The Astrophysical Journal, 937:83 (11pp), 2022 October 1 https://doi.org/10.3847/1538-4357/ac8b09
© 2022. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

HIFLOW: Generating Diverse HI Maps and Inferring Cosmology while Marginalizing over Astrophysics Using Normalizing Flows

Sultan Hassan[1,2,3,15], Francisco Villaescusa-Navarro[1,3], Benjamin Wandelt[1,4], David N. Spergel[1,3], Daniel Anglés-Alcázar[1,5], Shy Genel[1,6], Miles Cranmer[3], Greg L. Bryan[1,7], Romeel Davé[2,8,9], Rachel S. Somerville[1], Michael Eickenberg[10], Desika Narayanan[11,12], Shirley Ho[1,3,13], and Sambatra Andrianomena[2,14]

[1] Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York, NY, 10010, USA
[2] Department of Physics & Astronomy, University of the Western Cape, Cape Town 7535, South Africa
[3] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ, 08544, USA
[4] Sorbonne Universite, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, F-75014 Paris, France
[5] Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT, 06269, USA
[6] Columbia Astrophysics Laboratory, Columbia University, New York, NY, 10027, USA
[7] Department of Astronomy, Columbia University, 550 West 120th Street, New York, NY 10027, USA
[8] Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
[9] South African Astronomical Observatories, Observatory, Cape Town 7925, South Africa
[10] Center for Computational Mathematics, Flatiron Institute, 162 5th Ave, New York, NY, 10010, USA
[11] Department of Astronomy, University of Florida, Gainesville, FL, USA
[12] University of Florida Informatics Institute, 432 Newell Drive, CISE Bldg E251, Gainesville, FL, USA
[13] Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[14] South African Radio Astronomy Observatory (SARAO), Black River Park, Observatory, Cape Town, 7925, South Africa
Received 2021 October 6; revised 2022 May 2; accepted 2022 August 17; published 2022 September 30

Abstract

A wealth of cosmological and astrophysical information is expected from many ongoing and upcoming large-scale surveys. It is crucial to prepare for these surveys now and develop tools that can efficiently extract most information. We present HIFLOW: a fast generative model of the neutral hydrogen (HI) maps that is conditioned only on cosmology (Ω_m and σ_8) and designed using a class of normalizing flow models, the masked autoregressive flow. HIFLOW is trained on the state-of-the-art simulations from the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project. HIFLOW has the ability to generate realistic diverse maps without explicitly incorporating the expected two-dimensional maps structure into the flow as an inductive bias. We find that HIFLOW is able to reproduce the CAMELS average and standard deviation HI power spectrum within a factor of $\lesssim 2$, scoring a very high $R^2 > 90\%$. By inverting the flow, HIFLOW provides a tractable high-dimensional likelihood for efficient parameter inference. We show that the conditional HIFLOW on cosmology is successfully able to marginalize over astrophysics at the field level, regardless of the stellar and AGN feedback strengths. This new tool represents a first step toward a more powerful parameter inference, maximizing the scientific return of future HI surveys, and opening a new avenue to minimize the loss of complex information due to data compression down to summary statistics.

Unified Astronomy Thesaurus concepts: Reionization (1383); Early universe (435); Cosmological parameters (339); Intergalactic medium (813); Bayesian statistics (1900)

1. Introduction

Extracting the maximum amount of cosmological and astrophysical information remains a challenge in upcoming large-scale surveys such as the Square Kilometer Array (SKA; Mellema et al. 2013), the Hydrogen Epoch of Reionization Array (DeBoer et al. 2017), the Low Frequency Array (van Haarlem et al. 2013), the Vera C. Rubin Observatory Legacy Survey of Space and Time (Ivezić et al. 2019), Nancy Grace Roman Space Telescope (Roman, Spergel et al. 2015), Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (Doré et al. 2014), and Euclid (Racca et al. 2016). In particular, evaluating the exact likelihood of the expected high-dimensional data sets from these surveys remains intractable. Because of the large memory requirements associated with the upcoming data sets, many analyses use summary statistics, which in many cases (such as the commonly used power spectrum) results in throwing away a large amount of information. Some recent works have presented methods to search for the optimal summary statistic that can successfully capture the non-Gaussian complex features, such as the information maximizing neural networks (Charnock et al. 2018) and wavelet scattering transform (Mallat 2011), to reduce dimensions while minimizing loss of information. While these methods show different levels of success, a natural way to prevent loss of information is to perform inference at the field level.

Convolutional neural networks (CNNs) have been very successful in extracting information from high-dimensional data sets by capturing non-Gaussian features. A few examples of the successful use of CNNs include the following: constraining cosmology and astrophysics (Hassan et al. 2020; Villaescusa-Navarro et al. 2021a, 2021b), identifying sources driving cosmic reionization (Hassan et al. 2019), constraining the reionization history (Mangena et al. 2020), learning galaxy properties from 21 cm lightcones (Prelogović et al. 2022),

[15] Hubble fellow.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

recovering astrophysical parameters (Gillet et al. 2019), painting HI on the matter field from *N*-body simulations (Wadekar et al. 2021), removing astrophysical effects (Villanueva-Domingo & Villaescusa-Navarro 2021), and providing optimal summary statistics for simulation-based inference (Zhao et al. 2022).

