The impact of the WHIM on the IGM thermal state determined from the low-z Lyman-α forest

Hu, T.; Khaire, V.; Hennawi, J.F.; Oñorbe, J.; Walther, M.; Lukic, Z.; Davies, F.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **527,** 11338–11359 (2024) https://doi.org/10.1093/mnras/stad3846
Advance Access publication 2023 December 14

The impact of the WHIM on the IGM thermal state determined from the low-z Lyman α forest

Teng Hu [1]★ Vikram Khaire [1,2] Joseph F. Hennawi [1,3] Jose Oñorbe [4] Michael Walther [5,6] Zarija Lukic[7] and Frederick Davies [8]

[1]*Physics Department, Broida Hall, University of California, Santa Barbara, CA 93106-9530, USA*
[2]*Indian Institute of Space Science & Technology, Thiruvananthapuram, Kerala 695547, India*
[3]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[4]*Facultad de Física, Universidad de Sevilla, Avda. Reina Mercedes s/n, Campus de Reina Mercedes, E-41012 Sevilla, Spain*
[5]*Faculty of Physics, Ludwig-Maximilians-Universität München, University Observatory, Scheinerstr. 1, D-81679 Munich, Germany*
[6]*Excellence Cluster ORIGINS, Boltzmannstr. 2, D-85748 Garching, Germany*
[7]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[8]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*

Accepted 2023 December 8. Received 2023 December 7; in original form 2023 September 18

ABSTRACT

At $z \lesssim 1$, shock heating caused by large-scale velocity flows and possibly violent feedback from galaxy formation, converts a significant fraction of the cool gas ($T \sim 10^4$ K) in the intergalactic medium (IGM) into warm–hot phase (WHIM) with $T > 10^5$ K, resulting in a significant deviation from the previously tight power-law IGM temperature–density relationship, $T = T_0(\rho/\bar{\rho})^{\gamma-1}$. This study explores the impact of the WHIM on measurements of the low-z IGM thermal state, $[T_0, \gamma]$, based on the b–$N_{\rm H\,I}$ distribution of the Ly α forest. Exploiting a machine learning-enabled simulation-based inference method trained on Nyx hydrodynamical simulations, we demonstrate that $[T_0, \gamma]$ can still be reliably measured from the b–$N_{\rm H\,I}$ distribution at $z = 0.1$, notwithstanding the substantial WHIM in the IGM. To investigate the effects of different feedback, we apply this inference methodology to mock spectra derived from the IllustrisTNG and Illustris simulations at $z = 0.1$. The results suggest that the underlying $[T_0, \gamma]$ of both simulations can be recovered with biases as low as $|\Delta\log(T_0/{\rm K})| \lesssim 0.05$ dex, $|\Delta\gamma| \lesssim 0.1$, smaller than the precision of a typical measurement. Given the large differences in the volume-weighted WHIM fractions between the three simulations (Illustris 38 per cent, IllustrisTNG 10 per cent, and Nyx 4 per cent), we conclude that the b–$N_{\rm H\,I}$ distribution is not sensitive to the WHIM under realistic conditions. Finally, we investigate the physical properties of the detectable Ly α absorbers, and discover that although their T and Δ distributions remain mostly unaffected by feedback, they are correlated with the photoionization rate used in the simulation.

Key words: galaxy: formation – intergalactic medium – quasars: absorption lines.

1 INTRODUCTION

Being the largest reservoir of baryons in the Universe, the intergalactic medium (IGM) plays a crucial role in the evolution of the Universe and the formation of structures. Based on the canonical cosmological model constrained by many observational studies, the thermal evolution of the IGM is dominated by two major phase transition events of the Universe. The first phase transition is hydrogen reionization caused by the first generation of galaxies at redshift $6 < z < 20$ (Becker et al. 2001; Fan et al. 2006; McGreer, Mesinger & D'Odorico 2015; Robertson et al. 2015). The second one is the double reionization of helium (He II→He III) driven by quasi-stellar objects (QSOs; see e.g. Madau & Meiksin 1994; McQuinn et al. 2009; Khaire 2017), which is believed to occur at $z \sim 3$ (Worseck et al. 2011, 2018; Syphers & Shull 2014), where the quasar luminosity function reaches its peak (see e.g. Hopkins, Richards & Hernquist 2007; Khaire & Srianand 2015; Kulkarni, Worseck & Hennawi 2019). These two-phase transition events heat up the IGM dramatically to a maximum of 15 000 K while ionizing the IGM.

After the completion of hydrogen reionization ($\Delta z \sim$1–2), the IGM thermal state is shaped by the quasi-equilibrium balance between photoionization heating from the extragalactic UV background (Haardt & Madau 2012; Khaire & Srianand 2019; Faucher-Giguère 2020) and various cooling processes including recombinations, excitation, cooling due to Hubble expansion, and inverse Compton scattering of electrons off of the cosmic microwave background (CMB; see e.g. McQuinn 2016). All these processes together drive the IGM to follow a power-law temperature–density (T–Δ) relation:

$$T(\Delta) = T_0\Delta^{\gamma-1}, \tag{1}$$

where $\Delta = \rho/\bar{\rho}$ is the overdensity, T_0 is the temperature at mean density, and γ is the adiabatic index (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016). These two parameters $[T_0, \gamma]$ thus characterize the thermal state of the IGM, making it feasible to constrain the IGM thermal history (Lidz et al. 2010; Becker

★ E-mail: tenghu@ucsb.edu

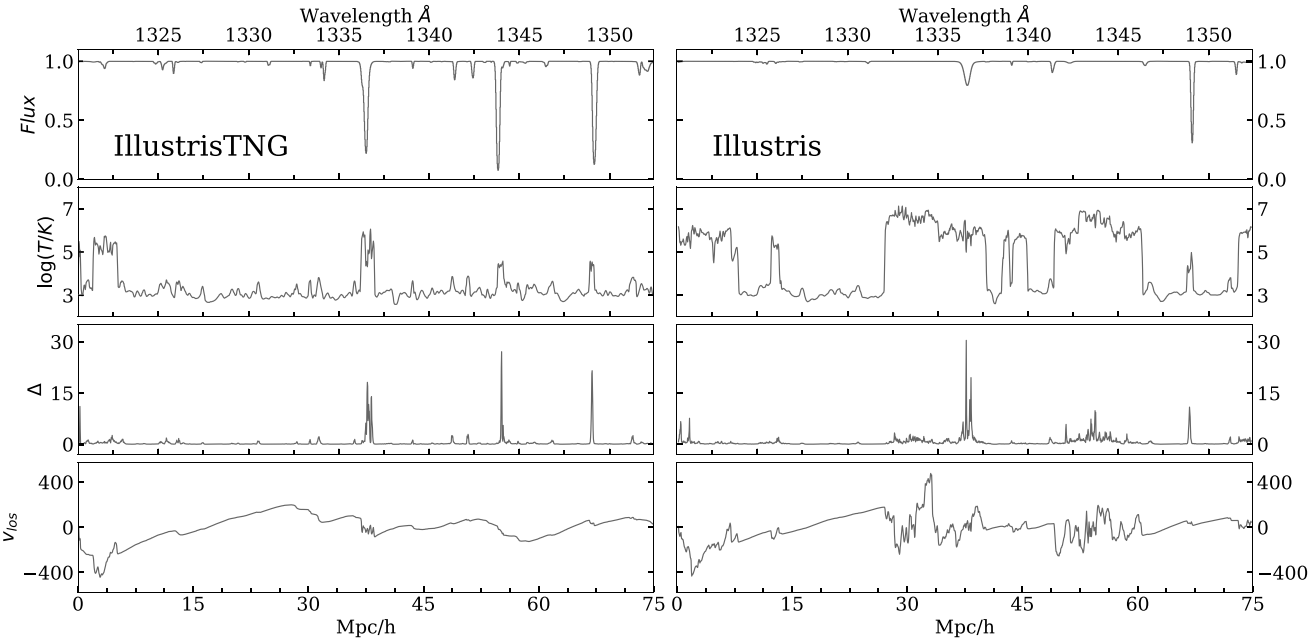

Figure 1. Examples of simulation skewers for IllustrisTNG (left) and Illustris (right) simulations, probing the structure generated by the same initial condition, while the two simulations are post-processed to share the same UV backgrounds photoionization rate, $\Gamma_{\rm H\,I}$. The flux is plotted in black on the top panel, while the temperature T, overdensity Δ, and velocity along LOS $v_{\rm los}$ are shown in black in the second, third, and bottom panels consecutively.

et al. 2011; Rorai et al. 2017; Hiss et al. 2018; Walther et al. 2019; Gaikwad et al. 2021) by measuring $[T_0, \gamma]$ at different epochs. These measurements improve our knowledge of the IGM thermal evolution and shed light on the underlying heating and cooling processes involved.

Nevertheless, the aforementioned power-law T–Δ relationship for the IGM could potentially break down at $z \lesssim 1$, where shock heating caused by large-scale velocity flows (Nath & Silk 2001; Cen & Ostriker 2006) and various feedback mechanisms become more common (Scannapieco, Silk & Bouwens 2005; Khaire et al. 2024). Specifically, shock heating at low-z converts a notable fraction of the cool IGM into warm hot intergalactic medium (WHIM) with $T > 10^5$ K (Shull et al. 2012), causing a substantial dispersion in the IGM T–Δ distribution (Davé et al. 2001; Cen & Ostriker 2006). As a result of such dispersion, the IGM T–Δ distribution can no longer be fully described by the typical power-law relationship (see Fig. 1), which introduces additional complexities in the measurement of the IGM thermal state (Hu et al. 2022, hereafter Hu22). The imperative question is whether the significant shock heating at low-z influences the observable, i.e. the Ly α forest, which serves as the primary probe of the IGM, and if it does, how might such impacts affect measurements of the IGM thermal state $[T_0, \gamma]$?

In practice, the IGM thermal state can be measured through various statistical properties of the Ly α forest. Particularly, at $z \lesssim 3$, the Ly α forest is amenable to Voigt profile decomposition (see Hiss et al. 2018), where each line can be fit by three parameters: redshift $z_{\rm abs}$, Doppler broadening b, and neutral hydrogen column density $N_{\rm H\,I}$. The IGM thermal state at these redshifts can thus be measured using the 2D joint b–$N_{\rm H\,I}$ distribution (Schaye et al. 1999, 2000; Bolton et al. 2014; Hiss et al. 2018; Rorai et al. 2018). Hu22 introduced a new inference method to measure the thermal state $[T_0, \gamma]$ and the photoionization rate $\Gamma_{\rm H\,I}$ of the IGM based on the b–$N_{\rm H\,I}$ distribution and Ly α line density, dN/dz, of the Ly α forest. Such a method performs Bayesian inference with the help of neural networks and Gaussian emulators, trained on a suite of Nyx simulations (Almgren et al. 2013; Lukić et al. 2015), making it possible to measure the thermal state of the IGM to high precision for realistic mock data sets.

Moreover, the thermal state of the IGM at $z < 1.7$ remains poorly constrained, since the Ly α transition below such redshift lies below the atmospheric cutoff ($\lambda \sim 3300$ Å), requiring UV observations from space with Hubble Space Telescope (HST). After He II reionization ($z < 3$), the thermal state of the IGM is considered to be dominated by adiabatic cooling from Hubble expansion, which leads to an IGM thermal state with $T_0 \sim 5000$ K and $\gamma \sim 1.6$ at the current epoch $z = 0$ (McQuinn & Upton Sanderbeck 2016). However, such a prediction of low temperatures has not yet been confirmed observationally. Meanwhile, recent studies have suggested that the Ly α lines appear broader than predicted by numerical simulations at $z < 0.5$ (Gaikwad et al. 2017; Nasir et al. 2017; Viel et al. 2017; Burkhart et al. 2022; Bolton et al. 2022a). This observation is based on the b parameters acquired from the HST Cosmic Origins Spectrograph (COS) spectra (Danforth et al. 2016, referred hereafter as D16) data set. While it has been argued that such a mismatch might be resolved by additional sources of turbulence, an alternative explanation would be that the IGM is actually hotter than previously presumed, with T_0 conceivably approaching 10 000 K, implying the existence of unexpected sources of heating (Bolton et al. 2022a, b), which, if true, would change our understanding of the IGM physics thoroughly.

In this paper, we adopt the Hu22 inference method to investigate the impact of the WHIM on measurements of the IGM thermal state, $[T_0, \gamma]$, based on the b–$N_{\rm H\,I}$ distribution of the Ly α forest. First, we assess the effectiveness of $[T_0, \gamma]$ as IGM parameters at low-z by comparing its performance as neural network training labels against the photoheating labels $[A, B]$ (see Section 2.3). These latter labels are photoheating rate rescaling factors used to generate the Nyx simulation suite with various thermal histories (see e.g Becker et al. 2011). Since our emulators are trained on these Nyx simulations generated by varying $[A, B]$, the inference method is naturally inclined to retrieve these photoheating labels. On the

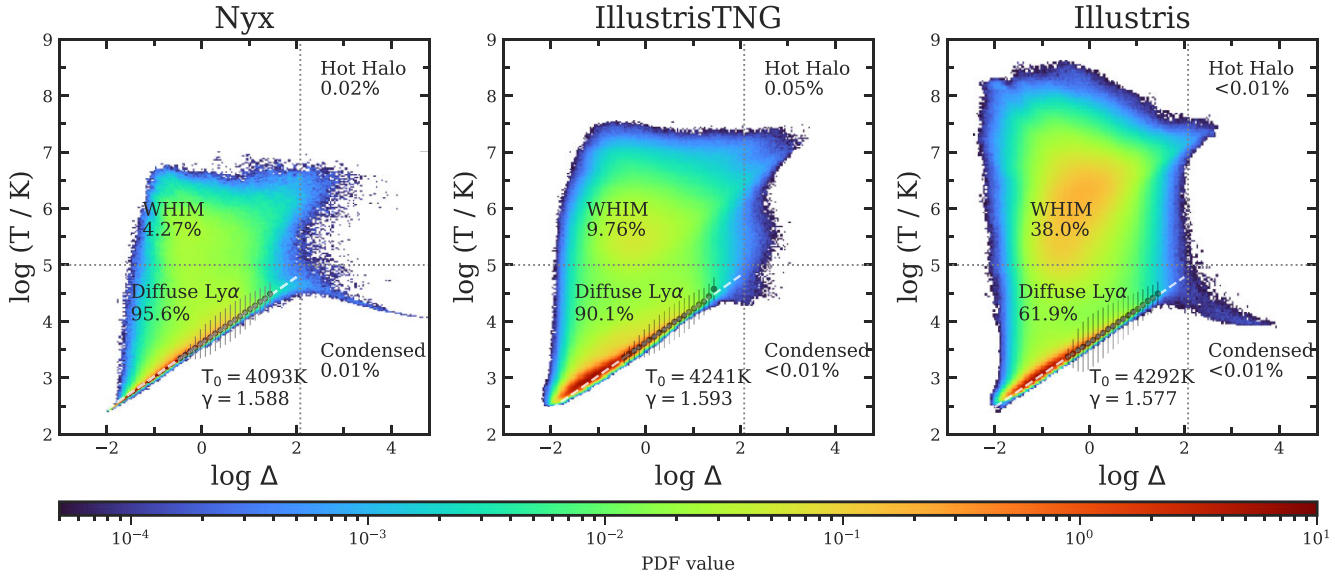

Figure 2. Volume weighted T–Δ distribution for all three simulations at $z = 0.1$. The $\log T$ for each bin are plotted as solid dots with 1–σ_T error bars. The best-fitting power-law relationship is shown as dashed lines. The Nyx (left) model is the default model which has $\log(T_0/\mathrm{K}) = 3.612$, $\gamma = 1.588$; and IllustrisTNG (middle) yields $\log(T_0/\mathrm{K}) = 3.627$, $\gamma = 1.593$; whereas Illustris (right) has $\log(T_0/\mathrm{K}) = 3.633$, $\gamma = 1.577$. The gas phase fractions are shown in the annotation.

other hand, if shock heating at low-z causes the T–Δ distribution of the Ly α absorbers to deviate from the power-law relationship, the effectiveness of $[T_0, \gamma]$ as labels could be compromised. Thus, our comparison between these two sets of labels provides insight into the robustness of $[T_0, \gamma]$ as IGM parameters at low-z, in the presence of substantial shock heating.

Afterwards, we explore the potential effects of different feedback mechanisms, which are associated with WHIM, on measurements of the IGM thermal state, $[T_0, \gamma]$. In terms of our inference methodology, the question becomes: what would happen if we used a simulation grid without feedback to interpret a Universe that includes feedback? Would this lead to unbiased $[T_0, \gamma]$? To answer these questions, we apply the Hu22 inference methodology to mock data drawn from the Illustris (Genel et al. 2014) and IllustrisTNG (Weinberger et al. 2017) simulations at $z = 0.1$. These two simulations incorporate galaxy formation models and feedback mechanisms that are not included in the Nyx simulation, which heat up the IGM substantially at low-z, and transform the cool diffuse Ly α gas into WHIM more effectively compared with Nyx simulation (see Fig. 2). We examine the inference results based on these two simulations and explore whether feedback biases the measurement of the thermal state $[T_0, \gamma]$.

To further investigate this problem, we explore the specific impacts of shock heating and other astrophysical processes, such as active galactic nucleus (AGN) feedback and UV background photoionization, on the physical properties of the Ly α forest at $z = 0.1$. Within the three aforementioned simulations, we identify simulated Ly α absorbers in the simulations and establish a direct correlation between the physical properties of these absorbers (including temperature T, overdensity Δ, and line-of-sight velocity $v_{\rm los}$) and the observed Ly α line parameters (b, $N_{\rm H\,I}$) derived from the absorption lines detected in corresponding mock spectra. We then examine the distributions of Δ and T of these simulated Ly α absorbers across the three aforementioned simulations to study the detailed effects of the feedback and UV background photoionization rate, $\Gamma_{\rm H\,I}$, on the Ly α forest.

This paper is organized as follows: In Section 2, we outline the simulations and associated processes applied to generate synthetic Ly α forest. It includes post-processing, forward-modelling, and Voigt profile fitting. The inference framework and results for all three simulations are then presented in Section 3. Section 4 is dedicated to the investigation of the physical characteristics of low-redshift Ly α forest absorbers in all three simulations. Finally, in Section 5, we present a summary and discussion of our findings. For the sake of brevity, we use log as a shorthand to denote $\log_{10}$ throughout the paper.

2 SIMULATIONS

In this paper, we utilize the inference framework described in Hu22, which employs the b–$N_{\rm H\,I}$ distribution emulator built on neural networks trained on a set of Nyx simulations. We also use galaxy formation simulations IllustrisTNG and Illustris to investigate the low-z Ly α forest under different feedback mechanisms. Since this work focuses on the low redshift Ly α forest, we use $z = 0.1$ simulation snapshots for all three simulations. In this section, we first provide a description of the simulations and the implemented physical models, followed by the (mock) data processing procedures employed in our study. This includes the generation of simulated line-of-sight (LOS) of Ly α forest (hereafter referred to as skewers for simplicity), forward modelling, and the Voigt profile fitting of Ly α lines. The cosmological parameters and thermal states of the three simulations are summarized in Table 1.

2.1 Nyx

Nyx is an adaptive mesh, massively parallel, cosmological simulation code primarily developed to simulate the IGM (Almgren et al. 2013; Lukić et al. 2015). Nyx simulates the dark matter evolution by treating the dark matter as self-gravitating Lagrangian particles, while it models baryons as an ideal gas on a uniform Cartesian grid following an Eulerian approach. The Eulerian hydrodynamics equations are

Table 1. Parameters of cosmology and T–Δ relation (at $z = 0.1$).

