

Our analysis, while comprehensive in its comparison of statistics, has limitations. We have not included other significant sources of observational contaminants, such as foreground residuals or data calibration errors. These are known to be important; for example, Greig et al. (2024) showed that foreground mitigation could reduce the constraints from the PS2D by a factor of a few. While we refrained from including them here to reduce the complexity of the data model, we acknowledge their importance. With several state-of-the-art foreground mitigation methods being developed (e.g., Mertens et al. 2024; Acharya et al. 2024), a crucial next step will be to understand the impact of their residuals on non-Gaussian statistics, a task that is not yet well understood and is beyond the scope of this work.

Furthermore, our inference framework was trained and tested using the same simulation code, 21cmFAST. It is possible, and indeed likely, that the true universe is not perfectly represented by this simulation. This "model misspecification" is a known challenge (e.g. Zhou & La Plante 2022). Classical likelihood-based studies (e.g., Greig & Mesinger 2015, 2017; Ghara et al. 2020, 2021; Giri & Schneider 2022; Schneider et al. 2023) often account for this by including a "fudge factor" or "modelling error" term to account for theoretical uncertainties, including differences between different simulation codes; this however delivers conservative posterior contours, which weaken the constraining power of the analysis, and is not guaranteed to produce accurate inferences. Our current SBI framework does not

include such a term, and has not been tested on observations from different simulators. Exploring the impact of model misspecification on our higher-order statistics constraints is a critical problem that we will explore in a future paper.

A key motivation for our implicit inference approach is the nature of the reionization history itself. The 21-cm signal is a direct probe of the IGM, and $\bar{x}_{HI}$ is one of the most important properties to infer. However, the reionization history is a derived quantity of the underlying astrophysical model, not a direct input parameter. This makes it challenging to employ classical MCMC methods, which are designed to sample input parameters. Consequently, previous studies (e.g., Ghara et al. 2020, 2021, 2025; Greig et al. 2021a,b) interpreting the available upper limits on the 21-cm power spectrum have typically inferred the ionising source parameters first, and then mapped those posterior constraints onto a derived reionization history. While some new frameworks have been developed to parameterise IGM properties directly (e.g., Mirocha et al. 2022; Ghara et al. 2024), they are at an early stage and do not yet capture the complex mechanisms, such as feedback, that modulate the reionization scenarios. Our SBI framework bypasses this step by learning the latent space that maximises the mutual information between the reionization history and the observed data, thereby directly inferring the reionization history and optimising the information extraction.

Our work joins a growing body of research using implicit inference to move beyond the power spectrum. For instance, Semelin et al. (2025) also demonstrated the power of combining statistics, showing that SBI could effectively merge the power spectrum and the Pixel Distribution Function (PDF) to achieve tighter constraints than either statistic individually. This parallels our own findings with the PS2D+Betti combination, which similarly combined a Fourier-space and a real image-space summary statistic. Other studies have focused on different non-Gaussian compressors, such as applying SBI to the full 3D image compressed by a 3D Convolutional Neural Network (CNN, in Zhao et al. 2022) or, more recently, by a Wavelet Scattering Transform (WST, Zhao et al. 2024). These works found that non-Gaussian compressors outperformed an SBI analysis applied to the power spectrum alone. These studies, along with our own, converge on a key conclusion: while SBI is a powerful framework, its true potential is unlocked by combining higher-order statistics (like Betti numbers, bispectra, or WST), that capture the complex, topological information of reionization, with the power spectrum.

Finally, our implicit inference framework is one of several promising deep learning approaches. In a recent, related study by our collaboration (de Salis et al. 2025), we directly compared different deep learning methodologies, testing various architectures including CNNs, Multi-Layer Perceptron (MLP) Mixers, and an SBI framework. That work focused on the task of inferring $\bar{x}_{HI}$ from the 2D power spectrum alone. We found that the SBI approach was highly competitive, achieving a performance (e.g., $R^2 \approx 97.4\%$) nearly on par with the best-performing, specialized 2D CNNs. This reinforces our finding that implicit inference is a robust and effective method for this inference task. The fact that our collaboration has already demonstrated its performance on the PS2D alone further supports the present results, which show that its predictive power is enhanced when incorporating higher-order summary statistics.