However, preparing these large-scale data sets requires running thousands of cosmological volumes using state-of-the-art hydrodynamic galaxy formation simulations by varying the cosmological and astrophysical parameters, which comes with a large computational expense. In addition, exploring the full parameter space controlling the astrophysical and cosmological observables is challenging. For instance, the state-of-the-art CAMELS (Villaescusa-Navarro et al. 2021c) project, which is the largest data set designed to train machine-learning models, provides only 1000 simulations per subgrid model, that includes all the different recipes of modeling stellar and AGN feedback below the resolution limit, for exploring its six-dimensional parameter space. Furthermore, linking these simulations directly to statistical tools, such as EMCEE (Foreman-Mackey et al. 2013) or PYDELFI (Alsing et al. 2019), to perform inference is beyond the reach of current computing capability. A powerful alternative approach is directly learn a tractable likelihood from data sets or simulations to perform parameter inference with a minimal cost. This is the goal of this paper.

Currently, there are several competing machine-learning techniques to generate new examples of large-scale data sets. This includes generative adversarial networks (GANs; Goodfellow et al. 2014), variational autoencoders (Kingma & Welling 2013), and normalizing flows (NF; Dinh et al. 2014; Jimenez Rezende & Mohamed 2015). Generating new diverse examples is crucial to ensure capturing the wide range of features present in the training sample. One common problem with GANs is called mode collapse, in which the generator always produces the same realization of the output. However, the advantages of using NF over other methods is the ability to learn the exact likelihood function to perform either inference or generate new diverse examples by inverting the flow transformations. NF methods have been very successful in generating random cosmological fields (Rouhiainen et al. 2021), simulating galaxy images (Lanusse et al. 2021), performing likelihood-free inference (e.g., Alsing et al. 2019), and modeling color–magnitude diagrams (Cranmer et al. 2019). NF attempts to learn the mapping between a standard Gaussian field and the more complex density distribution of the observable (in this case the HI maps). Once the mapping is found, new examples can be sampled simply from the initial Gaussian field. This is, in fact, somewhat conceptually similar to the flow within the most sophisticated galaxy formation and cosmological simulations. Most simulations in astrophysics and cosmology apply a series of recipes (i.e., to evolve the density and form stars) in order to transform an initial Gaussian distribution (i.e., the density field) to a complex nonlinear observable (e.g., large-scale structure, reionization morphology, cosmic evolution of star formation). Similar to galaxy formation and cosmological simulations, NF models are able to generate new diverse examples for the same set of parameters, and hence they naturally capture cosmic variance effects.

In this paper, we present HIFLOW: a fast generative model of the neutral hydrogen (HI) maps by the end of reionization at $z \sim 6$. We choose the HI fields since many of the upcoming future surveys aim to map out the HI distribution in the early universe to trace the large-scale structure. To train our emulator, we use the HI maps that are generated in a similar way as described in the CAMELS Multifield Data set[16] (Villaescusa-Navarro et al. 2022) but at $z = 6$ with a lower resolution. The state-of-the-art CAMELS simulation contains thousands of HI maps generated using SIMBA (Davé et al. 2019) and IllustrisTNG (Weinberger et al. 2017; Pillepich et al. 2018) simulations. We here focus on the HI maps generated using the ILLUSTRISTNG simulations. We first train unconditional HIFLOW and test its performance by comparing with maps from CAMELS in terms of several summary statistics such as the probability density functions (pdfs) and the power spectrum. We next train a conditional HIFLOW on the two cosmological parameters, namely Ω_m and σ_8, and validate the results at the power spectrum level. We show several examples for posterior distributions using the conditional HIFLOW. We finally demonstrate the ability of HIFLOW to perform cosmological inference while marginalizing over astrophysics at the field level.

This paper is organized as follows: We briefly discuss the simulations in Section 2 and present the NF method used in Section 3. The unconditional and conditional HIFLOW results are presented in Sections 4 and 5, respectively. We show several examples of how to perform efficient inference with HIFLOW in Section 6, and marginalize over astrophysics in Section 7. We summarize and make our concluding remarks in Section 8.

2. Simulations

We use simulations from the CAMELS project, which have been recently introduced in Villaescusa-Navarro et al. (2021c). Here, we briefly describe CAMELS and refer the reader to Villaescusa-Navarro et al. (2021c) for further details on the different simulations and data sets. CAMELS is a suite of thousands of simulations run with state-of-the-art cosmological hydrodynamic galaxy formation models, namely SIMBA (Davé et al. 2019) and ILLUSTRISTNG (Weinberger et al. 2017; Pillepich et al. 2018), by varying two cosmological parameters (Ω_m and σ_8) and four other parameters that modify the strength of stellar feedback ($A_{\rm SN1}$, $A_{\rm SN2}$) and black hole feedback ($A_{\rm AGN1}$, $A_{\rm AGN2}$) relative to the original ILLUSTRISTNG and SIMBA simulations.