Parameters	Nyx	IllustrisTNG	Illustris
Ω_m	0.3192	0.3089	0.2726
Ω_Λ	0.6808	0.6911	0.7274
Ω_b	0.0496	0.0486	0.0456
h	0.670	0.677	0.704
σ_8	0.8288	0.8159	0.809
n_s	0.96	0.97	0.963
T_0	4093 K	4241 K	4292 K
γ	1.588	1.593	1.577

solved using a second-order piece-wise parabolic method, which is capable of accurately capturing shocks.

Nyx includes the major astrophysical processes relevant to the evolution of the the Ly α forest. First, gas in the Nyx simulation is treated as having a primordial composition with a hydrogen mass fraction of 0.76 and helium mass fraction of 0.24, and zero metallicity. Nyx takes into account the process of inverse Compton cooling off the microwave background and tracks the total thermal energy loss due to atomic collisional processes. Nyx also implements recombination, collisional ionization, dielectric recombination, and cooling following the prescription given in Lukić et al. (2015). Ionizing radiation in Nyx is modelled by a spatially uniform but time-varying ultraviolet background radiation field of Haardt & Madau (2012), while assuming all cells in the simulation are optically thin. Furthermore, following standard practice, we allow the UV background photoionization rate, $\Gamma_{\rm H\,I}$, to be a free parameter in post-processing while generating mock Ly α skewers. Lastly, Nyx does not implement any galaxy formation or feedback, which simplification reduces the required computational resources significantly, allowing us to run a large ensemble of simulations with different thermal histories (see Section 2.3), which is required for accurate statistical inference.

Each Nyx simulation model used in this study was initialized with the same initial condition at $z = 159$ and evolved down to $z = 0.03$ in a $L_{\rm box} = 20$ cMpc h^{-1} simulation box with $N_{\rm cell} = 1024^3$ Eulerian cells and 1024^3 dark matter particles. The box size is a compromise between computational cost and the need for convergence at least to < 10 per cent on small scales (large k). In short, such choices of box size and resolution should not affect the line parameters of the Ly α forest significantly. More discussion of the resolution, box size, and convergence issues can be found in Lukić et al. (2015) and Hu et al. (2022).

2.2 IllustrisTNG and Illustris

To evaluate the effectiveness of the IGM thermal state $[T_0, \gamma]$ as the IGM parameters and test the efficacy of our inference framework on the realistic IGM, which can be affected by astrophysical processes that are not included in Nyx simulation such as galaxy formation and AGN feedback, we employ Illustris (Genel et al. 2014) and IllustrisTNG (Weinberger et al. 2017; Nelson et al. 2019) simulations, and use them as mock observational data in our inference method.

The IllustrisTNG and Illustris are cosmological hydrodynamic simulations powered by the AREPO code (Springel 2010). This code employs a moving mesh approach to solve hydrodynamics through the Euler equations, and it computes gravitational forces on a quasi-Lagrangian moving Voronoi mesh via the tree-PM method. Both simulations incorporate a wide range of astrophysical processes for galaxy formation, such as star formation, stellar and AGN feedback, galactic winds, and chemical enrichment (Marinacci et al. 2018; Naiman et al. 2018; Springel et al. 2018). They utilize the UV background detailed in Faucher-Giguère et al. (2009) for photoionization heating and cooling. Other processes for modelling the Ly αforest, like collisional ionization and inverse Compton cooling from the cosmic microwave background, are also taken into account.

The primary distinction between IllustrisTNG and Illustris lies in their AGN feedback mechanisms, especially regarding AGN feedback. Both simulations implement AGN feedback in two modes based on the gas accretion rate on to the central supermassive black hole: the 'quasar-mode' at high accretion rates (Springel 2005; Hopkins et al. 2008; Debuhr, Quataert & Ma 2011) and the 'radio-mode' at low rates (Bower et al. 2006; Croton et al. 2006; Sijacki et al. 2007). While both use continuous thermal feedback in 'quasar-mode', their 'radio-mode' implementations differ. Illustris employs a bubble model for radio-mode feedback, accumulating substantial feedback energy for explosive release, often ejecting excessive hot gas (Genel et al. 2014). Conversely, IllustrisTNG models this feedback as a kinetic wind, injecting momentum into neighbouring regions from the central black hole. This approach better replicates astrophysical properties like star formation rates and galaxy colour distributions (Nelson et al. 2018; Pillepich et al. 2018a, b).

Both the IllustrisTNG and Illustris simulations we used in this study have box sizes of 75 cMpc h^{-1} and 1820^3 baryon and dark matter particles. Since AREPO is a moving mesh code, we convert the Voronoi mesh outputs to 1820^3 cartesian grids by dumping the smoothed quantities such as temperature.[1], density, and velocities on grids to generate Ly α forest skewers. A Gaussian kernel with a size equal to 2.5 times the radius of each Voronoi cell is applied for the smoothing, assuming each Voronoi cell is spherical. We then generate skewers for IllustrisTNG and Illustris simulations following the approach discussed in Section 2.4. In Fig. 1, we plot two simulation skewers for IllustrisTNG and Illustris respectively, while the two simulations are post-processed to share the same UV backgrounds photoionization rate, $\Gamma_{\rm H\,I}$ (see Section 2.5 for more discussion). The flux ($e^{-\tau}$) is plotted in the top panel, and the temperature, over-density, and LOS velocity profiles are shown in the second, third, and bottom panels consecutively. It is worth mentioning that the two skewers probe the structure generated by the same initial condition, suggesting that the differences in T, Δ, and $v_{\rm los}$ are primarily caused by different feedback strengths. Specifically, the Illustris exhibits higher temperatures due to its stronger feedback, which results in weaker absorption features given the same UV backgrounds. More discussion on the differences between Ly α forest in IllustrisTNG and Illustris simulations can be found in Khaire et al. (2024) and Khaire et al. (2023).

2.3 IGM thermal state and parameter grid

Following Hu22, we make use of the Thermal History and Evolution in Reionization Models of Absorption Lines (THERMAL[2]) suite of Nyx simulations (see Hiss et al. 2018; Walther et al. 2019) to model the IGM with various thermal histories. The suite consists of 48 models with varying thermal histories, each generated by changing

[1]As presented in Appendix A of Martizzi et al. (2019), there exists a bug that affects the IGM temperature of the IllustrisTNG simulation. However, its effect on the Ly α forest is minimal because the bug predominantly impacts the gas with the lowest density. Consequently, we continue to use the uncorrected temperature for the IllustrisTNG simulation.
[2]For details of THERMAL suite, see http://thermal.joseonorbe.com

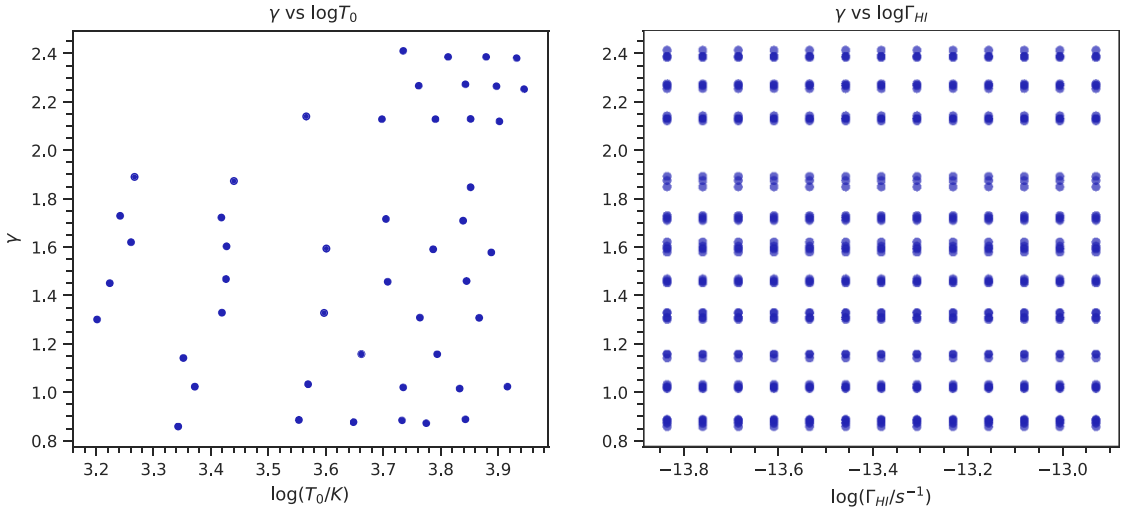

Figure 3. Parameters grid (solid circles) from snapshots of Nyx simulations from the THERMAL suite at $z = 0.1$, parametrized by the thermal state $[T_0, \gamma]$. The left-hand panel shows the γ–T_0 grid, whose shape is determined by the photoheating labels $[A, B]$ (see Fig. A3) and the evolution of the thermal state of the IGM. The right-hand panel is γ–$\Gamma_{\rm H\,I}$ grid, showing the 13 $\Gamma_{\rm H\,I}$ values for each point on the γ–T_0 grid.

the photoheating rate of the simulation following the prescription described in Becker et al. (2011), in which the photoheating rate, ϵ is assumed as a function of overdensity, i.e.

$$\epsilon = \epsilon_{\rm HM12}(z) A \Delta^B, \tag{2}$$

where $\epsilon_{\rm HM12}(z)$ stands for the time-varying photoheating rate per H II ion tabulated in Haardt & Madau (2012), and the constants A and B are free photoheating parameters that are varied in the different Nyx runs to achieve different thermal histories, which results in different thermal states at $z = 0.1$. The distribution of parameters in our thermal grid, i.e. the different values of $[T_0, \gamma]$ are illustrated in Fig. 3, and the corresponding values of $[A, B]$ are presented in Fig. A3 (see Appendix A for more discussion).

Conventionally, the thermal parameters $[T_0, \gamma]$ are obtained by fitting a power law to the T–Δ relationships (see equation 1). Such a fitting procedure is straightforward at higher redshift ($z \gtrsim 2$) where the T–Δ distributions of the IGM are tight. However, in low-z, the distributions of the IGM temperatures are noticeably broader due to the extensive shock heating, which heats up the IGM, resulting in more WHIM. The T–Δ distributions for all three simulations (Nyx default model with A = 1, B = 0, and IllustrisTNG and Illustris) are shown in Fig. 2. For each simulation, the gas is divided into four phases depending on the temperature and density, namely the WHIM, Diffuse Ly α, hot halo gas, and condensed, where the cutoffs are set to be $T = 10^5$ K and $\Delta = 120$[3]. It can be seen that there exist significant dispersion in the T–Δ distributions of the low-z IGM, i.e. the shock-heated WHIM, for all three simulations, and the fractions of the WHIM are directly proportional to the strength of the feedback. Specifically, $f_{\rm WHIM,\ Illustris} > f_{\rm WHIM,\ IllustrisTNG} > f_{\rm WHIM,\ Nyx}$, while Illustris implements extreme feedback, IllustrisTNG employs mild feedback and Nyx has no feedback.

In order to fit the power-law relationship in the presence of dispersion in the IGM T–Δ distribution, we utilize the fitting procedure presented in Hu22, which fits the power-law T–Δ relationship by binning the Diffuse Ly α gas ($T < 10^5$ K and $\Delta < 120$) into 20 bins based on $\log \Delta$, and applying a least-squares linear fit to the mean temperatures of the gas in each bin. Here, we modify the fitting range to $-0.5 < \log \Delta < 1.5$.[4], which provides a more accurate representation of the Δ range of the Ly α absorbers at $z \sim 0.1$, which is the principal subject of this paper.

Such a fitting procedure is applied to all simulations used in this study, including all Nyx models and IllustrisTNG and Illustris simulations. The best-fitting power-law relationship based on $[T_0, \gamma]$ and the T–Δ distributions are illustrated in Fig. 2. The figure shows that although the three simulations yield very different overall T–Δ distributions, their thermal state T_0 and γ are however similar.

Furthermore, as described in Hu22, we vary the UV background photoionization rate, $\Gamma_{\rm H\,I}$, of the Nyx simulations in post-processing when the simulation skewers are generated, extending the parameter grid to $[\log T_0, \gamma, \log \Gamma_{\rm H\,I}]$. The value of $\Gamma_{\rm H\,I}$ we used in this study spans from $\log(\Gamma_{\rm H\,I}/{\rm s}^{-1}) = -13.834$ to -12.932 in logarithmic steps of 0.075 dex, which gives 13 values in total (see the right-hand panel of Fig. 3). The range of $\Gamma_{\rm H\,I}$ used here covers more than twice the range obtained by UV background models of Khaire & Srianand (2019) at $z = 0.1$, achieved by varying the spectral energy distribution of quasars. Note that the range also coves more than 2σ uncertainty in the $\Gamma_{\rm H\,I}$ measurements (Gaikwad et al. 2017; Khaire et al. 2019). In total, the 3D thermal grid consists of $48 \times 13 = 624$ Nyx models.

As mentioned earlier, T_0 and γ characterize the IGM thermal state at $z \gtrsim 2$, where the IGM is dominated by the power-law T–Δ relationship. However, their efficacy as parameters for the IGM thermal state remains uncertain at $z \lesssim 1$, where a significant fraction of the gas deviates from the power-law T–Δ relationship due to shock heating and feedback. In this paper, we evaluate the effectiveness of the thermal state $[T_0, \gamma]$ as IGM parametrization at low-z using the inference framework presented in Hu22, and we make use of the photoheating parameters $[A, B]$ as an alternative set of labels as

[3]Here, we adopt the cutoff $T = 10^5$ K, and $\Delta = 120$ for different gas phases following Davé et al. (2010), and more discussion about the different cutoff can be found in Gaikwad et al. (2017).

[4]Such a choice of fitting range of the power law T–Δ relationship leads to slightly different thermal states $[T_0, \gamma]$ for the three simulations compared with those presented in previous works (Hu22; Khaire et al. 2024), but the difference is minor.

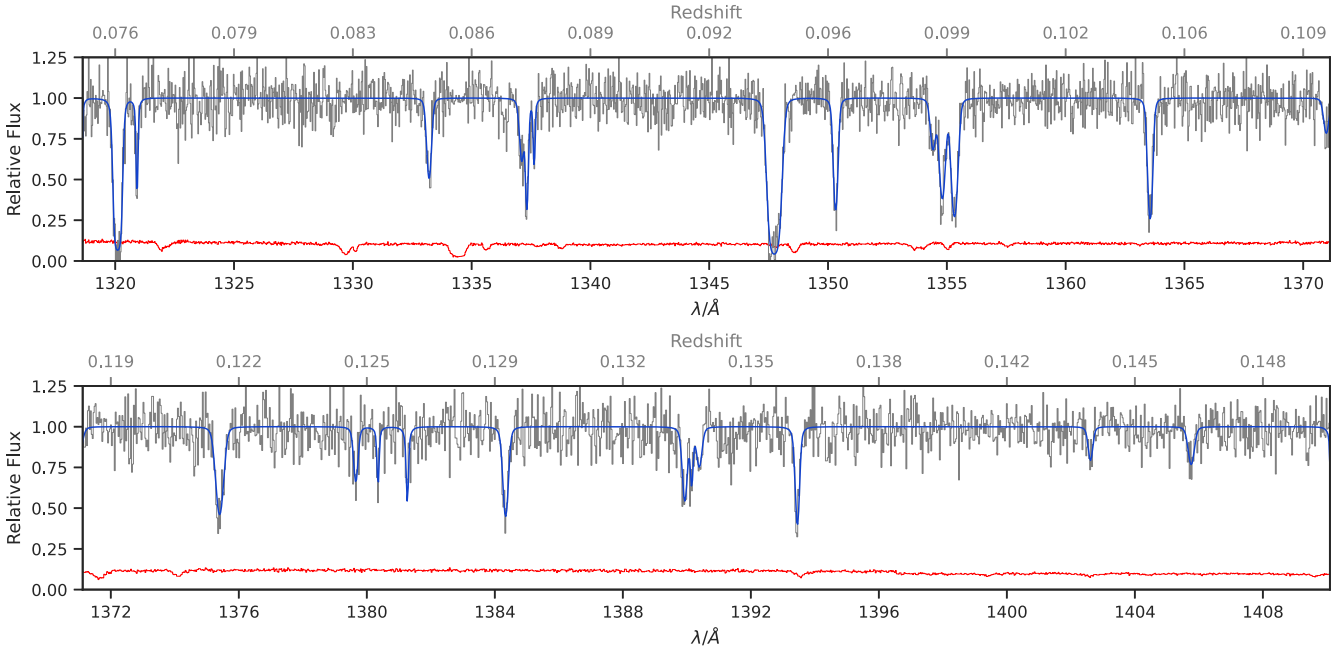

Figure 4. One of the forward-modelled mock spectra. The simulated spectrum is shown in black, where the model spectrum determined from VPFIT is shown in blue, and the noise vector is plotted in red.

a comparison. These labels are particularly relevant since all Nyx models used in the training procedure of our neural network, which is the major component of our inference method, are generated by varying $[A, B]$. This suggests that our inference framework should be capable of recovering the values of $[A, B]$ efficiently. Therefore, $[A, B]$ are particularly useful in the evaluation of the $[T_0, \gamma]$. More information about the photoheating labels $[A, B]$ is presented in Appendix. A.

2.4 Mock spectra, forward-modelling, and VPFIT

We follow the procedure described in Hu22 and generate mock spectra by calculating the Ly α optical depth (τ) array along the mock LOS. For each simulation, including all Nyx models and IllustrisTNG and Illustris, an ensemble of 20 000 skewers is created.

In this study, we include observational noise and instrumental effects to conduct our analysis under realistic conditions. We generate mock data sets with properties consistent with the D16 compilation of low redshift Ly α forest spectra, which comprises 82 unique quasar spectra observed with the COS on the HST. Among them, 34 spectra cover our targeted redshift range of $0.06 < z < 0.16$ and have signal-to-noise ratio (SNR)s greater than 5 per pixel. These spectra segments contribute to a total observational pathlength of $\Delta z_{\rm ob} = 2.136$. We follow the forward-modelling procedure described in Hu22. For each mock spectrum, we select one of the 34 spectra segments randomly, stitching skewers to match the wavelength grid of the selected segment[5]. We then convolve these combined skewers with the COS 130M LSF, and add Gaussian noise based on the noise vectors associated with the chosen spectrum segment. This process ensures that our mock data sets closely replicate the observational characteristics of the COS spectra, allowing for a more accurate and realistic analysis.

For each simulation model, including all Nyx simulation models as well as both IllustrisTNG and Illustris simulations, we generated 1000 forward-modelled mock spectra. The total pathlength for each simulation is approximately $\Delta z_{\rm tot} \sim 60$, which ensures that our training set for the b–$N_{\rm H\,I}$ distribution emulator (see Section 3.1) is large enough and the resulting b–$N_{\rm H\,I}$ distribution is not biased by the choice of (mock) spectra.

We then use VPFIT (Carswell & Webb 2014)[6] to fit the Ly α lines in our simulated spectra to obtain a set of $\{b, N_{\rm H\,I}\}$ pairs for all of the mock data sets, following the prescription given in Hu22. In this paper, as is the convention in low-z Ly α forest analysis, we apply a filter for both b and $N_{\rm H\,I}$, and uses only b–$N_{\rm H\,I}$ pairs with $12.5 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 14.5$ and $0.5 \leq \log(b/{\rm kms}^{-1}) \leq 2.5$ in our analysis (Schaye et al. 2000; Rudie, Steidel & Pettini 2012; Hiss et al. 2018). A segment of one of the forward-modelled mock spectra is shown in Fig. 4. The simulated spectrum is shown in grey, where the model spectrum determined from VPFIT is shown in blue, and the noise vector is plotted in red.