In conclusion, we have demonstrated that an implicit inference framework capable of combining diverse summary statistics is a powerful and necessary tool for 21-cm cosmology. The combination of the power spectrum with non-Gaussian probes—specifically the Betti numbers and, in most regimes, the bispectrum—provides significantly tighter and more robust constraints on the reionization history than the power spectrum alone. This hybrid, multi-statistic

approach, which our work validates, will be crucial for maximizing the scientific return from the 21-cm tomographic data expected from the SKA-Low.

DATA AVAILABILITY

The simulations for the training set were produced with public versions of 21cmFAST (Murray et al. 2020) and Tools21cm (Giri et al. 2020). The specific training dataset used in this study is available from the authors upon reasonable request.

REFERENCES

Acharya A., et al., 2024, Monthly Notices of the Royal Astronomical Society, 527, 7835
Barry N., Hazelton B., Sullivan I., Morales M., Pober J., 2016, Monthly Notices of the Royal Astronomical Society, 461, 3135
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, Monthly Notices of the Royal Astronomical Society, 447, 3402
Bianco M., Giri S. K., Iliev I. T., Mellema G., 2021, Monthly Notices of the Royal Astronomical Society, 505, 3982
Bianco M., Giri S. K., Prelogović D., Chen T., Mertens F. G., Tolley E., Mesinger A., Kneib J.-P., 2024, Monthly Notices of the Royal Astronomical Society, 528, 5212
Bolan P., et al., 2022, Monthly Notices of the Royal Astronomical Society, 517, 3263
Bosman S. E., et al., 2022, Monthly Notices of the Royal Astronomical Society, 514, 55
Breitman D., Mesinger A., Murray S. G., Prelogović D., Qin Y., Trotta R., 2024, Monthly Notices of the Royal Astronomical Society, 527, 9833
Cerardi N., Pierre M., Lanusse F., Corap X., 2025, Astronomy & Astrophysics, 701, A110
Choudhury T. R., Paranjape A., Maity B., 2024, J. Cosmology Astropart. Phys., 2024, 027
Cooper N., Norregaard C., Meriot R., Pritchard J. R., 2025, arXiv preprint arXiv:2508.16329
Cranmer K., Brehmer J., Louppe G., 2020, Proceedings of the National Academy of Science, 117, 30055
Curtis-Lake E., et al., 2023, Nature Astronomy, 7, 622
Datta K. K., Mellema G., Mao Y., Iliev I. T., Shapiro P. R., Ahn K., 2012, MNRAS, 424, 1877
Davies F. B., et al., 2018, The Astrophysical Journal, 864, 142
Elbers W., Weygaert R. v. d., 2019, Monthly Notices of the Royal Astronomical Society, 486, 1523
Fan X., Carilli C., Keating B., 2006, Annu. Rev. Astron. Astrophys., 44, 415

Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T., 2011, Monthly Notices of the Royal Astronomical Society, 413, 1353
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, The Astrophysical Journal, 613, 1
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Physics reports, 433, 181
Gan H., et al., 2023, Astronomy & Astrophysics, 669, A20
Georgiev I., Mellema G., Giri S. K., Mondal R., 2022, Monthly Notices of the Royal Astronomical Society, 513, 5109
Georgiev I., Mellema G., Giri S. K., 2025, Monthly Notices of the Royal Astronomical Society, 536, 3689
Ghara R., Choudhury T. R., Datta K. K., Choudhuri S., 2017, MNRAS, 464, 2234
Ghara R., Mellema G., Giri S. K., Choudhury T. R., Datta K. K., Majumdar S., 2018, Monthly Notices of the Royal Astronomical Society, 476, 1741
Ghara R., et al., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4728
Ghara R., Giri S. K., Ciardi B., Mellema G., Zaroubi S., 2021, Monthly Notices of the Royal Astronomical Society, 503, 4551
Ghara R., et al., 2024, Astronomy & Astrophysics, 687, A252
Ghara R., et al., 2025, Astronomy & Astrophysics, 699, A109
Giri S. K., Mellema G., 2021, Monthly Notices of the Royal Astronomical Society, 505, 1863
Giri S. K., Schneider A., 2022, Physical review D, 105, 083011
Giri S. K., Mellema G., Dixon K. L., Iliev I. T., 2018a, Monthly Notices of the Royal Astronomical Society, 473, 2949
Giri S. K., Mellema G., Ghara R., 2018b, Monthly Notices of the Royal Astronomical Society, 479, 5596
Giri S. K., D'Aloisio A., Mellema G., Komatsu E., Ghara R., Majumdar S., 2019, Journal of Cosmology and Astroparticle Physics, 2019, 058
Giri S., Mellema G., Jensen H., 2020, The Journal of Open Source Software, 5, 2363
Giri S. K., Bianco M., Schaeffer T., Iliev I. T., Mellema G., Schneider A., 2024, Monthly Notices of the Royal Astronomical Society, 533, 2364
Greig B., Mesinger A., 2015, Monthly Notices of the Royal Astronomical Society, 449, 4246
Greig B., Mesinger A., 2017, Monthly Notices of the Royal Astronomical Society, 472, 2651
Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2021a, Monthly Notices of the Royal Astronomical Society, 500, 5322
Greig B., et al., 2021b, Monthly Notices of the Royal Astronomical Society, 501, 1
Greig B., Mesinger A., Davies F. B., Wang F., Yang J., Hennawi J. F., 2022, Monthly Notices of the Royal Astronomical Society, 512, 5390
Greig B., Prelogović D., Qin Y., Ting Y.-S., Mesinger A., 2024, Monthly Notices of the Royal Astronomical Society, 533, 2530
HERA Collaboration et al., 2022, The Astrophysical Journal, 925, 221
HERA Collaboration et al., 2023, The Astrophysical Journal, 945, 124
Hoag A., et al., 2019, The Astrophysical Journal, 878, 12
Huberty M., Scarlata C., Hayes M. J., Gazagnes S., 2025, The Astrophysical Journal, 987, 82
Iliev I. T., Mellema G., Pen U.-L., Merz H., Shapiro P. R., Alvarez M. A., 2006, Monthly Notices of the Royal Astronomical Society, 369, 1625
Jeffrey N., Alsing J., Lanusse F., 2020, Monthly Notices of the Royal Astronomical Society, 501, 954–969
Jensen H., et al., 2013, MNRAS, 435, 460
Jones G. C., et al., 2025, Monthly Notices of the Royal Astronomical Society, 536, 2355
Kapahtia A., Chingangbam P., Ghara R., Appleby S., Choudhury T. R., 2021, Journal of Cosmology and Astroparticle Physics, 2021, 026
Kern N. S., Liu A., Parsons A. R., Mesinger A., Greig B., 2017, The Astrophysical Journal, 848, 23
Koopmans L., et al., 2015, arXiv preprint arXiv:1505.07568
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E., Puchwein E., Chardin J., Aubert D., 2019, Monthly Notices of the Royal Astronomical Society: Letters, 485, L24
Lanzieri D., Zeghal J., Lucas Makinen T., Boucaud A., Starck J.-L., Lanusse F., 2025, Astronomy & Astrophysics, 697, A162

Lemos P., Coogan A., Hezaveh Y., Perreault-Levasseur L., 2023, Sampling-Based Accuracy Testing of Posterior Estimators for General Inference (arXiv:2302.03026), https://arxiv.org/abs/2302.03026

Lewis A., Challinor A., Hanson D., 2011, Journal of Cosmology and Astroparticle Physics, 2011, 018

Liu A., Tegmark M., 2011, Physical Review D—Particles, Fields, Gravitation, and Cosmology, 83, 103006

Majumdar S., Pritchard J. R., Mondal R., Watkinson C. A., Bharadwaj S., Mellema G., 2018, Monthly Notices of the Royal Astronomical Society, 476, 4007

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., De Barros S., Vanzella E., 2018, The Astrophysical Journal, 856, 2