We focus our analysis on two CAMELS sets, namely, the 1 parameter (1P) set, which varies a single parameter at a time with the same initial seed number, and the Latin-hyper-cube (LH) set, which is a set of 1000 simulations that explores this six-dimensional parameter space with uniform prior ranges defined as follows: $\Omega_m \in (0.1,\ 0.5)$, $\sigma_8 \in (0.6,\ 1.0)$, $A_{\rm SN1} \in (0.25,\ 4.0)$, $A_{\rm AGN1} \in (0.25, 4.0)$, $A_{\rm SN2} \in (0.5,\ 2.0)$, and $A_{\rm AGN2} \in (0.5,\ 2.0)$, with different initial seeds.

3. Masked Autoregressive Flow

HIFLOW is designed following closely the method presented in Papamakarios et al. (2017), Germain et al. (2015). NF are a class of generative models, which allow for tractable and efficient density estimation. The core principle of NF is the change-of-variable formula, which constructs a mapping (f) between a base distribution ($\pi_u(\boldsymbol{u})$, usually a Gaussian) and a more complex distribution $p(\boldsymbol{x})$ (i.e., the observable). Having obtained f, a new example of $\boldsymbol{x}$ can be generated using $\boldsymbol{x} = f(\boldsymbol{u})$,

[16] https://camels-multifield-dataset.readthedocs.io

where $\boldsymbol{u}$ is randomly drawn from the base distribution ($\boldsymbol{u} \sim \pi_u(\boldsymbol{u})$). This transformation ($f$) is required to be invertible as well as differentiable so that the target density $p(\boldsymbol{x})$ can be exactly evaluated as

$$p(\boldsymbol{x}) = \pi_u(f^{-1}(\boldsymbol{x}))\left|\det\left(\frac{\partial f^{-1}}{\partial \boldsymbol{x}}\right)\right|, \quad (1)$$

where a tractable Jacobian is needed for easy computation of the determinant. For instance, if f is a series of transformations (e.g., $f_i = (\exp(\alpha_i))^2$), then it is straightforward to find the determinant of its Jacobian as follows:

$$\left|\det\left(\frac{\partial f^{-1}}{\partial \boldsymbol{x}}\right)\right| = \prod_{i=1}^{N}\left|\det\left(\frac{\partial f_i^{-1}}{\partial \boldsymbol{x}}\right)\right| = \exp\left(-2\sum_i \alpha_i\right). \quad (2)$$

We choose to design HIFLOW using the masked autoregressive flow (MAF), which has been shown, in Papamakarios et al. (2017), to outperform many successful density estimation methods and generative models, such as the real-valued nonvolume preserving flow (Dinh et al. 2016). It is worth noting that there are recently more improved models, such as neural spline flows (Durkan et al. 2019) and generative flow with Invertible 1×1 Convolutions (Kingma & Dhariwal 2018), that achieve higher performance as compared with MAF. Autoregressive models (e.g., Uria et al. 2016; Kingma et al. 2016) can be used to estimate densities and generate new examples by decomposing a joint density $p(x)$ into a product of conditionals such as $p(\boldsymbol{x}) = \prod_i p(x_i|\boldsymbol{x}_{1:i-1})$, which ensures that future conditional transformations are only a function of the previous values, and hence satisfies the autoregressive property. If the flow is modeled using an autoencoder (i.e., series of layers), then masking is required to remove connections between different units in different layers to preserve the ordering and the autoregressive property. This approach is called the masked autoencoder for distribution estimation (MADE; Germain et al. 2015), which is the building block of the flow in MAF. MAF increases the flexibility to learn more complex distributions by stacking several autoregressive models (MADE_i, $i = 1...k$) into a deeper flow, where the density of the random numbers u_1 of MADE_1 is modeled with MADE_2, and those of MADE_2 with MADE_3 and so on, up to linking MADE_k with the base (Gaussian) density. To evaluate whether MAF is able to learn the target density, the training data set would be converted back into random numbers to test whether they represent a standard Gaussian.