The top panels of Fig. 5 display 1D histograms of both b (left) and $N_{\rm H\,I}$ (right) for all three simulations, and the bottom panels illustrate their relative differences when compared to the Nyx simulation (as discussed in Section 2.5, the three simulations used here are dN/dz matched). The median value for $\log b$ and $\log N_{\rm H\,I}$ are indicated by dashed vertical lines for each simulation. Notably, while the median values of both b and $N_{\rm H\,I}$ are comparable across the three simulations, there are distinct differences in the distributions of both parameters across the three simulations. We also notice that the differences in the b parameters are more significant across the three simulations compared with $N_{\rm H\,I}$.

[5]The length of the Nyx skewers is 25 Mpc h^{-1}, corresponding to rough Δz = 0.01. Roughly 10 Nyx skewers are needed for one forwarded-modelled mock spectrum. Similarly, for Illustris and IllustrisTNG, which have box sizes of 75Mpc h^{-1}, about 3 skewers are used to generate one mock spectrum.

[6]VPFIT: http://www.ast.cam.ac.uk/~rfc/vpfit.html

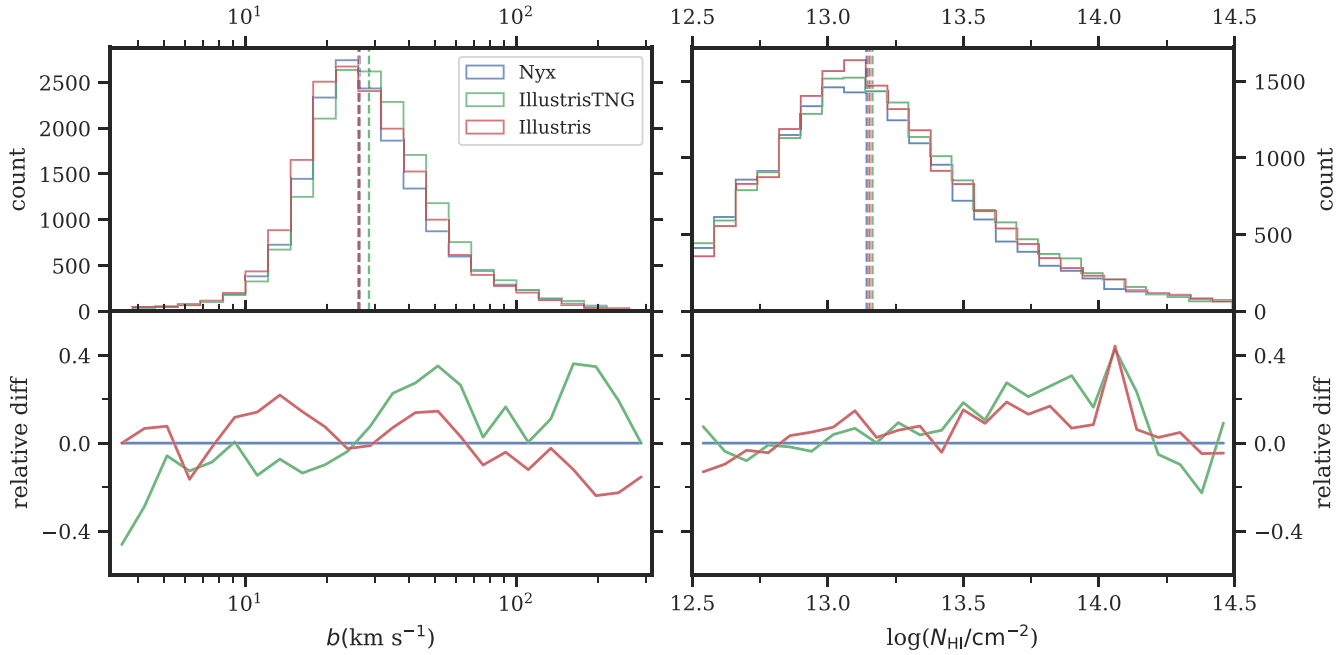

Figure 5. Top: Marginalized 1D b (left) and $N_{\rm H\,I}$ (right) distributions for all three simulations. For each simulation, the $\{b,\ N_{\rm H\,I}\}$ data set is obtained by VP-fitting an ensemble of 1000 forward-modelled mock spectra. The median values for $\log b$ and $\log N_{\rm H\,I}$ are indicated by dashed vertical lines. bottom: The relative difference compared with Nyx simulation. The three simulations used here are dN/dz matched.

2.5 Photoioniztion rate $\Gamma_{\rm H\,I}$ and dN/dz

It is noteworthy that the three simulations used in this study by default have different UV background photoionization rates $\Gamma_{\rm H\,I}$ (for Nyx, here we are referring to the default model with $\log(T_0/{\rm K}) = 3.612$ and $\gamma = 1.588$.) This is because the photoionization rate $\Gamma_{\rm H\,I}$ are tuned in post-processing across all three simulations to ensure they exhibit the same absorber density dN/dz as the one we measured from D16 data set at $z = 0.1$. Specifically, we apply the aforementioned VP-fitting procedure to D16 spectra (segments) with $0.06 < z < 0.16$, and obtain dN/dz = 167.3 for absorbers within the limits $12.5 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 14.5$ and $0.5 \leq \log(b/{\rm km s}^{-1}) \leq 2.5$. Such matching of d$N$/d$z$ is analogous to the matching of the mean flux of simulations at high-z for optically thin absorbers (Lukić et al. 2015). To match this dN/dz, we tune the photoionization rate, following the prescription described in Section 2.3, and set $\log(\Gamma_{\rm H\,I}, /{\rm s}^{-1}) = -13.093, -13.021, -13.414$ for Nyx, IllustrisTNG, and Illustris, respectively (see Fig. 6). Such mismatch in $\Gamma_{\rm H\,I}$ is caused by the degeneracy between the photoionization rate and different feedback recipes used in the simulations. Since both the UV background and feedback suppress the formation of Ly α absorbers (Khaire et al. 2024). More specifically, the feedback heats up the IGM, converting a significant amount of the diffuse Ly α gas into WHIM, which reduces the Ly α transmission caused by the neutral hydrogen H I in the cool diffuse Ly α gas. To this end, simulations with stronger feedback exhibit lower dN/dz under the same $\Gamma_{\rm H\,I}$.

We measure the dN/dz for the three simulations, including all Nyx simulation models and IllustrisTNG and Illustris, each based on its respective set of 1000 forward-modelled mock spectra. The relationships between UV background photoionization rate and dN/dz for all three models are shown in Fig. 6, where the dN/dz for Nyx is plotted in blue, IllustrisTNG in green, and Illustris in red, while the dN/dz for the D16 data at $z = 0.1$ is shown as the horizontal dashed–dotted grey line. Fig. 6 demonstrates that while Illustis has the strongest feedback, which causes more gas to be collisionally

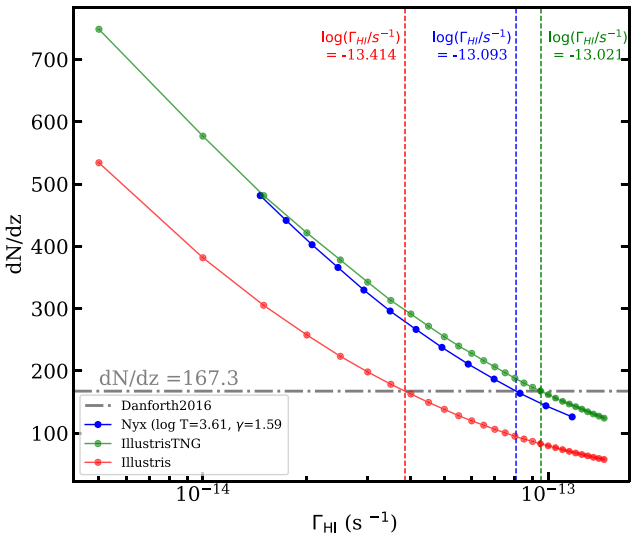

Figure 6. dN/dz versus $\Gamma_{\rm H\,I}$ for all three simulations at $z = 0.1$. Nyx (default model) is shown in blue, IllustrisTNG in green, and Illustris in red, while the observed dN/dz calculated from D16 data set for the corresponding redshift are shown in the horizontal grey dashed–dotted line. The $\Gamma_{\rm H\,I}$ values used for each simulation to match the observed dN/dz are indicated by vertical dashed lines with the corresponding colour.

ionized, reducing the Ly α absorption, it requires the lowest $\Gamma_{\rm H\,I}$ to match the dN/dz to the observed value, and IllustrisTNG, with mild feedback, has higher for the same UV background. Interestingly, whereas Nyx employs no feedback mechanism, it requires slightly lower $\Gamma_{\rm H\,I}$ compared with IllustrisTNG, which implements feedback, to achieve the same dN/dz. Such a trend, which is opposite to the correlation between dN/dz and feedback strength (as seen between Illustris and IllustrisTNG), is caused by the small difference in

the Δ, T distributions in Nyx and IllustrisTNG. More specifically, IllustrisTNG exhibits a slightly higher mass fraction of the diffuse Ly α gas in the particular Δ, T range that is probed by the Ly α forest[7]. It is possible that while the mild feedback in the IllustrisTNG simulation results in a slightly higher WHIM fraction compared with Nyx, it also produces more gas with $T \sim 10^{4.5}$ K and $\Delta \sim 10$, which is the Δ, T range probed by the Ly α forest at $z = 0.1$ (see Section 4). However, the detailed astrophysical mechanism leading to this specific Δ–T distribution in IllustrisTNG at $z = 0.1$ is still unclear, and we leave it to our future work.

If not otherwise specified, the three simulations used in this study, including Illustris, IllustrisTNG, and Nyx default model, are tuned to have the same Ly α line densities, with $\mathrm{d}N/\mathrm{d}z = 167.3$, which is the same value we measured from the D16 data set.

3 INFERENCE METHOD

3.1 Emulating the $\{b, N_{\rm H\,I}\}$ distribution

In this work, we make use of the inference framework following Hu22, which measures the thermal state and the photoionization rate $\Gamma_{\rm H\,I}$ of the low redshift IGM using its b–$N_{\rm H\,I}$ distribution and absorber line density $\mathrm{d}N/\mathrm{d}z$. The b–$N_{\rm H\,I}$ distribution emulator is built on density-estimation likelihood-free inference (DELFI), which turns inference into a density estimation task by learning the distribution of a data set as a function of the labels or parameters (Papamakarios & Murray 2016; Alsing, Wandelt & Feeney 2018; Lueckmann et al. 2018; Papamakarios, Sterratt & Murray 2018; Alsing et al. 2019). Following Hu22, we make use of `pydelfi`, the publicly available `python` implementation of DELFI,[8] which makes use of neural density estimation (NDE) to learn the sampling conditional probability distribution $P(\mathbf{d} \mid \boldsymbol{\theta})$ of the data summaries $\mathbf{d}$, as a function of labels/parameters $\boldsymbol{\theta}$, from a training set of simulated data. Here, the data summaries $\mathbf{d}$ are $[\log N_{\rm H\,I}, \log b]$, and our two sets of label parameters $\boldsymbol{\theta}$ are the thermal state $[\log T_0, \gamma, \log \Gamma_{\rm H\,I}]$ and photoheating labels $[A, B, \log \Gamma_{\rm H\,I}]$. The $\Gamma_{\rm H\,I}$ grids are identical for the two sets of labels.

We generate two training data sets by labelling the $\{b, N_{\rm H\,I}\}$ pairs obtained from our simulated spectra with the two sets of labels respectively. We train the neural network on the summary-parameter pairs for each training data set separately. Our b–$N_{\rm H\,I}$ distribution emulator learns the conditional probability distribution $P(b\ , N_{\rm H\,I} \mid T_0, \gamma, \log \Gamma_{\rm H\,I})$ and $P(b\ , N_{\rm H\,I} \mid A, B, \log \Gamma_{\rm H\,I})$ from the corresponding training data set. These conditional b–$N_{\rm H\,I}$ distributions are then used in our inference algorithm, where we try to find the best-fitting model given the observational/mock data set, which is described in the following section.

3.1.1 Likelihood function

In Bayesian inference, a likelihood $\mathcal{L} = P(\mathrm{data}|\mathrm{model})$ is used to describe the probability of observing the data for any given model. We adopt the likelihood formalism introduced in Hu22, which is summarized as follows:

$$\ln \mathcal{L} = \sum_{i=1}^{n} \ln(\mu_i) - \left(\frac{\mathrm{d}N}{\mathrm{d}z}\right)_{\rm model} \Delta z_{\rm data}, \tag{3}$$

where μ_i is the Poisson rate of an absorber occupying a cell in the b–$N_{\rm HI}$ plane with area $\Delta \mathrm{N}_{\mathrm{H\,I}i} \times \Delta bi$, i.e.

$$\mu_i = \left(\frac{\mathrm{d}N}{\mathrm{d}z}\right)_{\rm model} P(b_i, N_{\mathrm{H\,I},i} \mid \boldsymbol{\theta})\, \Delta N_{\rm H\,I}\, \Delta b\, \Delta z_{\rm data}. \tag{4}$$

The $P(bi, N_{\mathrm{H\,I}}i \mid \boldsymbol{\theta})$ in the equation is the probability distribution function at the point $(bi, N_{\rm H\,I}, i)$ for any given model parameters $\boldsymbol{\theta}$ evaluated by the DELFI b–$N_{\rm H\,I}$ distribution emulator described in Section 3.1. The $\Delta z_{\rm data}$ is the total redshift pathlength covered by the quasar spectra from which we obtain our $\{b, N_{\rm H\,I}\}$ data set, and $(\mathrm{d}N/\mathrm{d}z)_{\rm model}$ is the absorber density which is evaluated for any given set of parameters using a Gaussian process emulator (based on `George`, see Ambikasaran et al. 2016), which is also trained on our training data sets obtained from the Nyx simulation suite.

To perform our analysis under realistic conditions, all tests performed in this paper are based on mock data sets consisting of 34 forward-modelled spectra, each corresponding to one of the 34 D16 quasar spectra, which gives these data sets the same pathlength as the observation data set with $\Delta z_{\rm ob} = 2.136$. Each of the mock data sets is constructed by randomly selecting 34 spectra from the 1000 forward-modelled spectra, while making sure that each of the 34 D16 quasar spectra is represented exactly once, thereby maintaining the integrity and representativeness of our mock data sets.

An example of the MCMC posterior obtained based on the aforementioned likelihood function is given in Fig. 7. The inference is conducted using the labels $[T_0, \gamma, \log \Gamma_{\rm H\,I}]$. The posterior appears compact, with the medians of the marginalized posteriors landing close to the true parameters for all three parameters, i.e. within 1σ errors for marginalized 1D distributions. The b–$N_{\rm H\,I}$ distribution recovered from the mock data set is presented in Fig. 8, which is emulated by our b–$N_{\rm H\,I}$ distribution emulator, trained on $[T_0, \gamma, \log \Gamma_{\rm H\,I}]$, based on the inferred parameters, i.e. median values of the marginalized 1D MCMC posterior. The plot exhibits a good match between the mock data set (black dots) and the recovered b–$N_{\rm H\,I}$ distribution (colour map).

As a comparison to the IGM parametrization based on the thermal state, $[T_0, \gamma, \log \Gamma_{\rm H\,I}]$, the inference result derived from the same mock data set using the photoheating labels $[A, B, \log \Gamma_{\rm H\,I}]$ is given in Appendix. A.

3.2 Inference test

An inference test is an effective method to evaluate the robustness of a given inference algorithm, which usually consists of approximations and emulation/interpolation procedures that might induce additional uncertainties, altering the error budget. In practice, an inference test can be conducted by performing a set of realizations of the inference method using mock data sets and evaluating the robustness of the resulting posterior probability distributions, which can be quantified by the coverage probability $P_{\rm cov}$ (Prangle et al. 2014; Ziegel & Gneiting 2014; Morrison & Simon 2018; Sellentin & Starck 2019), the proportion of the time that the true parameters used to generate a mock data set are contained within the posterior contour corresponding to a certain probability level $P_{\rm inf}$. Such calculations can be performed for many different probability levels, resulting in a series of coverage probabilities. For perfect inference, this coverage probability $P_{\rm cov}$ is always equal to the probability level of the chosen posterior contour $P_{\rm inf}$ (shown as the black dashed line in Fig. 9).

[7]This is different from the mass-weighted or volume-weighted diffuse Ly α fraction, $f_{\mathrm{Ly}\,\alpha}$, which is defined to include all gas with $T \leq 10^5$ K and $\Delta \leq 120$ following Davé et al. (2010). Instead, the gas probed by the observed HST COS Ly α Forest has a narrow range of Δ and T which also depends on the $\Gamma_{\rm H\,I}$ value used in the simulation (see Section 4 for more details).
[8]See https://github.com/justinalsing/pydelfi

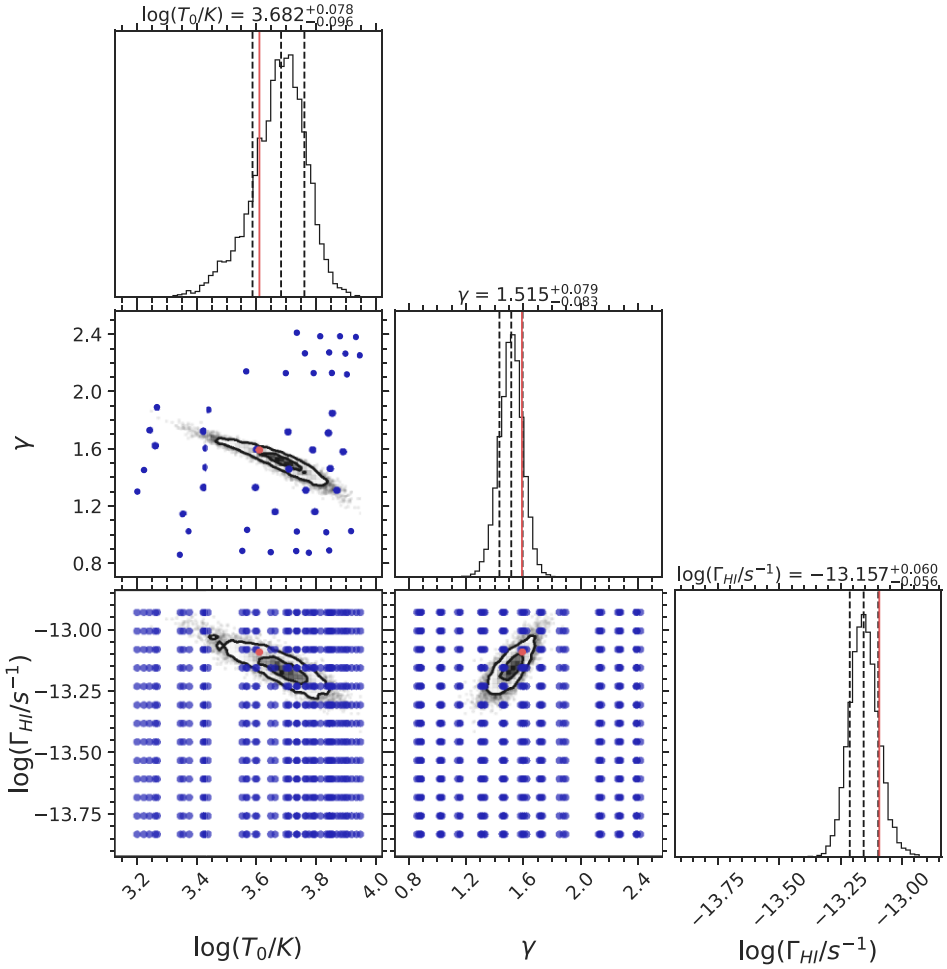

Figure 7. An example of posterior obtained by our inference method based on inference labels $[T_0, \gamma, \log \Gamma_{\rm H\,I}]$. Projections of the thermal grid used for generating models are shown as blue dots, while the true model is shown as red dots. The inner (outer) black contour represents the projected 2D 1(2)σ interval. Red lines in the marginal distributions indicate the parameters of true models, while the dashed black lines indicate the 16, 50, and 84 percentile values of the marginalized 1D posterior. The true parameters are: $\log(T_0/{\rm K}) = 3.612$ and $\gamma = 1.588$, while $\log(\Gamma_{\rm H\,I}/{\rm s}^{-1}) = -13.093$.