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

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

Mellema G., Koopmans L., Shukla H., Datta K. K., Mesinger A., Majumdar S., 2015, arXiv preprint arXiv:1501.04203

Meriot R., Semelin B., 2024, A&A, 683, A24

Meriot R., Semelin B., Cornu D., 2025, A&A, 698, A80

Mertens F. G., et al., 2020, Monthly Notices of the Royal Astronomical Society, 493, 1662

Mertens F. G., Bobin J., Carucci I. P., 2024, Monthly Notices of the Royal Astronomical Society, 527, 3517

Mertens F., et al., 2025, Astronomy & Astrophysics, 698, A186

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

Mevius M., et al., 2022, Monthly Notices of the Royal Astronomical Society, 509, 3693

Mirocha J., Munoz J. B., Furlanetto S. R., Liu A., Mesinger A., 2022, Monthly Notices of the Royal Astronomical Society, 514, 2010

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

Nunhokee C., et al., 2025, arXiv preprint arXiv:2505.09097

Papamakarios G., Murray I., 2018, Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation (arXiv:1605.06376), https://arxiv.org/abs/1605.06376

Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933

Pietschke Y., Heneka C., Schlenker T., Ore A., Schosser B., 2025, arXiv preprint arXiv:2506.19925

Planck Collaboration et al., 2020, A&A, 641, A4

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

Pranav P., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 4167

Prelogović D., Mesinger A., 2023, Monthly Notices of the Royal Astronomical Society, 524, 4239

Prelogović D., Mesinger A., 2024, Astronomy & Astrophysics, 688, A199

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

Qin Y., et al., 2025, Publications of the Astronomical Society of Australia, 42, e049

Saoulis A. A., Piras D., Jeffrey N., Spurio Mancini A., Ferreira A. M. G., Joachimi B., 2025, Monthly Notices of the Royal Astronomical Society, 542, 3231

Schaeffer T., Giri S. K., Schneider A., 2023, Monthly Notices of the Royal Astronomical Society, 526, 2942

Schmit C. J., Pritchard J. R., 2018, MNRAS, 475, 1213

Schneider A., Schaeffer T., Giri S. K., 2023, Physical Review D, 108, 043030

Scoccimarro R., 1998, Monthly Notices of the Royal Astronomical Society, 299, 1097

Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, The Astrophysical Journal, 496, 586

Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, MNRAS, 460, 3624

Semelin B., Mériot R., Mishra A., Cornu D., 2025, A&A, 698, A35

Tejero-Cantero A., Boelts J., Deistler M., Lueckmann J.-M., Durkan C., Gonçalves P. J., Greenberg D. S., Macke J. H., 2020, Journal of Open Source Software, 5, 2505

Trott C. M., Midgley S., Greig B., Pindor B., Cook J., Tingay S., Ung D.,

et al., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4711

Umeda H., Ouchi M., Nakajima K., Harikane Y., Ono Y., Xu Y., Isobe Y., Zhang Y., 2024, The Astrophysical Journal, 971, 124

Vasist M., Rozet F., Absil O., Mollière P., Nasedkin E., Louppe G., 2023, Astronomy & Astrophysics, 672, A147

Villaescusa-Navarro F., 2018, Pylians: Python libraries for the analysis of numerical simulations, Astrophysics Source Code Library, record ascl:1811.008 (ascl:1811.008)

Watkinson C. A., Majumdar S., Pritchard J. R., Mondal R., 2017, MNRAS, 472, 2436

Watkinson C. A., Giri S. K., Ross H. E., Dixon K. L., Iliev I. T., Mellema G., Pritchard J. R., 2019, Monthly Notices of the Royal Astronomical Society, 482, 2653

Zhao X., Mao Y., Cheng C., Wandelt B. D., 2022, ApJ, 926, 151

Zhao X., Mao Y., Zuo S., Wandelt B. D., 2024, ApJ, 973, 41

Zhou Y., La Plante P., 2022, Publications of the Astronomical Society of the Pacific, 134, 044001

de Salis E., et al., 2025, in EPIA Conference on Artificial Intelligence. pp 426–438

This paper has been typeset from a TeX/LaTeX file prepared by the author.