We generate HI column density ($\mathrm{N_{HI}}$) maps, along any two dimensions of size $25 \times 25\ h^{-1}$ cMpc, by projecting gas particles within 5 cMpc h^{-1} columns along the third dimension from the ILLUSTRISTNG LH set with 64×64 pixels, resulting in a resolution of $\sim$0.4 h^{-1} cMpc at $z \sim 6$. These N_{HI} maps are basically generated by adding up the neutral hydrogen masses of gas particles within each column, and dividing by the pixel area and proton mass. This means we can generate five distinct HI maps along each of the three directions x, y, and z per simulation. While these 15 maps, from each simulation, are not entirely independent as expected during HIFLOW training, we assume they are due to the small size of the training set. In total, our data set contains 15,000 HI maps from the 1000 simulations of the CAMELS LH set. We use 900 simulations (or 13,500 maps) for training, and 50 simulations (or 750 maps) for validation and testing each. We convert all maps from two-dimensional to one-dimensional representation and transform the data to have a range from -1 to 1. Our best-performing MAF, as evaluated on both validation and testing sets to show the highest averaged likelihood probability, consists of 10 autoregressive layers (10 MADE). Each MADE consists of 3 hidden layers of sizes 1024, 2048, and 4096, and each conditional is parameterized as a mixture of 10 Gaussians. Training takes approximately 30 minutes on a single graphics processing unit (GPU). Following the terminology by Papamakarios et al. (2017), our design is called MAF mixture of Gaussians (MoG) (10). We use the hyperbolic tangent as an activation function throughout, Adam (Kingma & Ba 2014) as an optimizer to perform stochastic gradient descent via maximum likelihood, with a minibatch size of 100, a learning rate of 10^{-4}, a small weight decay rate of 10^{-6}, and early stopping is applied if no improvement is observed for the 30 consecutive epochs on the validation set. It is straightforward to extend this flow to learn the target density conditioned on a set of parameters ($\boldsymbol{y}$). The conditional density would be decomposed as follows: $p(\boldsymbol{x}|\boldsymbol{y}) = \prod_i p(x_i|\boldsymbol{x}_{1:i-1},\boldsymbol{y})$. A visual summary of the conditional HIFLOW on cosmological parameters (Ω_m and σ_8) is shown in Figure 1. In this analysis, we design both unconditional and conditional HIFLOW and discuss their performance in the next section.

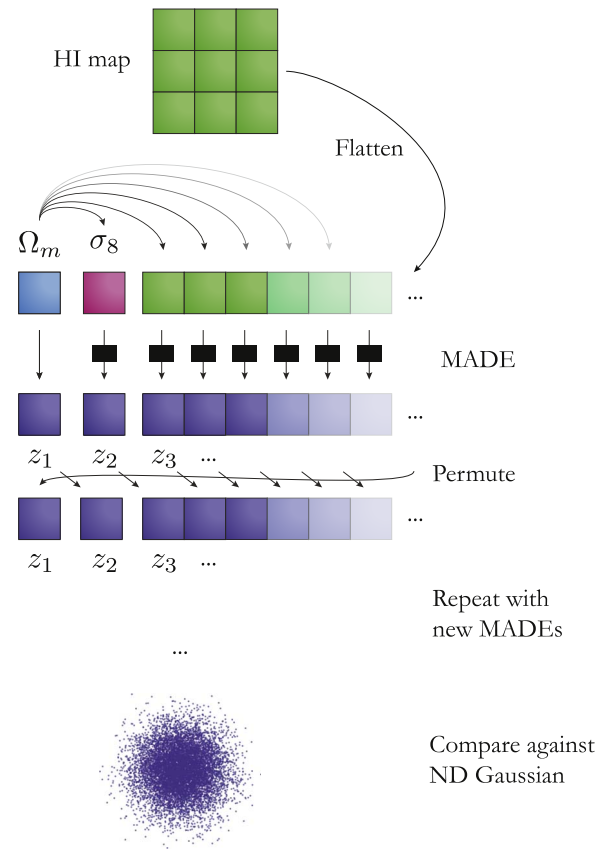

Figure 1. Diagram of the inference scheme by the conditional HIFLOW on the cosmological parameters (Ω_m and σ_8). Arrows indicate conditional dependence between variables, which constructs the flow between all MADEs in order to produce the joint density from the N-dimensional Gaussian distribution. The z variables denote the latent spaces in MADEs.

4. Unconditional HIFLOW

We now test the performance of HIFLOW, without conditioning on parameters. We first show a visual comparison