In this study, we make use of the inference test described in Wolfson et al. (2022), which calculates the coverage probability based on the MCMC posteriors. Compared with the one used in Hu22, this inference test algorithm is more precise and automatically returns full coverage probabilities from 0 to 1 rather than coverage probabilities at only a few specific probability levels.

To evaluate the effectiveness of $[T_0, \gamma, \log \Gamma_{\rm H\,I}]$ as IGM parameters for inference at low-z, where the IGM T-Δ distribution is no longer characterized by the power-law relationship, we perform inference tests based on different sets of labels. We compare the result of the inference test based on labels $[T_0, \gamma, \log \Gamma_{\rm H\,I}]$ with the one based on the photoheating labels $[A, B, \log \Gamma_{\rm H\,I}]$. As discussed in Section 2.3, the comparison between these two sets of labels sheds light on the efficacy of $[T_0, \gamma]$ as IGM parameters at low-z, where the pervasive shock heating causes significant dispersion in the IGM T–Δ distribution.

For each set of labels, we ran 300 realizations of our inference method, each based on a model randomly chosen from the grid. We exclude models that are close to the boundaries to mitigate boundary effects caused by the hard cutoff of the inference prior, which leads to the truncation of the posteriors. For $[T_0, \gamma, \log \Gamma_{\rm H\,I}]$ gird, we specify $3.3 < \log(T_0/{\rm K}) < 3.9$, $1.0 < \gamma < 2.3$, $-13.75 < \log(\Gamma_{\rm H\,I}/{\rm s}^{-1}) < -13.0$. We then calculate the full coverage probabilities based on all 300 MCMC posteriors.

The results of the inference tests are shown in Fig. 9, where the x-axis stands for inferred probability $P_{\rm inf}$, and the y-axis shows the coverage probability, $P_{\rm cov}$. The shaded regions indicate the 1-σ error for $P_{\rm cov}$, which is calculated based on the binomial distribution. The $y = x$ black-dashed line represents a perfect inference test. It can be seen that for Nyx simulations, our inference method is mildly overconfident, and the thermal state $[T_0, \gamma]$ (blue) performs slightly better than the photoheating labels $[A, B]$ (red), i.e. $P_{\rm cov}/P_{\rm inf}$ is closer to unity. Quantitatively, for inference based on the thermal state $[T_0, \gamma]$, the 68 per cent contour contains the true parameters 61.2 $\pm$ 2.8% of the time, and the 95 per cent contour contains the true parameters 90.4 $\pm$ 1.6% of the time. The results show that the $[T_0, \gamma]$ are still robust inference labels for the IGM at low-z, although the shock heating alters the T-Δ distribution. This further suggests that shock heating alone does not significantly change our understanding in determining the thermal state of the IGM using the Lyα forest. Lastly, while the general efficacy of the inference framework remains robust, we attribute its imperfections to two primary sources: deficiencies within the neural network used in our inference algorithm, and the boundary

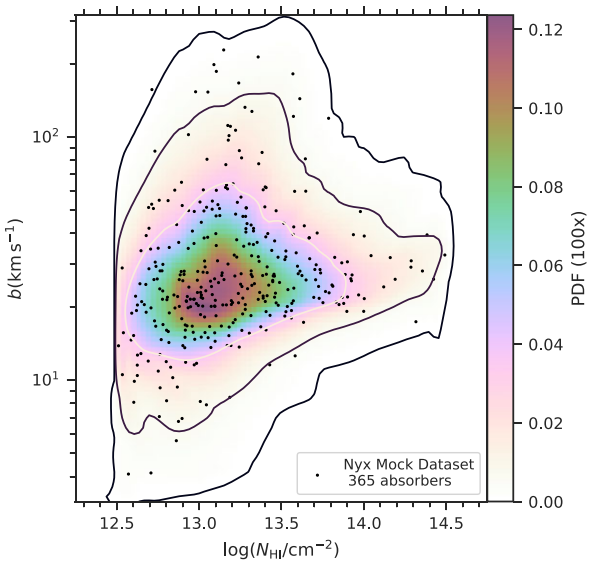

Figure 8. The colour map is the full b–$N_{\rm H\,I}$ distribution recovered from the Nyx mock data set, which is emulated by our DELFI emulator based on the best-fitting parameters (median values of the marginalized MCMC posterior), where $\log(T_0/{\rm K}) = 3.682$ (3.612) and $\gamma = 1.515$ (1.588) and $\log(\Gamma_{\rm H\,I}/{\rm s}^{-1}) = -13.157$ (-13.093), the true parameters are given in parentheses. Black dots are the mock data sets we used in the inference. The contours correspond to cumulative probabilities of 68, 95, and 99.7 per cent. For illustration purposes, the values of pdf are multiplied by 100 in the colour bar.

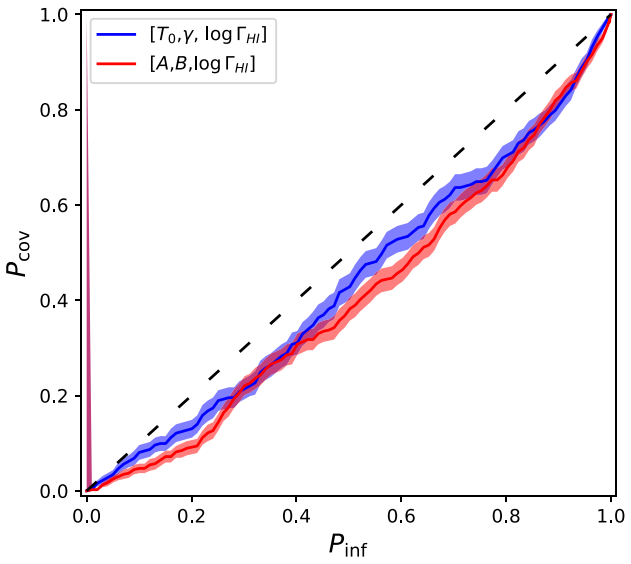

Figure 9. Coverage probability $P_{\rm cov}$ for inference tests based on different labels. The x-axis stands for the inferred probability $P_{\rm inf}$, and the y-axis shows the coverage probability $P_{\rm cov}$ for the true parameters to fall in the contour corresponding to $P_{\rm inf}$. Blue: Inference test based on the thermal state [T_0, γ, $\log\Gamma_{\rm H\,I}$], Red: Inference test based on the photoheating label [A, B, $\log\Gamma_{\rm H\,I}$]. The shaded regions indicate the 1σ error for $P_{\rm cov}$.

effects caused by the truncation of the posteriors when hitting the boundary.

3.3 Inference results for IllustrisTNG and Illustris

In this section, we employ the IllustrisTNG and Illustris simulations as mock observational data to explore the impacts of feedback, mainly AGN feedback, on the IGM thermal state [T_0, γ]. More specifically, we evaluate the robustness of our inference method, built on the Nyx thermal grid without galaxy formation and feedback, when applied to observational data derived from a (mock) Universe with substantial feedback associated with galaxy formation and AGN activities. The investigation is broken down into two separate inquiries. First, it explores the extent to which feedback associated with galaxy formation and AGN activities impacts the Ly α forest. Second, it investigates how, given the presence of these effects, the feedback influences the inferred parameters [T_0, γ].

Following the forward-modelling prescription described in Section 2.4, we generate mock data sets with $\Delta z = 2.136$, the pathlength of D16 data set at $z = 0.1$, for both simulations (see Fig. 11), and run our inference method on each data set. As discussed in Section 2.5, a degeneracy exists between the strength of the AGN feedback implemented in the simulations and the UV background photoionization rate $\Gamma_{\rm H\,I}$, both of which suppress the abundance of absorbers, hence reducing the dN/dz(see Khaire et al. 2024, for more details). Given that our inference method primarily derives the photoionization rate $\Gamma_{\rm H\,I}$ based on the dN/dz, the resulting $\Gamma_{\rm H\,I}$ always aligns with the value that generates the equivalent dN/dz in the Nyx simulation (see Section 2.5). Since here we use IllustrisTNG and Illustris simulations with their dN/dz matched to the D16 low-z data set, the inferred $\Gamma_{\rm H\,I}$ always disagrees with the true values used to generate the IllustrisTNG and Illustris simulations. To this end, we conduct our inference test in 2D without considering the accuracy with which we recover the photoionization rate $\Gamma_{\rm H\,I}$. Posterior distributions for the thermal parameters obtained from our inference applied to Illustris and IllustrisTNG are shown in Fig. 10, where we have marginalized over $\Gamma_{\rm H\,I}$. For these two mock data sets, we infer that $[\log(T_0/{\rm K}),\ \gamma] = [3.586^{+0.149}_{-0.074}, 1.658^{+0.060}_{-0.128}]$ for IllustrisTNG ([3.627,1.593]), and $[\log(T_0/{\rm K}),\ \gamma] = [3.696^{+0.080}_{-0.102}, 1.485^{+0.093}_{-0.094}]$ for Illustris ([3.633,1.577]), while the true parameters for the two simulations, $[T_0, \gamma]_{\rm fit}$, are given in parentheses respectively.

It can be seen that the true parameters $[T_0, \gamma]_{\rm fit}$, obtained by fitting the T–Δ distributions of the simulations, are within 1σ errors (1D marginalized) for both simulations, and the 1σ errors for both the IllustrisTNG and Illustris simulations are slightly larger than those for Nyx simulations, which is caused by the intrinsic difference between Nyx, IllustrisTNG, and Illustris simulations, where the latter two are based on completely different hydrodynamic codes. In Fig. 11, we present both the mock data sets used for inference and the b–$N_{\rm H\,I}$ distributions emulated based on the inference results. The plots highlight strong agreement between the emulated b–$N_{\rm H\,I}$ distributions and the respective mock data set for each simulation.

Nevertheless, it is worth mentioning that the inferred thermal states for IllustrisTNG and Illustris presented above are based on realistic conditions, with total pathlength $\Delta z = 2.136$. Such a small Δz makes the inference result vulnerable to randomness induced by the selection of mock data sets. To address this issue, here we conduct our inference on IllustrisTNG and Illustris simulations, using data sets with much larger pathlength, specifically with $\Delta z = 42.47$, which is 20 times the size of the observational data set. The inference results yield $[\log(T_0/{\rm K}),\ \gamma] = [3.605^{+0.031}_{-0.027}, 1.657^{+0.022}_{-0.024}]$ for IllustrisTNG ([3.627,1.59]), and $[\log(T_0/{\rm K}),\ \gamma] = [3.680^{+0.019}_{-0.020}, 1.483^{+0.021}_{-0.021}]$ for Illustris ([3.633,1.58]), while the true parameters for the two simulations, $[T_0, \gamma]_{\rm inf}$, are given in parentheses. The resulting corner plots are presented in Fig. 12. These results are used as our inferred thermal states $[T_0, \gamma]_{\rm inf}$ for IllustrisTNG and Illustris simulations in the following part of this study. It is noticeable that the inferred T_0 for Illustris is higher than the true value with an error $\Delta\log(T_0/{\rm K}) = 0.047$ dex, while the γ is below the true value, with $\Delta\gamma = -0.094$.

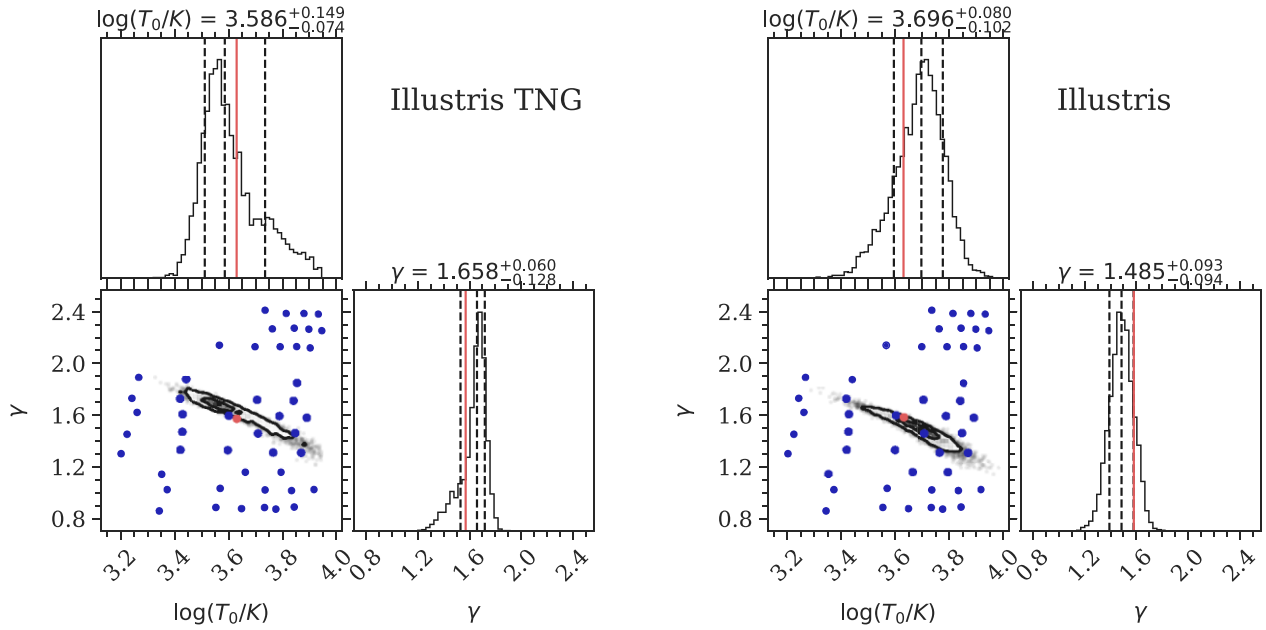

Figure 10. Posteriors obtained by using IllustrisTNG (left) and Illustris (right) as mock observational data in our inference method. Projections of the thermal grid used for generating models are shown as blue dots. The inner (outer) black contour represents the projected 2D 1(2)σ interval. The true parameters for the simulations, obtained by fitting the T–Δ distributions of the simulations, are indicated by the red dot (lines) in the (marginal) distributions, while the dashed black lines indicate the 16, 50, and 84 percentile values of the marginalized 1D posterior.

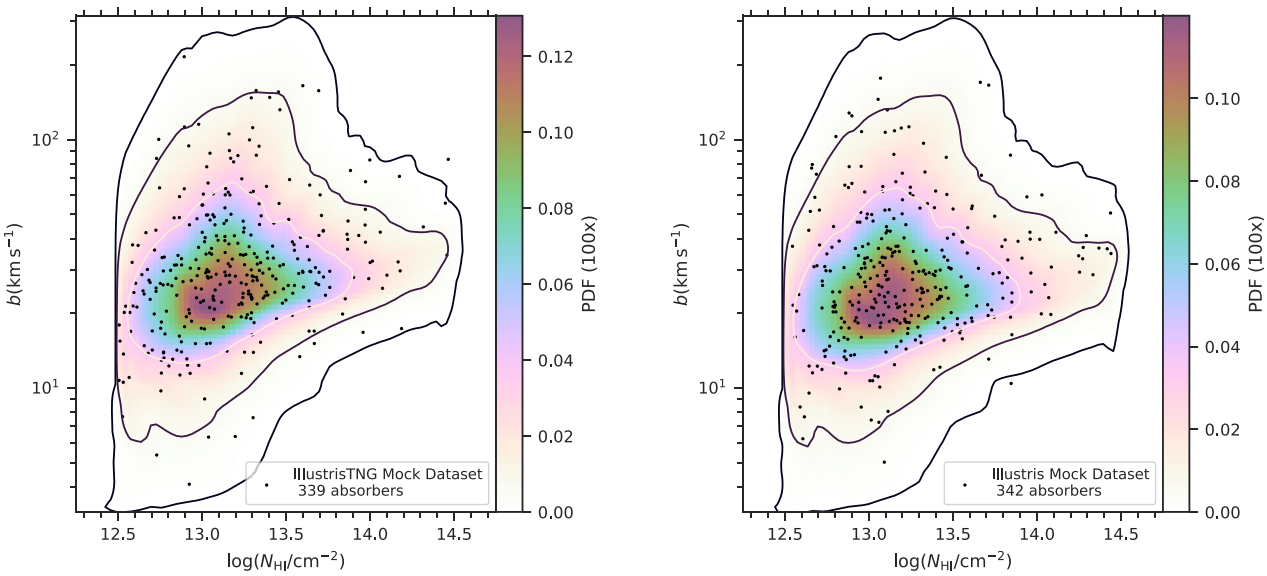

Figure 11. Joint b–$N_{\rm H\,I}$ distributions recovered from the inference results for IllustrisTNG (left) and Illustris (right) simulations, emulated by our DELFI emulator based on the median values of the marginalized MCMC posterior. Black dots are the mock data sets we used in the inference. The contours correspond to cumulative probabilities of 68, 95, and 99.7 per cent. For illustration purposes, the values of the pdf are multiplied by 100 in the colour bar.

For IllustrisTNG, the offsets between the $[T_0, \gamma]_{\rm inf}$ and $[T_0, \gamma]_{\rm fit}$ are smaller, with $\Delta\log(T_0/{\rm K}) = -0.022$ dex, $\Delta\gamma = 0.064$. We notice that these offsets are smaller than the typical inference precision obtained based on realistic data sets, as shown in Figs 7 and 10, which report the marginalized 1D 1σ error in $\log T_0$, $\sigma_{\log T_0}$, $\sim$0.1 dex and the marginalized 1D 1σ error in γ, σ_γ, $\sim$0.1. For both simulations, we observe the offsets $\Delta\log T_0 \lesssim 0.5\sigma_{\log T_0}$, and $\Delta\gamma \lesssim \sigma_\gamma$.