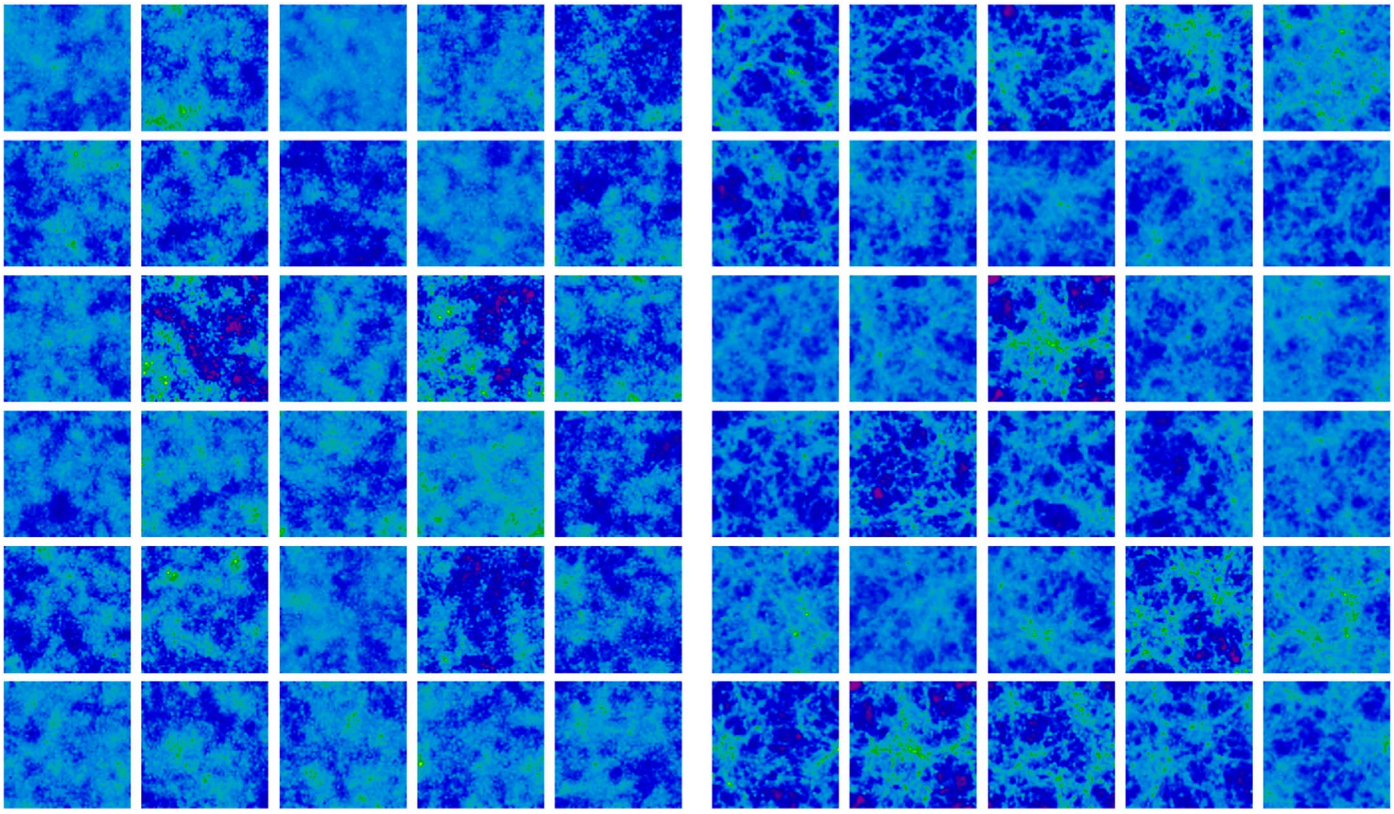

Figure 2. Random representative examples of diverse HI maps from the CAMELS testing set (real, right) and generated using the unconditional HIFLOW (fake, left). These maps cover an area of $25 \times 25\ h^{-1}$ cMpc with 64 pixels on a side, resulting in a resolution of $\sim 0.4\ h^{-1}$ cMpc. The color scale in these maps show the column density range $\log_{10} N_{\rm HI}/{\rm cm}^{-2} = 14$–22.

between the true HI maps (right) from CAMELS versus the fake HI maps generated by the unconditional HIFLOW (left) in Figure 2. We see that the real and fake maps look very similar on large scales. On the other hand, it is clear that HIFLOW does not accurately capture the small-scale features such as filaments. This is expected since the model does not explicitly incorporate the two-dimensional structure of maps as an inductive bias. For instance, using invertible 1×1 convolutions would increase the model flexibility to capture the locality and reproduce the small-scale features (e.g., Kingma & Dhariwal 2018). However, it is still promising to see that, without explicitly taking advantages of the two-dimensional information, our model is able to generate diverse new examples that capture the expected large-scale features reasonably well.

We now attempt to quantify the accuracy of the unconditional HIFLOW in terms of the power spectrum (P) and the one-dimensional pdf of the column density pixels of the generated HI maps against the CAMELS real maps. We generate 750 *fake* HI maps from the unconditional HIFLOW and compare them with the 750 HI maps from CAMELS testing set. We then compute the power spectra and pdfs over these 750 maps and compare the results in Figure 3. We show the average μ(pdf) and $\mu(P)$ in the top panels and the corresponding standard deviation (σ(pdf) and σ(P)) in the bottom panels. Comparing the average and standard deviation of the pdfs and power spectra, we see that the HIFLOW is able to reproduce CAMELS in the column density range $N_{\rm HI} \sim 10^{14-21}$ cm^{-2}, and the power spectrum within a factor of $\sim$2. The HIFLOW is able to recover the large-scale power at $<1.5\ h\,{\rm Mpc}^{-1}$ in wavenumber with a high accuracy. The disagreement on small scales (around a wavenumber of 3 $h\,{\rm Mpc}^{-1}$) is expected since the model only sees flattened maps during training, and hence it would be more challenging to model the highly nonlinear small scales without additional information such as the two-dimensional structure. This effect can be clearly seen in Figure 2. Nevertheless, for studies that focus on large scales, the HIFLOW still is an accurate, efficient tool for generating new diverse examples of the HI maps.

5. Conditional HIFLOW

We now focus on our other aim, which is to learn the HI maps conditioned on parameters. We first attempt to identify which parameters strongly affect the HI distribution. To do so, we make use of the CAMELS 1P set, in which a single parameter is changed at a time while keeping other parameters fixed. We generate maps from the 1P set, compute their average pdfs (μ(pdf)), and show the results in Figure 4. We find that the stellar ($A_{\rm SN1,\ SN2}$) and AGN ($A_{\rm AGN1,\ AGN2}$) feedback parameters have no impact on the pdf of the HI maps at $z \sim 6$. This might be due to the fact that at these early epochs there are fewer galaxies and AGN for their feedback to make a noticeable impact on HI maps on 25 h Mpc^{-1} scales. On the other hand, the HI distribution is quite sensitive to the variation in the cosmological parameters (Ω_m, σ_8), and hence we choose to condition our HIFLOW only on the cosmological parameters, while marginalizing over the astrophysical parameters. In fact, we have initially conditioned HIFLOW on all six parameters, and found worse performance (i.e., lower averaged likelihood probability over all testing set). This is due to the degeneracies between feedback parameters as seen in Figure 4. We will

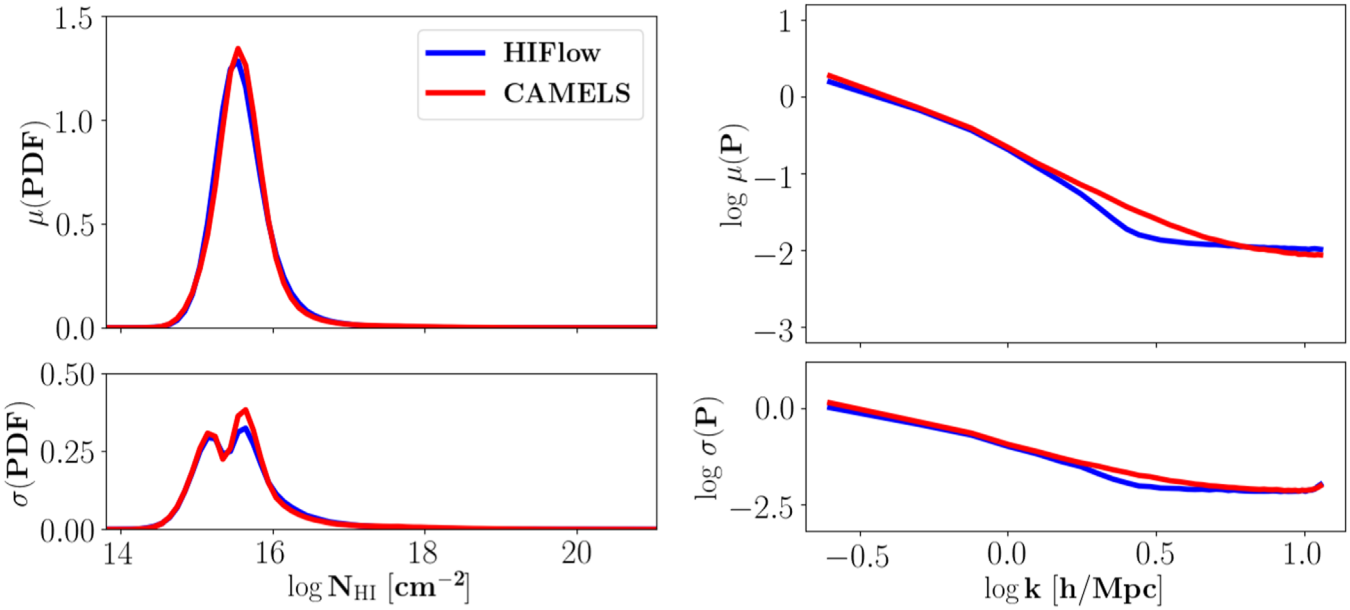

Figure 3. Comparison between CAMELS (red) and the unconditional HIFLOW (blue), in terms of the probability density distribution of the HI column density pixels (pdf, left) and power spectra of the HI maps (P, right). Top and bottom panels show the average μ(pdf) and $\mu(P)$ and standard deviation σ(pdf) and σ(P), respectively, over the 750 HI real maps from CAMELS testing set and the fake maps from HIFLOW. It is evident that HIFLOW recovers the expected pdf properties (average and standard deviation as a function of HI column density). While HIFLOW underpredicts the small-scale power by a factor of $\sim$2, it predicts the expected large-scale power very accurately. This effect is visible in Figure 2.