To check the robustness of these results, we use the IllustrisTNG and Illustris simulations as mock observational data and perform inference tests using two different sets of 'true parameters': the $[T_0, \gamma]_{\rm fit}$ obtained from our power-law fits the $\Delta - T$ distribution of the simulations (see Fig. 2), and the $[T_0, \gamma]_{\rm inf}$ given by our inference method when applied to an extremely large mock data set, as described above. Given that the inferred $\Gamma_{\rm H\,I}$ for both IllustrisTNG and Illustris simulations consistently deviates from

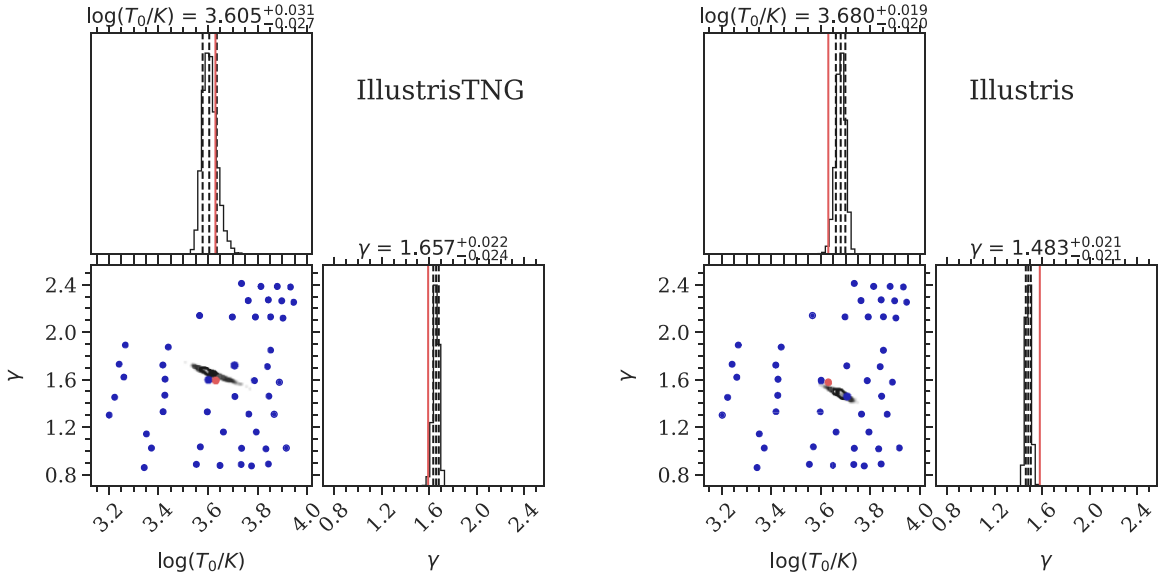

Figure 12. Corner plots for IllustrisTNG (left) and Illustris (right), based on the larger mock data set, with $\Delta z = 42.72$, corresponding 20 times the observational data set. Projections of the thermal grid used for generating models are shown as blue dots. The inner (outer) black contour represents the projected 2D 1(2)σ interval. The true parameters for the simulations, obtained by fitting the T–Δ distributions of the simulations, are indicated by the red dot (lines) in the (marginal) distributions, while the dashed black lines indicate the 16, 50, and 84 percentile values of the posterior.

the actual values, owing to the previously mentioned degeneracy between the photoheating rate and feedback strength, any inference tests incorporating the $\Gamma_{\rm H\,I}$ from these two simulations will surely fail. To this end, we focus on the inference results on the T_0–γ plane and conduct marginalized inference tests by marginalizing the posteriors over the $\Gamma_{\rm H\,I}$, in which 2D marginalized contours levels are modelled by Gaussian mixture models. For each simulation, we run 100 realizations on each set of 'true parameters', and run inference tests on the obtained posteriors. The results are shown in Fig. 13, indicating that our inference method is overconfident for both sets of 'true parameters'. While the inference method is not able to recover the thermal state $[T_0, \gamma]_{\rm fit}$, the thermal state $[T_0, \gamma]_{\rm inf}$ significantly improves the outcome of the inference test. These results suggest that our inference method is able to robustly recover the $[T_0, \gamma]$ with small biases, for simulations that include feedback mechanisms.

The inference tests imply that there exist offsets for the inferred parameters $[T_0, \gamma]_{\rm inf}$ for IllustrisTNG and Illustris, where $\Delta\log(T_0/{\rm K}) = -0.022$ dex, $\Delta\gamma = 0.064$ for IllustrisTNG and $\Delta\log(T_0/{\rm K}) = 0.047$ dex, $\Delta\gamma = -0.094$ for Illustris. However, these offset are insignificant, with $\Delta\log T_0 \lesssim 0.5\sigma_{\log T_0}$, and $\Delta\gamma \lesssim \sigma_\gamma$. However, it is unclear whether the observed differences between $[T_0, \gamma]_{\rm inf}$ and $[T_0, \gamma]_{\rm fit}$ can be attributable to the intrinsic difference between the Nyx, IllustrisTNG, and Illustris simulations, or if they arise from potential degeneracy between the IGM thermal state $[T_0, \gamma]$ and the feedback mechanism implemented in the simulation. Nevertheless, the latter hypothesis seems to contrast with the results based on the various statistics of the low-z Ly α forest presented in Khaire et al. (2024), which suggests that the impacts from different feedback models are not distinguishable via the Ly α forest under realistic scenarios, i.e. forward-modelled using the D16 COS data set. The only exception is the case of the Lyα flux power spectrum at small scales, where minor deviations are observed in both simulations(see Khaire et al. 2023, for the dN/dz around massive haloes). To further explore this problem, we examine the physical properties of low -z Ly α absorbers in the following section.

4 LOW-z LY α FORESTS AND SIMULATED ABSORBERS

4.1 Identifying the simulated Ly α absorbers

To understand whether the low-z Ly α forest effectively probes the WHIM, we attempt to identify the simulated Ly α absorbers, i.e. the $n_{\rm H\,I}$ peaks in the simulation skewers, that give rise to the Ly α lines detected in the mock spectra. This approach allows us to directly examine the physical properties (T, Δ, and $n_{\rm H\,I}$) of these simulated Ly α absorbers and draw a direct correspondence between them and the line parameters ($\{b, N_{\rm H\,I}\}$) of their corresponding Ly α lines detected in the mock spectra. In this section, we chose to focus on the simulated Ly α absorbers in the Nyx simulation at $z = 0.1$ (with default thermal history, i.e. $T_0 = 3.612$, and $\gamma = 1.588$ at $z = 0.1$). For clarification, within the context of this study, the 'terms simulated Ly α absorbers' or 'simply simulated absorbers' are used to denote the $n_{\rm H\,I}$ peaks that give rise to the Ly α absorption lines in the mock spectra detected by `VPFIT`.

Our approach for identifying simulated Ly α absorbers works as follows. First, we include the physical properties, including temperature T, overdensity Δ, velocity along LOS $v_{\rm los}$, and the neutral fraction $x_{\rm H\,I}$ in our skewers and stitch them in the forward-modelling procedure (see Section 2.4). We interpolate the stitched skewer on to the forward-modelled wavelength grid, and calculated the neutral hydrogen density $n_{\rm H\,I}$ for each simulation cell, based on the neutral fraction $x_{\rm H\,I}$, overdensity Δ, and the mean hydrogen density $\bar{n}_{\rm H}$. Subsequently, we scan the stitched skewers (in real space) for $n_{\rm H\,I}$ peaks, and classify these with $n_{\rm H\,I} > 10^{-12}$ cm^{-3} as potential simulated Ly α absorbers. The minimal peak H I density is derived from both the minimal H I column density for the detected lines $N_{\rm H\,I,min} = 10^{12.5}{\rm cm}^{-2}$ (see Section 2.4) and the maximal length for simulated absorbers $l_{\rm abs,\,max} = 0.5$ Mpc h^{-1}, which is consistent with previous studies that attempt to characterize the structures giving rise to the Ly α forest at $z = 0.1$ (Bolton et al. 2022b; Tillman et al. 2023b).

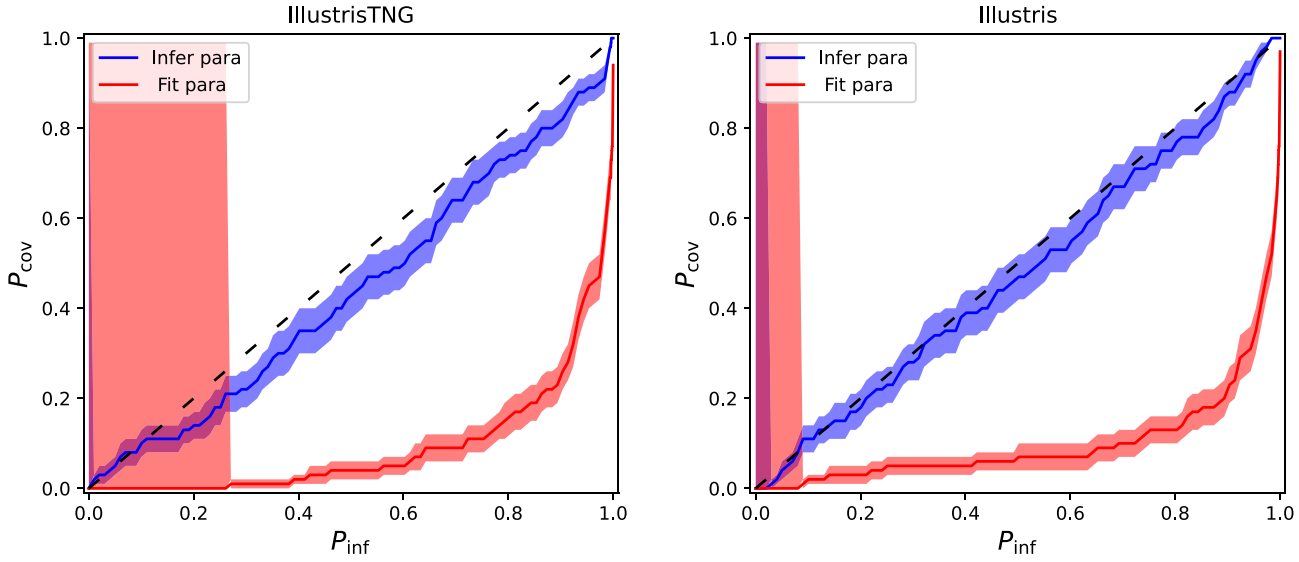

Figure 13. Marginalized coverage probability $P_{\rm inf}$ for inference tests using IllustrisTNG (left) and Illustris (right) simulations as mock observational data. The x-axis stands for the inferred probability $P_{\rm inf}$, and the y-axis shows the probability for the parameters of the true model to fall in the contour corresponding to $P_{\rm cov}$. The shaded regions indicate the $1-\sigma$ error for $P_{\rm cov}$. Inference tests with the true parameters set by $[T_0, \gamma]_{\rm inf}$ are shown in blue, while inference tests with the true parameters set by $[T_0, \gamma]_{\rm fit}$ are shown in red.

Given these two parameters, the requisite minimum H I peak density for simulated absorbers to yield observable Ly α absorption lines is computed as $n_{\rm H\,I,\,min} = \frac{N_{\rm H\,I,min}}{l_{\rm abs,\,max}} = \frac{10^{12.5}{\rm cm}^{-2}}{0.5{\rm Mpc/h}} \sim 10^{-12}$ cm^{-3}, which effectively filters out $n_{\rm H\,I}$ peaks that give rise to Ly α absorption lines below our sensitivity. We then determine the physical size for each potential simulated absorber along the LOS, $l_{\rm abs}$, using a threshold at which $n_{\rm H\,I}$ drops below 1 per cent of its peak value, while restricting the maximal size to be $l_{\rm abs,\,max} = 0.5$ Mpc/h. We calculate the H I column densities of the simulated Ly α absorbers, $N_{\rm H\,I,sim}$, by integrating the $n_{\rm H\,I}$ over the ranges set by aforementioned threshold. We observed that the resulting $N_{\rm H\,I}$ is not particularly sensitive to $l_{\rm abs}$, because the $n_{\rm H\,I}$ peak is so narrow that the majority of the neutral hydrogen comes from the peak region (see Fig. 14).

After identifying the potential Ly α absorbers, we extract their LOS velocity from the simulation cells, and compute the central wavelength of the expected absorption lines in redshift space, accounting for the redshift caused by its LOS velocity. For each anticipated absorption line originating from an $n_{\rm H\,I}$ peak, we check whether its central wavelength lies within $\pm$ 50 km s^{-1} of the central wavelength of any Ly α lines detected in the mock spectrum. If so, we confirm the identification of a simulated Ly α absorber, and take the T and Δ at the $n_{\rm H\,I}$ peak as its values, which is valid since the $n_{\rm H\,I}$ peak is so narrow that the majority of the $N_{\rm H\,I}$ comes from the region close to the peak. While theoretically, the Ly α lines are expected to be caused by multiple $n_{\rm H\,I}$ peaks in real space (Garzilli, Theuns & Schaye 2015), we discover that at $z = 0.1$, each Ly α line detected in the mock spectra with $12.5 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 14.5$ predominantly originates from one single $n_{\rm H\,I}$ peak in the simulation. It is not a resolution effect, since the simulation (Nyx) has a grid length ~ 0.02 Mpc h^{-1}, while the simulated absorbers have sizes ~ 0.5 Mpc h^{-1}. As such, we only consider the $n_{\rm H\,I}$ peak with the highest $n_{\rm H\,I}$ value if multiple $n_{\rm H\,I}$ peaks contribute to the same detected absorption line.

Fig. 14 showcases examples of the simulated Ly α absorbers, alongside their corresponding absorption lines in the mock spectra and the related simulation skewers. The top panel depicts the flux (black), noise vector (green), and the model fitted by VPFIT (blue). The central wavelength of Ly α lines identified by VPFIT are indicated by orange vertical lines, and the corresponding simulated absorbers are indicated by orange dashed lines in the second panel (and all other panels below). The log $N_{\rm H\,I,fit}$, $b_{\rm fit}$ reported by VPFIT are given in the annotation, together with the log $N_{\rm H\,I,sim}$, $b_{\rm thermal}$ calculated based on the simulation, whereas the $b_{\rm thermal} = (2kT/m_{\rm H\,I})^{1/2}$ is the thermal component of the b-parameters computed based on the T of the simulated Ly α absorbers (see equation 5). The second panel depicts the neutral hydrogen density $n_{\rm H\,I}$, while the shaded regions represent the identified Ly α absorbers along LOS, which are used as the integral ranges while computing the $N_{\rm H\,I,sim}$. The orange vertical dashed lines show the $n_{\rm H\,I}$ peaks of the confirmed simulated Ly α absorbers, while the purple vertical dashed lines show the potential simulated Ly α absorbers that do not cause detectable Ly α lines. The grey horizontal dashed line represents the minimal H I peak density, $n_{\rm H\,I,min} = 10^{-12}$ cm^{-3}. The third, fourth and fifth panels show the overdensity Δ, temperature log T, and LOS velocity $v_{\rm los}$ (black solid lines). The brown horizontal dashed line in the fourth panel stands for $T = 10^5$ K, which divides the cool diffuse Ly α gas and the WHIM. The left-hand panel shows a simulated Ly α absorber in the diffuse Ly α phase, while the right left-hand panel shows a simulated Ly α absorber arising from the WHIM phase.

We perform the identification procedure for all 1000 mock spectra, discovering 34 011 potential simulated Ly α absorbers,i.e. $n_{\rm H\,I}$ peaks, among which 10 510 are identified as simulated Ly α absorbers and matched to their respective absorption lines identified by VPFIT. The discrepancy between potential and confirmed Ly α absorbers is due to the inclusion of minor $n_{\rm H\,I}$ peaks, that are too weak to cause any detectable Ly α line, which is indicated by purple vertical lines in the left panel of Fig. 14. Lastly, approximately 2 per cent of the lines detected by VPFIT could not be matched to any simulated Ly α absorber. These anomalies could potentially result from false identification of the VPFIT induced by noise. None the less, given the rarity of these cases, omitting them should not influence our statistical results or conclusions.

To validate our identification method, we compare the observed line parameters, reported by VPFIT, with the values calculated from

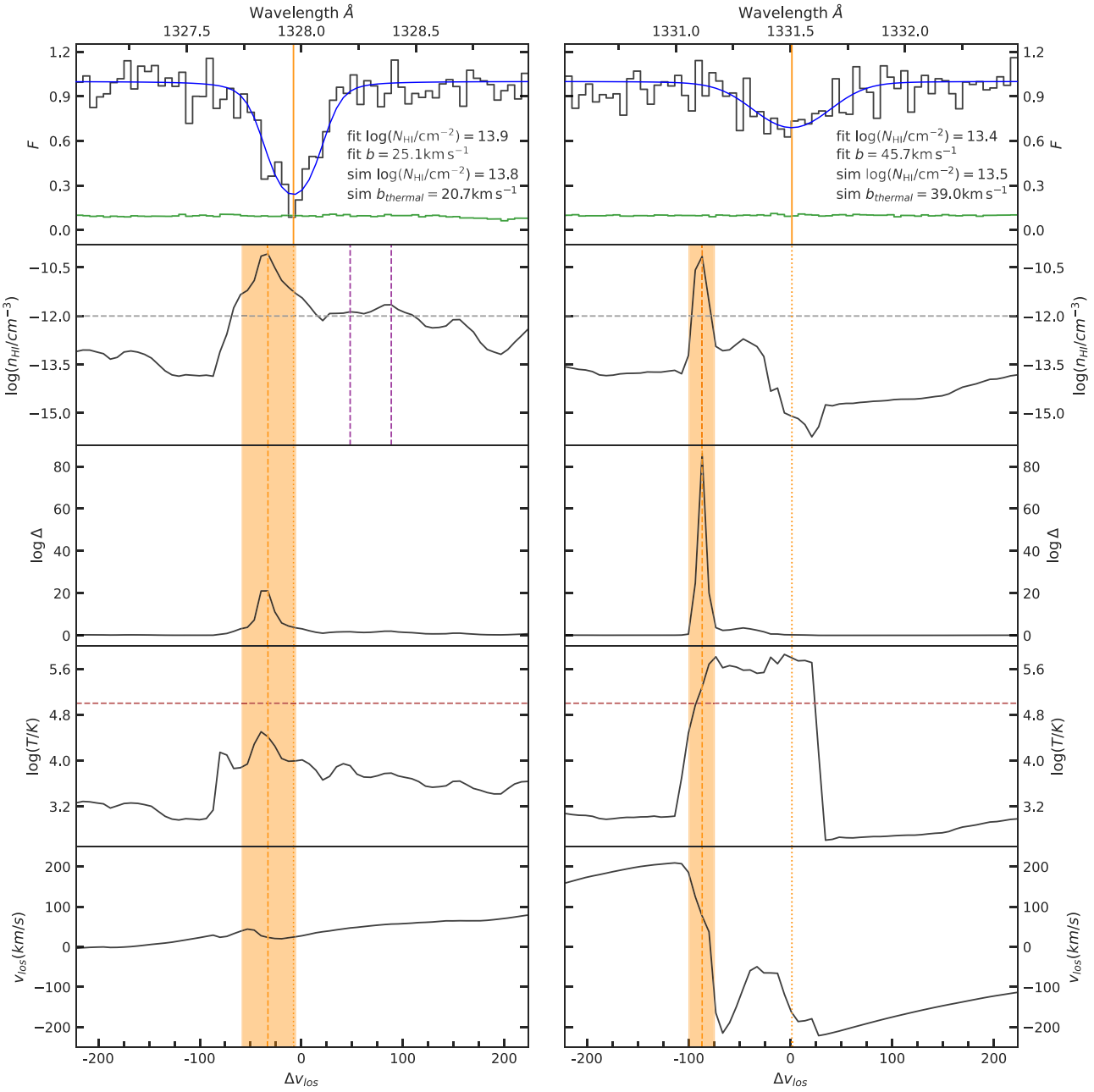

Figure 14. Illustration of a segment of one of the forward-modelled mock spectra (top panel) with the absorption lines detected by VPFIT and the corresponding skewer. The top panel depicts the flux (black), noise vector (green), and the model fitted by VPFIT (blue). The central wavelength of Ly α lines identified by VPFIT are indicated by orange vertical lines, and the corresponding simulated absorbers are indicated by orange dashed lines in the second panel (and all other panels below). The log $N_{\rm H\,I,fit}$, $b_{\rm fit}$ reported by VPFIT are given in the annotation, together with the log $N_{\rm H\,I,sim}$, $b_{\rm thermal}$ calculated based on the simulation. The second panel depicts the neutral hydrogen density $n_{\rm H\,I}$, while the shaded regions represent the identified Ly α absorbers along LOS, which are used as the integral ranges while computing the $N_{\rm H\,I,sim}$. The orange vertical dashed lines show the $n_{\rm H\,I}$ peaks of the confirmed simulated Ly α absorbers, while the purple vertical dashed lines show the potential simulated Ly α absorbers that do not cause detectable Ly α lines. The grey horizontal dashed line represents the minimal H I peak density, $n_{\rm H\,I,min} = 10^{-12}$ cm^{-3}. The third, fourth and fifth panels show the overdensity Δ, temperature T, and LOS velocity $v_{\rm los}$. The brown horizontal dashed line in the fourth panel stands for $T = 10^5$ K. left: A Ly α absorbers in the diffuse Ly α phase. right: A Ly α absorbers in the WHIM phase.

the simulation. In Fig. 15, we showcase the $N_{\rm H\,I,fit}$ (left) and $b_{\rm fit}$ (right) for all Ly α lines fitted by VPFIT, compared with the $N_{\rm H\,I,sim}$ and $b_{\rm thermal}$, respectively, both calculated from the corresponding simulated Ly α absorbers identified in the Nyx simulation. The left-hand panel indicates a strong correlation between the fitted $N_{\rm H\,I,fit}$ and the $N_{\rm H\,I,sim}$ calculated from the simulation, implying that the $n_{\rm H\,I}$ peaks identified by our method are indeed the simulated Ly α absorbers responsible for the Ly α lines detected in the mock spectra. The right panel demonstrates that the bulk of $b_{\rm fit}$ lies above the dashed line representing $b_{\rm fit} = b_{\rm thermal}$. This result aligns with the nature of the b-parameter, as given by

$$b = \sqrt{b_{\rm thermal}^2 + b_{\rm notherm}^2}, \tag{5}$$

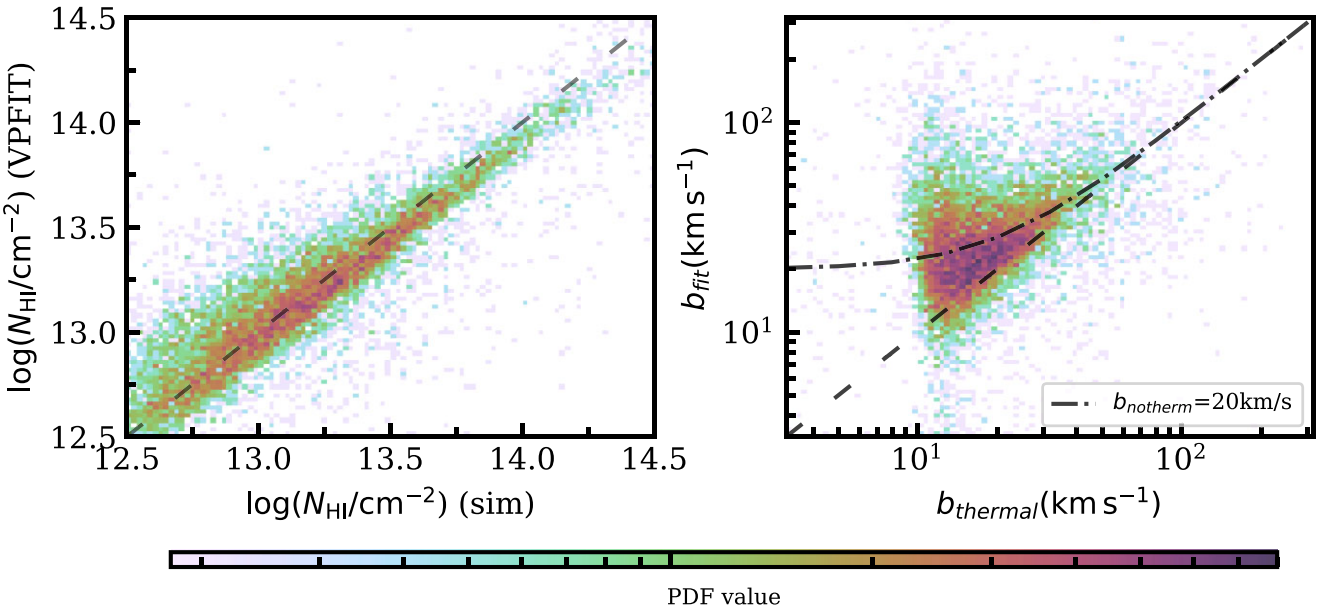

Figure 15. The observed variables $N_{\rm H\,I,fit}$, $b_{\rm fit}$ fitted by VPFIT compared with the physical quantities $N_{\rm H\,I,sim}$ and $b_{\rm thermal}$ of the simulated absorbers identified in the simulation skewers, where the $N_{\rm H\,I,sim}$ is calculated by integrating the $n_{\rm H\,I}$ of the absorbers along the LOS, and the $b_{\rm thermal}$ is computed by assuming the broadening of the Ly α lines are pure thermal. Left: $N_{\rm H\,Ifit}$ versus $N_{\rm H\,I,sim}$. Right: $b_{\rm fit}$ versus $b_{\rm thermal}$. The dashed–dotted line in the right-hand panel represents the b-parameter resulting from the combination of the thermal component $b_{\rm thermal}$ and a turbulence in the IGM with $b_{\rm notherm} = 20$ km s^{-1}.

where the $b_{\rm notherm}$ is the non-thermal component of the b-parameter resulting from combinations of Hubble flow, peculiar velocities and turbulence in the IGM. Equation (5) demonstrates that the $b_{\rm thermal}$ is the lower limit of the b-parameter, which corresponds to the lower right cutoff of the b–$N_{\rm H\,I}$ distribution (see the colour maps in Figs 8 and 11 as examples. More discussions on this topic can be found in Schaye et al. 1999, Rudie, Steidel & Pettini 2012, Bolton et al. 2014, and Hu22). Furthermore, the right-hand panel of Fig. 15 gives a rough correlation between the $b_{\rm fit}$ and $b_{\rm thermal}$ and provides an approximate estimation of the strength of the non-thermal broadening of the Ly α lines at $z = 0.1$. It suggests that for the Nyx simulation, the non-thermal contribution to the b-parameter can be modelled by a 'turbulent' motion in the IGM with $b_{\rm notherm} \sim 20$ km s^{-1} (indicated by the black dashed–dotted line in Fig. 15).

We summarize the (Δ, T) for the ensemble of simulated Ly α absorbers identified in the Nyx simulation in Fig. 16. Considering that we have established one-to-one correspondence between the simulated absorbers and observed (mock) absorption lines, we employ a consistent filter to both sets, which selects Ly α lines with, $12.5 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 14.5$ and $0.5 \leq \log(b/{\rm km\,s}^{-1}) \leq 2.5$ (see Section 2.4). In the upper panels, we plot the volume-weighted 1D marginal distributions of Δ and T for all simulation grid cells, juxtaposed with the 1D distributions of Δ and T for the simulated Ly α absorbers, showing that the simulated Ly α absorbers, in general, have higher temperature and overdensity, compared with the full simulation. The peaks of the Δ and T distributions of the simulated Ly α absorber highlight the specific range of Δ and T to which the Ly α forest is sensitive at $z = 0.1$. More specifically, the Ly α forest is most sensitive to the IGM characterized by $\log\Delta = 0.92$ and $T = 10^{4.27}$ K. It is worth mentioning that, as mentioned in Section 2.3, the Ly α optical depth $\tau_{{\rm Ly}\,\alpha}$ is dependent on $\Gamma_{\rm H\,I}$. Consequently, the regions to which the Ly α forest is sensitive also depend on $\Gamma_{\rm H\,I}$. This point will be fully discussed later in Section 4.3.

The bottom left panel of Fig. 16 shows the (volume-weighted) T–Δ distributions for simulated Ly α absorbers (left), and all grid cells in the simulation (right), while the volume-weighted gas fractions[9] are given in annotations for simulated absorbers and the whole simulation in the left-hand and the right-hand panel, respectively. The black contours in both panels illustrate the 1σ and 3σ (68 and 99.7 per cent) contours for the T–Δ distribution of the simulated Ly α absorbers. The T–Δ distribution of the simulated Ly α absorbers appears to be scattered at low-z, extending into the WHIM phase, due to the pervasive effects of shock heating. As per the gas phase fractions of the Ly α absorbers shown in the bottom left panel, approximately 7 per cent of the absorbers originate from the WHIM phase, suggesting that the low-z Ly α forest does probe the WHIM (see the right panel of Fig. 14 as an example), although its sensitivity is notably limited given the small fraction of lines arising from this phase. Such a result aligns with Tepper-García et al. (2012) regarding the detectability of the Broad Ly α Absorbers (BLAs) at low-z under realistic conditions.

4.2 Simulated Ly α absorbers in IllustrisTNG and Illustirs

To further study the effects of the feedback mechanisms on the Ly α forest at $z = 0.1$, we identify the simulated Ly α absorbers in both the IllustrisTNG and Illustris simulations, and pair them to the corresponding absorption lines present in the mock spectra, following the method outlined in Section 4.1. For each simulation, we carry out the identification process across 1000 mock spectra and

[9] As previously mentioned, for each simulated Ly α absorber, we use the T and Δ at its $n_{\rm H\,I}$ peak, which dominates the Ly α absorption. To this end, when calculating the volume-weighted gas fractions, we do not take the physical size into account, but instead, only consider the one simulation cell where the $n_{\rm H\,I}$ reaches its maximum. This is reasonable since typical $n_{\rm H\,I}$ peaks seen in this study are so narrow that most of the $N_{\rm H\,I}$ comes from the peak cell. As a result, the so-called volume-weighted gas fractions for simulated absorbers are effectively unweighted. This approach is used for all gas fractions related to simulated Ly α absorbers throughout this paper.

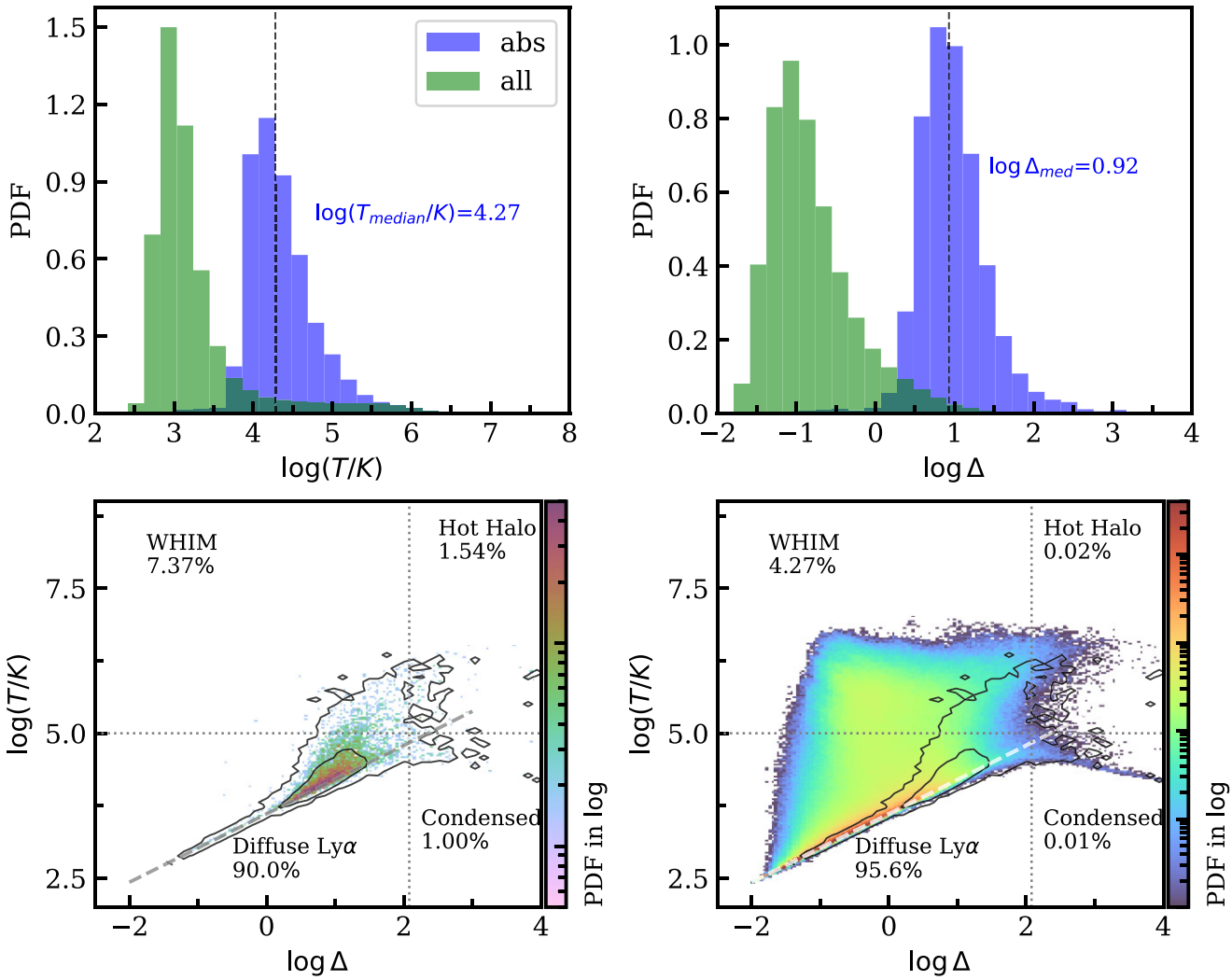

Figure 16. Distributions of Δ and T of the simulated Ly α absorbers in the Nyx simulation, compared with the full simulation. The ensemble consists of 10 510 absorbers, all obtained from the 1000 spectra discussed in Section 4.1. The top panels show the 1D distributions of T(right) and Δ(left) for the whole simulation (green) compared with simulated Ly α absorbers (blue). The medians of the T and Δ for the simulated absorbers are indicated by dashed black lines. The bottom panels plot the 2d T–Δ distributions for the Ly α absorbers(left) and for the whole simulation(right), while the contours for 1σ, 3σ (68 and 99.7 per cent) of the T–Δ distribution of the absorbers are shown in both panels. The volume-weighted gas phases for absorbers and the whole simulation are given in the left-hand panel and the right-hand panel, respectively. The best-fitting power-law T–Δ relationships are given in the bottom panels as comparisons.

summarize the physical properties of the simulated absorbers. It is worth mentioning that here the IllustrisTNG and Illustris simulations are tuned to have identical dN/dz, which requires different $\Gamma_{\rm H\,I}$ values (see Section 2.5).

We plot the marginalized 1D distributions of the Δ and T for both IllustrisTNG (top) and Illustris (bottom) in Fig. 17. The plots show that the overall distributions of T and Δ for the two simulations are evidently different due to their different feedback recipes. For instance, the extreme feedback in Illustris simulation results in much more WHIM compared with IllustrisTNG, causing a secondary peak in its T distribution. However, the distributions of T and Δ for the absorbers in both simulations are comparable, with $\log(T_{\rm med}/{\rm K})$ =4.33, $\log\Delta_{\rm med}$ = 0.97 for IllustrisTNG, and $\log(T_{\rm med}/{\rm K})$ = 4.16, $\log\Delta_{\rm med}$ = 0.72 for Illustris. Moreover, we discover that the differences in $\log(T_{\rm med}/{\rm K})$ and $\log\Delta_{\rm med}$ for the three simulations are actually caused by the different $\Gamma_{\rm H\,I}$ values used for the three simulations. The relevant discussion is presented in Section 4.3.

The (volume-weighted) 2D T–Δ distributions for simulated Ly α absorbers in both IllustrisTNG (top) and Illustris (bottom) simulations are shown in the left column of Fig. 18, whereas the (volume-weighted) 2D T–Δ distributions for the whole simulations are given in the right column as comparisons. The volume-weighted gas fractions are given in the annotation for simulated absorbers and full simulation in the left-hand and the right-hand panels, respectively. For the simulated Ly α absorbers, 12.2 per cent (10.7 per cent) of the Ly α absorbers arise from the WHIM for IllustrisTNG (Illustris), while the value for Nyx simulation is approximately 7 per cent. The 1σ and 3σ (68 and 99.7 per cent) contours for the T–Δ distributions for the simulated Ly α absorbers are also given in the Figure, showing that their T-Δ distributions are more scattered compared with these in Nyx simulation, especially for the WHIM phase absorbers. These differences are caused by stronger shock heating in IllustrisTNG and Illustris simulations compared with Nyx simulation, caused by their feedback mechanisms. However, while the (volume-weighted) WHIM fractions for the two simulations are remarkably different,

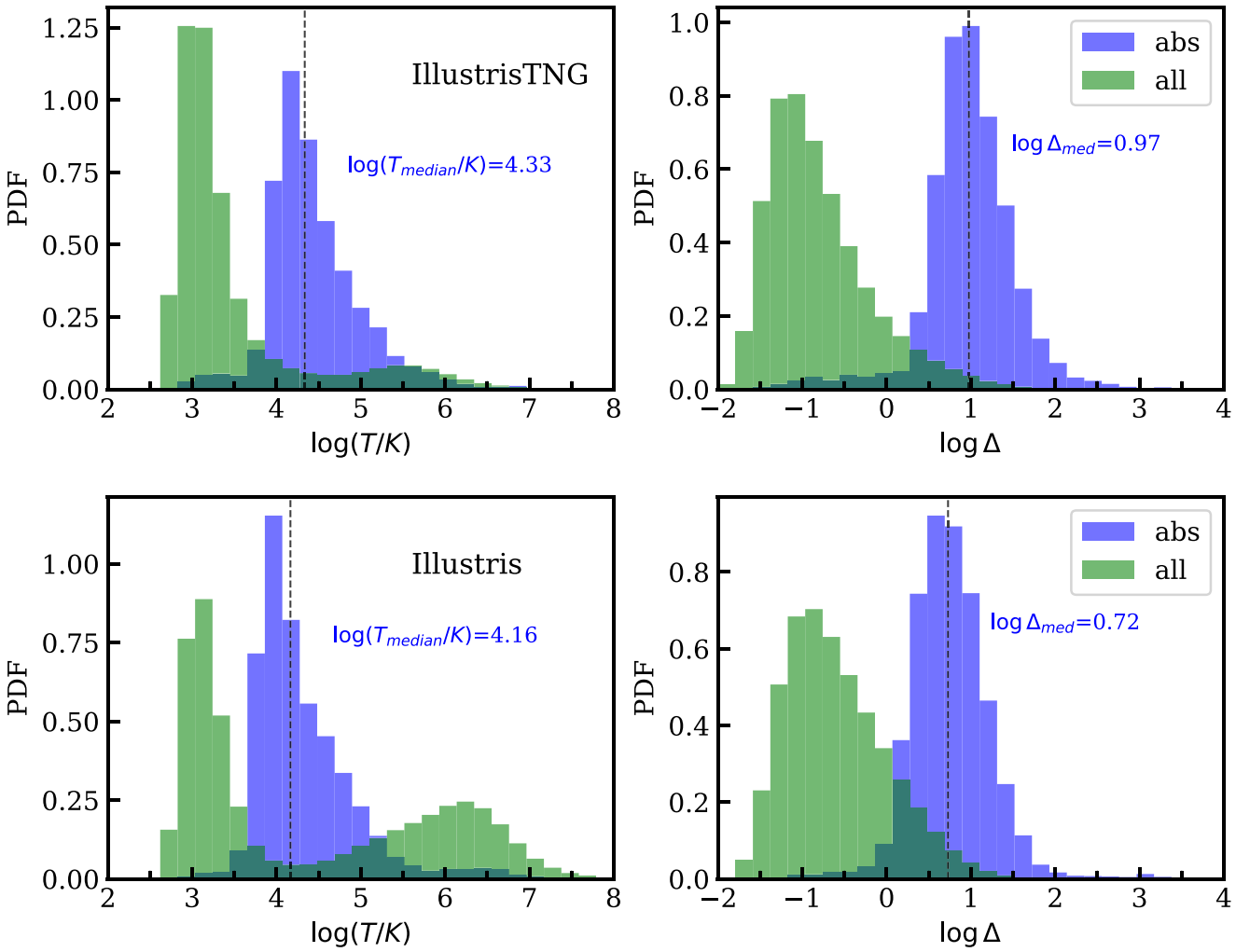

Figure 17. Marginalized 1D Δ and T distributions of the simulated Ly α absorbers in the IllustrisTNG (top) and Illustris (bottom) simulation. The medians of the T and Δ for the simulated absorbers are indicated by dashed black lines. The overall Δ and T distributions of the full simulations are plotted as comparisons. The two simulations share the same dN/dz, which is the same value observed in the D16 data set.