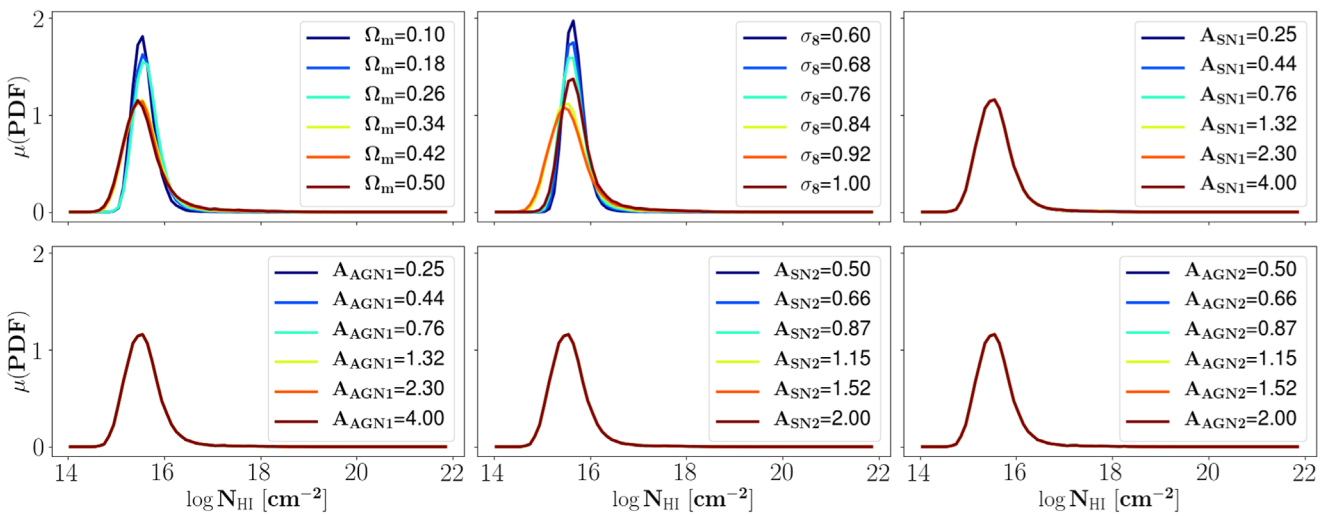

Figure 4. Impact of varying a single CAMELS parameter on the mean pdf over 15 HI maps, sharing the same parameters using the 1P set. As is evidenced by the amount of variation in the mean pdf (μ(pdf)), the HI maps are mostly affected only by the cosmological parameters (Ω_m and σ_8) and not the astrophysical parameters (A_{SN1}, A_{SN2}, A_{AGN1}, and A_{AGN2}); hence we choose to condition HIFLOW solely on cosmology.

come to this point later in Section 7. We next retrain HIFLOW, using the same architecture, to learn the following conditional density $p(\mathrm{HI}|\Omega_m, \sigma_8)$ as described earlier in Section 3.

Figure 5 shows a comparison between the conditional HIFLOW (blue) and CAMELS (red) for randomly selected values of Ω_m and σ_8 from the testing set in terms of the power spectrum. Because we extract 15 maps from each CAMELS simulation, we generate 15 new maps from the conditional HIFLOW using the same cosmology. Matching the number of samples (15) is crucial for a consistent comparison, since smaller variance is expected for a larger number of samples (i.e., $\sigma \rightarrow \sigma/\sqrt{N}$). We then compare them using the mean and standard deviation powers $\mu(P)$ and $\sigma(P)$ over all 15 maps. We quote the coefficient of determination R^2 in all panels to quantify the correlation between CAMELS and HIFLOW at the level of $\mu(P)$ and $\sigma(P)$ for the same set of the cosmological parameters. The R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum_i (\mathrm{X}_i - \mathrm{Y}_i)^2}{\sum_i (\mathrm{X}_i - \overline{\mathrm{X}})^2}, \tag{3}$$

where $X_i \equiv (\mu(P_{\mathrm{CAMELS,i}}), \sigma(P_{\mathrm{CAMELS,i}}))$ and $Y_i \equiv (\mu(P_{\mathrm{HIFlow,i}}), \sigma(P_{\mathrm{HIFlow,i}}))$. The R^2 measures the fraction of the total variance within CAMELS that HIFLOW can explain. It is worth noting that the CAMELS maps are generated using six parameters, while the conditional HIFLOW is based on two parameters. Hence, these other four parameters might provide additional sources of variance.