9.8 per cent for IllustrisTNG and 38.0 per cent for Illustris, the WHIM fractions for the Ly α absorbers are similar, both around 11 per cent. Furthermore, in Section 4.3 we demonstrate that the small difference in WHIM fractions for simulated absorbers in the two simulations actually arises from the different $\Gamma_{\rm H\,I}$ values used in the two simulations. Such a fact implies that the low-z Ly α forest does not probe the WHIM effectively under realistic conditions, which is consistent with the conclusion drawn by Khaire et al. (2024).

4.3 Simulations under the same $\Gamma_{\rm H\,I}$

Considering that the calculation of the Ly α optical depth $\tau_{\rm Ly\alpha}$ involves $\Gamma_{\rm H\,I}$, and given that the observed absorption feature (i.e. the Ly α forest) consistently probes regions with $\tau_{\rm Ly\,\alpha} \sim 1$, it follows that the T and Δ of these regions probed by the Ly α forest, are influenced by the $\Gamma_{\rm H\,I}$ values. Such an argument can be qualitatively demonstrated by the fluctuating Gunn–Peterson approximation (FGPA; see Weinberg et al. 1997)

$$\tau_{\rm Ly\alpha} \propto x_{\rm HI} n_{\rm H} \propto \frac{n_{\rm H}^2 T^{-0.7}}{\Gamma_{\rm HI}} \propto \frac{\Delta^{2.7-\gamma}}{\Gamma_{\rm HI}} \propto \frac{T^{2/(\gamma-1)-0.7}}{\Gamma_{\rm HI}}, \quad (6)$$

where the $\tau_{\rm Ly\,\alpha}$ denotes the Ly α optical depth and the $n_{\rm H}$ is the hydrogen number density. Since the Ly α forest always probes the region with $\tau_{\rm Ly\,\alpha} \sim 1$, the last two terms in equation (6) suggest that the $\Gamma_{\rm H\,I}$ is in positive correlation with Δ and T, respectively, given the $\gamma \sim 1.6$ at $z = 0.1$.

In our analysis, the three simulations are tuned to match dN/dz. However, due to the degeneracy between $\Gamma_{\rm H\,I}$ and feedback mechanisms, each simulation ends up with a distinct $\Gamma_{\rm H\,I}$ value (refer to Section 2.5). As a result, the T and Δ distributions of the simulated Ly α absorbers in these simulations are influenced not just by the feedback but also by the varying $\Gamma_{\rm H\,I}$ values. To isolate and examine solely the impact of feedback, we post-process the IllustrisTNG and Illustris simulations to align with the $\Gamma_{\rm H\,I}$ value used in Nyx, set at $\log(\Gamma_{\rm H\,I}/s^{-1}) = -13.093$. With this consistent $\Gamma_{\rm H\,I}$ across the three simulations, we re-perform the analysis from the prior section and summarize the results below. It is worth mentioning that the overall T–Δ distributions of simulations are determined by the cooling and heating processes during their evolution and are not altered by the post-processing of the $\Gamma_{\rm H\,I}$.

We plot the marginalized Δ and T distributions and their median values for Ly α absorbers in Nyx, IllustrisTNG, and Illustris simulations with the same dN/dz in Fig. 19. Interestingly, for simulations

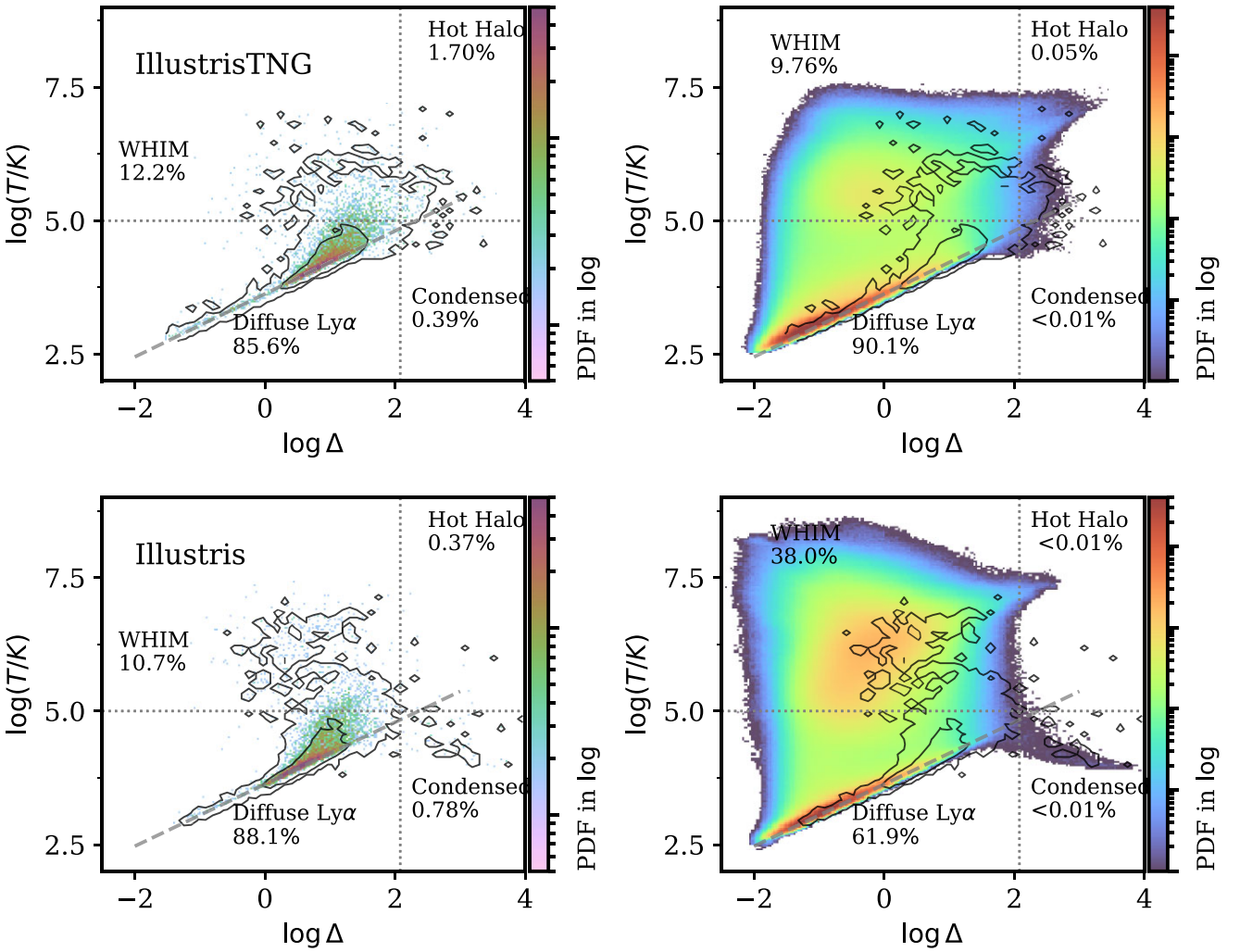

Figure 18. T–Δ distributions of the Ly α absorbers in the IllustrisTNG (top) and Illustirs (bottom) simulations, compared with the T–Δ distributions of full simulations (right). The contours for 1σ, 3σ (68 and 99.7 per cent) of the T–Δ distribution of the absorbers are shown in both panels. The volume-weighted gas phases for absorbers and the whole simulation are given in the left-hand panel and the right-hand panel, respectively. The best-fitting power-law T–Δ relationships are given as comparisons. The two simulations are dN/dz matched.

with the same dN/dz, the T and Δ for absorbers are correlated with its $\Gamma_{\rm H\,I}$. More specifically, with $\Gamma_{\rm H\,I,Illustris} < \Gamma_{\rm HI,Nyx} < \Gamma_{\rm H\,I,IllustrisTNG}$ (see Fig. 6), we obtain $T_{\rm med,\,Illustris} < T_{\rm med,\,Nyx} < T_{\rm med,\,IllustrisTNG}$ and $\Delta_{\rm med,\,Illustris} < \Delta_{\rm med,\,Nyx} < \Delta_{\rm med,\,IllustrisTNG}$.

We plot the marginalized Δ and T distributions and their median values for Ly α absorbers in the three simulations under the same $\Gamma_{\rm H\,I}$ in Fig. 20. Under the same $\Gamma_{\rm H\,I}$, the T and Δ distributions for simulated absorbers in all three simulations become almost identical, having nearly the same median values for T and Δ, respectively. Such a result suggests that while feedback evidently affects the overall T–Δ distributions of the low-z IGM (see Fig. 2), their impacts on the physical properties of the low-z Ly α forest (i.e. the T and Δ distributions) are not distinguishable under realistic conditions.

In Fig. 21, we plot the 2D T–Δ distributions of the simulated Ly α absorbers in IllustrisTNG (left) and Illustris (right), under the same $\Gamma_{\rm H\,I}$. While the overall IGM T–Δ distributions for the two simulations are evidently different (see Fig. 2), the T–Δ distributions of the simulated Ly α absorbers in these two simulations are similar, and the gas phase fractions for absorbers in both simulations are almost identical, suggesting that the small difference in the WHIM fractions of the simulated absorbers shown in Fig. 18 are caused by different $\Gamma_{\rm H\,I}$ values. Such results indicate that the $\Gamma_{\rm H\,I}$ has a much stronger impact on the Ly α forest compared with the feedback mechanisms implemented in IllustrisTNG and Illustris simulations.

5 SUMMARY AND DISCUSSION

In this paper, we explore the effects of the WHIM, which causes significant dispersion in the IGM T–Δ distribution, on the low-z Ly α forest and the IGM thermal state $[T_0, \gamma]$ measured from it. We first evaluate the effectiveness of $[T_0, \gamma]$ as IGM parameters under the inference framework presented in Hu22, and compare its performance with the photoheating labels $[A, B]$. We discover that the thermal state $[T_0, \gamma]$ still parametrizes the IGM effectively in spite of the dispersion in the IGM T–Δ distribution. We further apply the inference method to IllustrisTNG and Illustris simulations which implement different variants of feedback, potentially making them

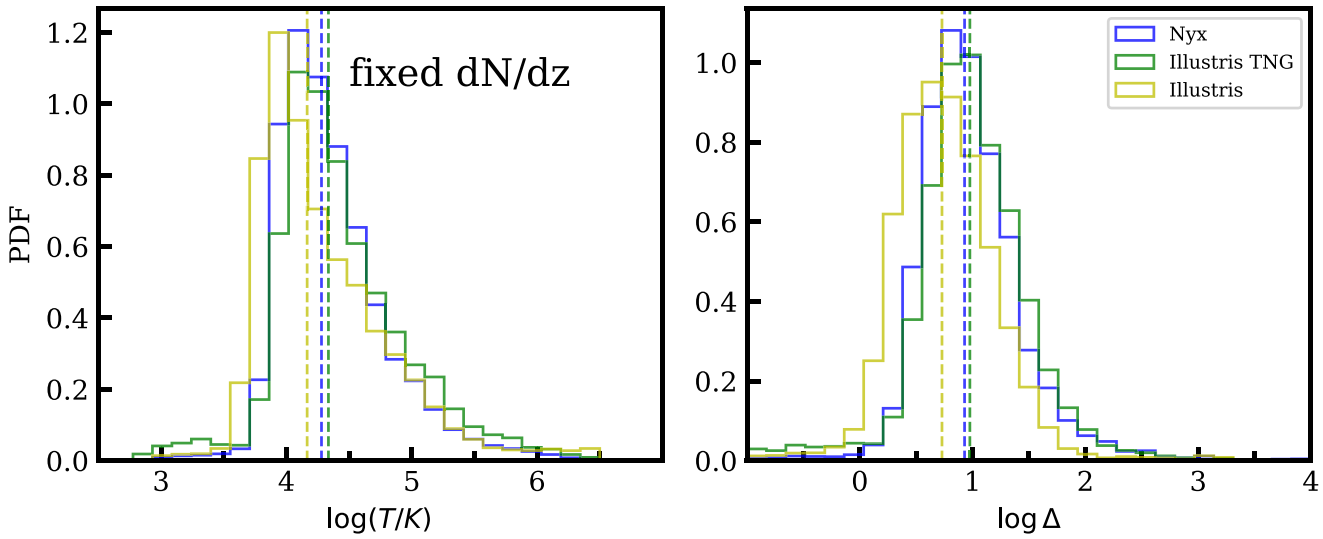

Figure 19. Marginalized 1d distribution of T (left), and Δ (right) of the simulated Ly α absorbers identified in Nyx (blue), IllustrisTNG (green), and Illustris (yellow) simulations. The medians of log T and log Δ are indicated by vertical dashed lines. The three simulations are turned to have identical dN/dz.

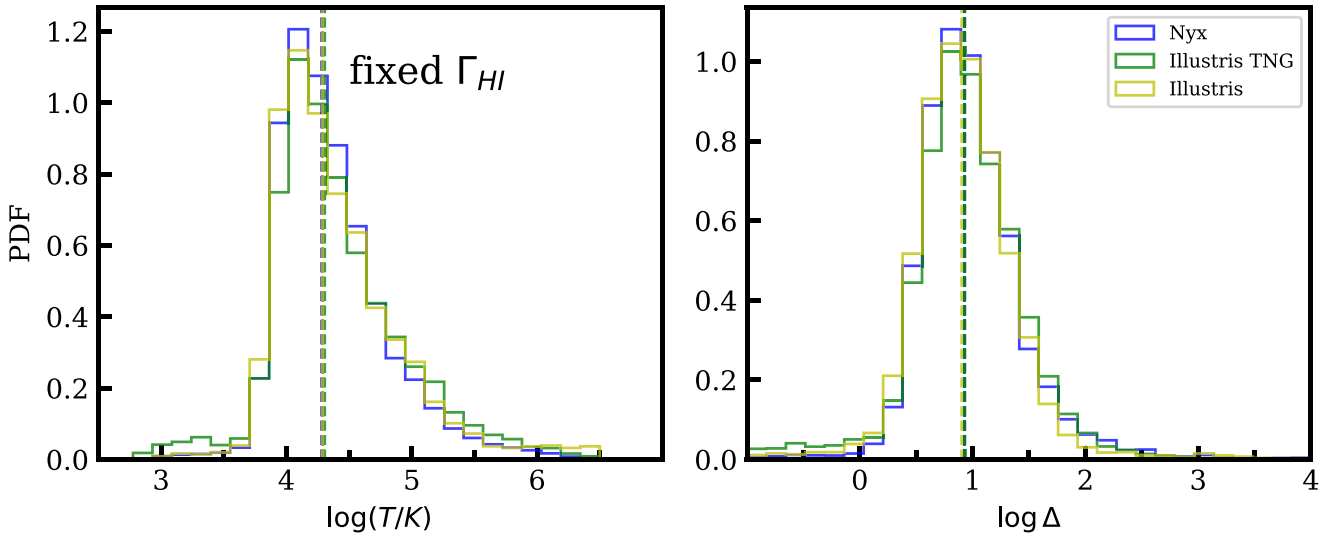

Figure 20. Marginalized 1d distribution of T (left), and Δ (right) of the simulated Ly α absorbers identified in Nyx (blue), IllustrisTNG (green), and Illustris (yellow) simulations. The medians of log T and log Δ are indicated by vertical dashed lines. The three simulations used here are post-processed to have the same UV background photoionization rate, with $\Gamma_{\rm H\,I} = -13.093$.

better approximations to the real Universe. We discover that the [T_0, γ] of these two simulations can be recovered using the inference method within reasonable offsets. Considering the inference results and the huge difference across the three simulations in the IGM WHIM fractions, we conclude that the Ly α forest does not probe the WHIM effectively under realistic conditions, and the IGM thermal state [T_0, γ] is not affected by the shock heating caused by AGN feedback and other astrophysical processes significantly at $z = 0.1$. To further confirm our conclusion, we identified the Ly α absorbers in all three simulations at $z = 0.1$, and pair them to the corresponding absorption lines identified in the mock spectra. The physical properties of the simulated Ly α absorbers support our conclusion that the observable Ly α forest are not affected by the substantial WHIM in the low-z, and the thermal state [T_0, γ] measured from the Ly α forest remains solid. In this section, we summarize our paper and present our discussion as follows.

(i) We compare the performance of [T_0, γ] as neural network training labels against the photoheating labels [A, B], i.e. the photoheating rate rescaling factors used to generate the Nyx simulation suite with various thermal histories. Given that the [A, B] parameters were actually used to generate the simulation outputs, one might expect that they would serve as a better set of labels than [T_0, γ]. However, the inference results show the efficacy of these two sets of labels are comparable, suggesting that the [T_0, γ], which parametrize the power law T–Δ relationship, still effectively characterize the Ly α observables at low-z, notwithstanding the dispersion in the T–Δ distribution induced by shock heating at low-z.

(ii) We explored the degree to which the presence of feedback can influence or bias the inference of the IGM thermal state parameters from the b–$N_{\rm H\,I}$ distribution. In the context of our inference framework, this question becomes: what would happen if we used a simulation grid without feedback to infer the thermal state of

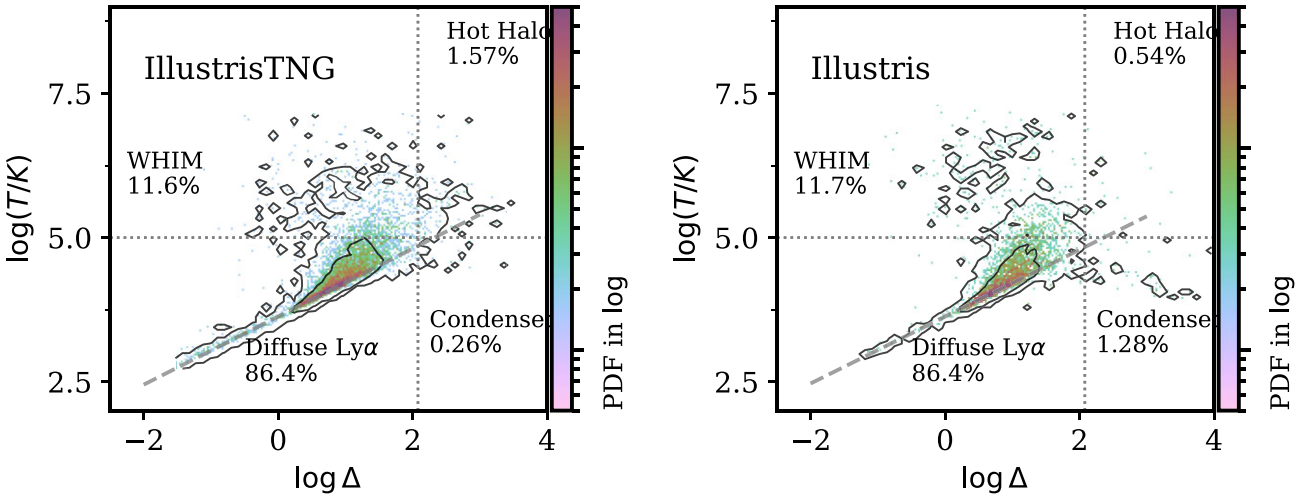

Figure 21. T–Δ distributions of the simulated Ly α absorbers in the IllustrisTNG (left) and Illustirs (right) simulations under the same $\Gamma_{\rm H\,I}$. The contours for 1σ, 3σ (68 and 99.7 per cent) of the T–Δ distribution of the absorbers are shown in both panels. The volume-weighted gas phase fractions for absorbers in both simulations are given, respectively.

a Universe that has strong feedback? Would the feedback lead to biased inference? To address these questions, we apply our inference procedure trained on Nyx simulations without feedback to mock data sets from the IllustrisTNG and Illustris simulations which include feedback, whereby the latter serve as potential proxies for the real Universe. We find that the $[T_0,\ \gamma]$ of IllustrisTNG and Illustris can be recovered within small offset, where $\Delta\log\,(T_0/{\rm K}) = -0.022$ dex, $\Delta\gamma = 0.064$ for IllustrisTNG and $\Delta\log\,(T_0/{\rm K}) = 0.047$ dex, $\Delta\gamma = -0.094$ for Illustris. These offsets are smaller than the typical precision afforded by a realistic data set, i.e. $\Delta\log T_0 \lesssim 0.5\ \sigma_{\log T_0}$, and $\Delta\gamma \lesssim \sigma_\gamma$.

(iii) We developed a method to identify regions in the simulation responsible for the Ly α absorption lines identified via Voigt-profile fitting, allowing us to determine their temperature T and overdensity Δ from the simulation skewers. For the Nyx simulations, the simulated Ly α absorbers have a median density $\log\,\Delta_{\rm median} = 0.92$, a median temperature $T_{\rm median} = 10^{4.27}$K, and about 7 per cent of the simulated Ly α absorbers have $T > 10^5$, making them outliers from the power-law T–Δ relationship. This low fraction is consistent with the previous study of Tepper-García et al. (2012) on the low-z BLAs.

(iv) As pointed out in previous work (Bolton et al. 2022a; Khaire et al. 2024; Tillman et al. 2023a), the Ly α forest is affected by the UV background, which impacts the dN/dz. Nevertheless, we observe that the temperature and overdensity of the region probed by the low-z Ly α forest are also affected by the UV background photoionization rate $\Gamma_{\rm H\,I}$ used in the simulation. For dN/dz matched simulations, the T and Δ of the simulated Ly α absorbers are correlated with its $\Gamma_{\rm H\,I}$ respectively. Specifically, the Ly α forest probes regions with higher T and Δ given a higher $\Gamma_{\rm H\,I}$. This is because for Ly α absorbers with $\tau_{{\rm Ly}\,\alpha} \sim 1$, the fluctuating Gunn–Peterson approximation implies that $\Gamma_{\rm H\,I} \propto \Delta^{2.7-\gamma} \propto T^{2/(\gamma-1)-0.7}$, where $\gamma \sim 1.6$.

(v) We post-processing the three simulations to share the same $\Gamma_{\rm H\,I}$, allowing us to explore the effects of different mechanisms. Under the same $\Gamma_{\rm H\,I}$, the T and Δ distributions of the simulated Ly α absorbers across all three simulations become almost indistinguishable, converging to nearly identical median values, while the overall IGM T–Δ distributions remain different among the simulations, due to their distinct feedback mechanisms. For the WHIM fractions, the volume-weighted WHIM fractions for IllustrisTNG and Illustris stand at 9.8 and 38.0 per cent, respectively, but the WHIM fractions for the simulated Ly α absorbers in both simulations are nearly identical, averaging around 11.6 per cent. This suggests that while feedback significantly alters the low-z IGM T–Δ distribution, especially the WHIM phase gas, their impacts on the low-z Ly α forest is indistinguishable under realistic conditions. Such a conclusion aligns with the results derived from various statistics of the low-z Ly α forest by Khaire et al. (2024).

We have thus far demonstrated the robustness of the thermal state $[T_0,\ \gamma]$ as IGM parameters at low-z, in spite of the dispersion in the T–Δ distribution induced by shock heating. We also proved that the Hu22 inference framework can effectively measure the thermal state $[T_0,\ \gamma]$ notwithstanding the feedback mechanisms implemented in the IllustrisTNG and Illustris simulations. Looking ahead, we plan to apply the Hu22 inference methodology to simulations with more flexible and sophisticated feedback mechanisms, including EAGLE (Schaye et al. 2015) and CAMELS suite (Villaescusa-Navarro et al. 2021). The outcomes will provide us with a deeper understanding of the impact of various feedback processes on low-z IGM. Moreover, by applying our methodology on archival HST COS and space telescope imaging spectrograph (STIS) data sets, we expect precise measurements of the low-z IGM thermal state. These results will pinpoint the onset of the discrepancy in the b parameters of the low-z Ly α forest between current simulations and observations, which is essential for unravelling the underlying physics and acquiring a comprehensive picture of the IGM thermal evolution at low-z after the epoch of helium reionization.

ACKNOWLEDGEMENTS

We thank the members of the ENIGMA[10], Siang Peng Oh, Timothy Brandt, and K.G. Lee for helpful discussions and suggestions.

Calculations presented in this paper used the hydra and draco clusters of the Max Planck Computing and Data Facility (MPCDF, formerly known as RZG). MPCDF is a competence center of the Max Planck Society located in Garching (Germany). This research

[10] http://enigma.physics.ucsb.edu/

also used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. In addition, we acknowledge PRACE for awarding us access to JUWELS hosted by JSC, Germany. JO acknowledges support from grants BEAGAL18/00057 and CNS2022-135878 from the Spanish Ministerio de Ciencia y Tecnologia.

DATA AVAILABILITY

The simulation data and analysis code underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Almgren A. S., Bell J. B., Lijewski M. J., Lukić Z., Van Andel E., 2013, ApJ, 765, 39
Alsing J., Wandelt B., Feeney S., 2018, MNRAS, 477, 2874
Alsing J., Charnock T., Feeney S., Wandelt B., 2019, MNRAS, 488, 4440
Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O'Neil M., 2016, IEEE Trans. Pattern Anal. Mach. Intell., 38, 252
Becker R. H. et al., 2001, AJ, 122, 2850
Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, MNRAS, 410, 1096
Bolton J. S., Becker G. D., Haehnelt M. G., Viel M., 2014, MNRAS, 438, 2499
Bolton J. S., Caputo A., Liu H., Viel M., 2022a, Phys. Rev. Lett., 129, 211102
Bolton J. S., Gaikwad P., Haehnelt M. G., Kim T.-S., Nasir F., Puchwein E., Viel M., Wakker B. P., 2022b, MNRAS, 513, 864
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Burkhart B., Tillman M., Gurvich A. B., Bird S., Tonnesen S., Bryan G. L., Hernquist L., Somerville R. S., 2022, ApJ, 933, L46
Carswell R. F., Webb J. K., 2014, Astrophysics Source Code Library, record ascl:1408.015
Cen R., Ostriker J. P., 2006, ApJ, 650, 560
Croton D. J. et al., 2006, MNRAS, 365, 11
Danforth C. W. et al., 2016, ApJ, 817, 111
Davé R. et al., 2001, ApJ, 552, 473
Davé R., Oppenheimer B. D., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 408, 2051
Debuhr J., Quataert E., Ma C.-P., 2011, MNRAS, 412, 1341
Fan X. et al., 2006, AJ, 132, 117
Faucher-Giguère C.-A., 2020, MNRAS, 493, 1614
Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Gaikwad P., Srianand R., Choudhury T. R., Khaire V., 2017, MNRAS, 467, 3172
Gaikwad P., Srianand R., Haehnelt M. G., Choudhury T. R., 2021, MNRAS, 506, 4389
Garzilli A., Theuns T., Schaye J., 2015, MNRAS, 450, 1465
Genel S. et al., 2014, MNRAS, 445, 175
Haardt F., Madau P., 2012, ApJ, 746, 125
Hiss H., Walther M., Hennawi J. F., Oñ orbe J., O'Meara J. M., Rorai A., Lukić Z., 2018, ApJ, 865, 42
Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731
Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, ApJS, 175, 356
Hu T. et al., 2022, MNRAS, 515, 2188 (Hu22)
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Khaire V., 2017, MNRAS, 471, 255
Khaire V., Srianand R., 2015, MNRAS, 451, L30
Khaire V., Srianand R., 2019, MNRAS, 484, 4174
Khaire V. et al., 2019, MNRAS, 486, 769
Khaire V., Hu T., Hennawi J. F., Walther M., Davies F., 2024, MNRAS, 527, 4545
Khaire V., Hu T., Hennawi J. F., Burchett J. N., Walther M., Davies F., 2023, preprint (arXiv:2311.08470)
Kulkarni G., Worseck G., Hennawi J. F., 2019, MNRAS, 488, 1035
Lidz A., Faucher-Giguère C.-A., Dall'Aglio A., McQuinn M., Fechner C., Zaldarriaga M., Hernquist L., Dutta S., 2010, ApJ, 718, 199
Lueckmann J.-M., Bassetto G., Karaletsos T., Macke J. H., 2018, preprint (arXiv:1805.09294)
Lukić Z., Stark C. W., Nugent P., White M., Meiksin A. A., Almgren A., 2015, MNRAS, 446, 3697
Madau P., Meiksin A., 1994, ApJ, 433, L53
Marinacci F. et al., 2018, MNRAS, 480, 5113
Martizzi D. et al., 2019, MNRAS, 486, 3766
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
McQuinn M., 2016, ARA&A, 54, 313
McQuinn M., Upton Sanderbeck P. R., 2016, MNRAS, 456, 47
McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Hopkins P. F., Dutta S., Faucher-Giguère C.-A., 2009, ApJ, 694, 842
Morrison J., Simon N., 2018, J. Comput. Graph. Stat., 27, 648
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nasir F., Bolton J. S., Viel M., Kim T.-S., Haehnelt M. G., Puchwein E., Sijacki D., 2017, MNRAS, 471, 1056
Nath B. B., Silk J., 2001, MNRAS, 327, L5
Nelson D. et al., 2018, MNRAS, 475, 624
Nelson D. et al., 2019, Computa. Astrophys. Cosmol., 6, 2
Papamakarios G., Murray I., 2016, Advances in Neural Information Processing Systems, The MIT Press, Cambridge, MA, p. 1028
Papamakarios G., Sterratt D. C., Murray I., 2018, preprint (arXiv:1805.07226)
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Prangle D., Blum M. G. B., Popovic G., Sisson S. A., 2014, Aust. N. Z. J. Stat., 56, 309
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJL, 802, L19
Rorai A. et al., 2017, Science, 356, 418
Rorai A., Carswell R. F., Haehnelt M. G., Becker G. D., Bolton J. S., Murphy M. T., 2018, MNRAS, 474, 2871
Rudie G. C., Steidel C. C., Pettini M., 2012, ApJ, 757, L30
Scannapieco E., Silk J., Bouwens R., 2005, ApJ, 635, L13
Schaye J., Theuns T., Leonard A., Efstathiou G., 1999, MNRAS, 310, 57
Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, MNRAS, 318, 817
Schaye J. et al., 2015, MNRAS, 446, 521
Sellentin E., Starck J.-L., 2019, J. Cosmol. Astropart. Phys., 2019, 021
Shull J. M., Harness A., Trenti M., Smith B. D., 2012, ApJ, 747, 100
Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877
Springel V., 2005, MNRAS, 364, 1105
Springel V., 2010, MNRAS, 401, 791
Springel V. et al., 2018, MNRAS, 475, 676
Syphers D., Shull J. M., 2014, ApJ, 784, 42
Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., 2012, MNRAS, 425, 1640
Tillman M. T. et al., 2023a, AJ,166,228
Tillman M. T., Burkhart B., Tonnesen S., Bird S., Bryan G. L., Anglés-Alcázar D., Davé R., Genel S., 2023b, ApJ, 945, L17
Viel M., Haehnelt M. G., Bolton J. S., Kim T.-S., Puchwein E., Nasir F., Wakker B. P., 2017, MNRAS, 467, L86
Villaescusa-Navarro F. et al., 2021, ApJ, 915, 71
Walther M., Oñorbe J., Hennawi J. F., Lukić Z., 2019, ApJ, 872, 13
Weinberg D. H., Hernsquit L., Katz N., Croft R., Miralda-Escudé J., 1997, in Petitjean P., Charlot S.eds, Structure and Evolution of the Intergalactic Medium from QSO Absorption Line System. p. 133 Paris: Editions Frontieres Paris, France
Weinberger R. et al., 2017, MNRAS, 465, 3291
Wolfson M., Hennawi J. F., Davies F. B., Oñorbe J., 2022, MNRAS,521,4056
Worseck G. et al., 2011, ApJ, 733, L24
Worseck G., Davies F. B., Hennawi J. F., Prochaska J. X., 2018, APJ,875,111
Ziegel J. F., Gneiting T., 2014, Electron. J. Stat., 8, 2619

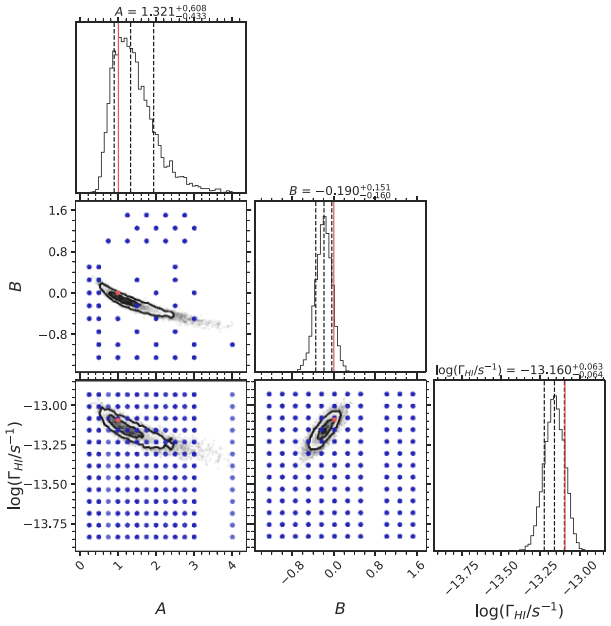

Figure A1. An example of posterior obtained by our inference method based on $[T_0, \gamma, \Gamma_{\rm HI()}]$. Projections of the thermal grid used for generating models are shown as blue dots, while the true model is shown as red dots. The inner (outer) black contour represents the projected 2D 1(2) σ interval. The parameters of true models are indicated by red lines in the marginal distributions, while the dashed black lines indicate the 16, 50, and 84 percentile values of the posterior. The true parameters are: $A = 1.0$ and $B = 0.0$, $\log(\Gamma_{\rm HI}/{\rm s}^{-1}) = -13.093$.

APPENDIX A: INFERENCE BASED ON THE PHOTOHEATING LABELS [*A*, *B*]

In this section, we present our inference results using the framework where different Nyx models are labelled by the photoheating parameters [A, B] instead of the thermal state [T_0, γ], and the inference method returns [A, B, $\log\Gamma_{\rm HI}$]. The inference is conducted following the procedures described in Section 3, based on the DELFI b–$N_{\rm HI}$ distribution emulator trained on training data set labelled by [A, B, $\log\Gamma_{\rm HI}$], which returns $P(b\ , N_{\rm HI}\ |\ A, B, \log\Gamma_{\rm HI})$.

The simulation grid, parametrized by the photoheating labels [A, B], is given in Fig. A3. An example of the MCMC posterior obtained based on the aforementioned likelihood function is given in Fig. A1. The inference method returns $A = 1.321$ (1.0), $B = -0.190$ (0.0), $\Gamma_{\rm HI} = -13.160$ (-13.093), whereas the true values are given in the parentheses. The posterior appears compact, with the medians of the marginalized posteriors landing within 1σ errors for all three parameters. The b–$N_{\rm HI}$ distribution recovered from the mock data set is presented in Fig. A2, which is emulated by our DELFI b–$N_{\rm HI}$ distribution emulator based on the inferred parameters.

We perform an inference test following the Section 3.2, in which we also exclude models that are too close to the parameter boundaries to avoid the truncation of the resulting posteriors. Specifically, we only use models with $3.3 < \log(T_0/{\rm K}) < 3.9$, $1.0 < \gamma < 2.3$, $-13.75 < \log(\Gamma_{\rm HI}/{\rm s}^{-1}) < -13.0$. The result of the inference test is shown in Fig. 9. The performance looks comparable to the one based on the thermal state [T_0, γ], suggesting that [T_0, γ] are still effective IGM parameters at low-z, notwithstanding the substantial dispersion in the IGM T-Δ distribution induced by pervasive shock heating at this redshift.

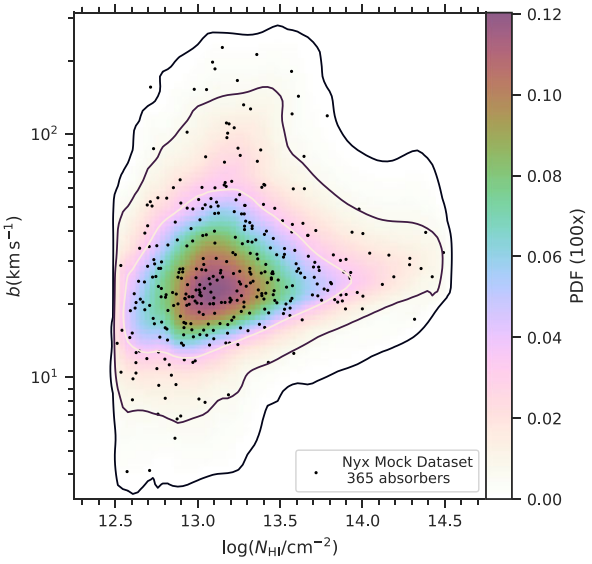

Figure A2. The colour map is the full b–$N_{\rm HI}$ distribution recovered from the Nyx mock data set, which is emulated by our DELFI emulator based on the best-fitting parameters (median values of the marginalized MCMC posterior), where $A = 3.695$ (1.0) and $B = 1.507$ (0.0) and $\log(\Gamma_{\rm HI}/{\rm s}^{-1}) = -13.237$ (-13.093), the true parameters are given in parentheses. Black dots are the mock data sets we used in the inference. The contours correspond to cumulative probabilities of 68, 95, and 99.7 per cent. For illustration purposes, the values of pdf are multiplied by 100 in the colour bar.

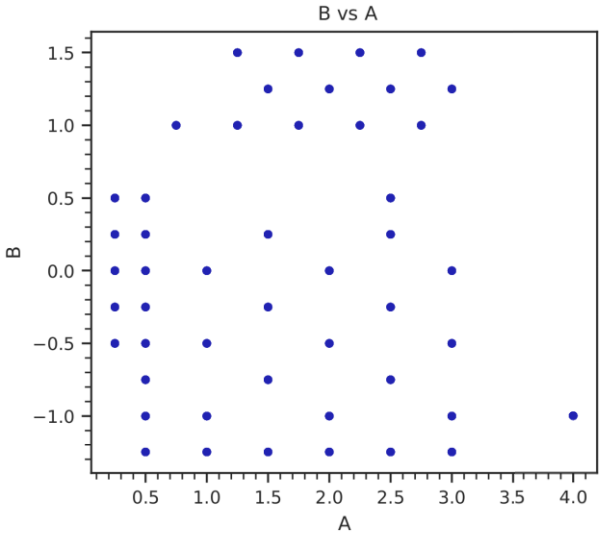

Figure A3. Parameters grid (blue circles) from snapshots of hydrodynamic simulations of the THERMAL suite at $z = 0.1$ parametrized by the thermal state [A, B].

This paper has been typeset from a TEX/LATEX file prepared by the author.