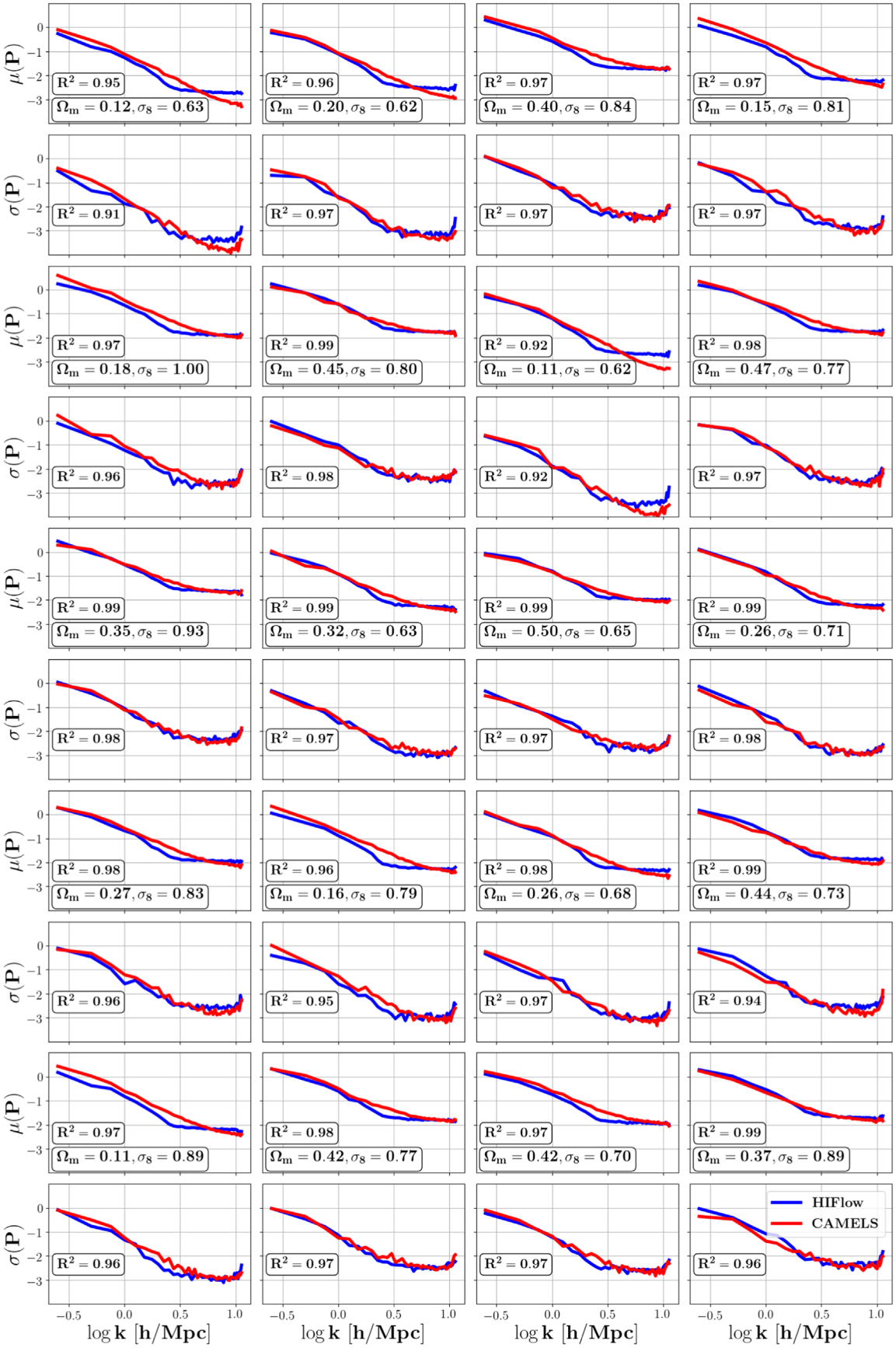

Figure 5. Comparison between the conditional HIFLOW (blue) and CAMELS (red) for randomly selected values of Ω_m and σ_8 from the testing set. The mean and standard deviation power $\mu(P)$ and $\sigma(P)$ are computed over the 15 maps. The $\sigma(P)$ panels share the same cosmology as their immediate top $\mu(P)$ panels. In all cases, the HIFLOW is able to reproduce CAMELS within a factor of $\leqslant 2$, scoring a very high $R^2 > 90\%$.

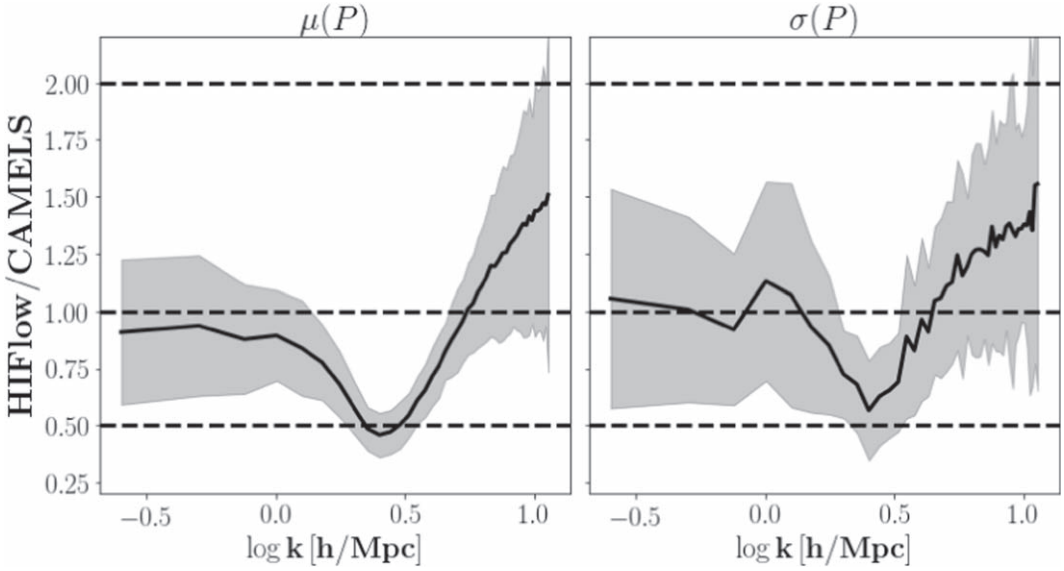

Figure 6. Ratio between the conditional HIFLOW and CAMELS of the mean (left) and standard deviation (right) power over all the testing set as quoted in the legend. Solid lines show the average and shaded areas reflect the standard deviation over all the prior range. HIFLOW predicts the large scale with a higher accuracy than the small-scale power. In all cases, the HIFLOW is able to predict the true mean and standard deviation power within a factor of $\lesssim 2$.

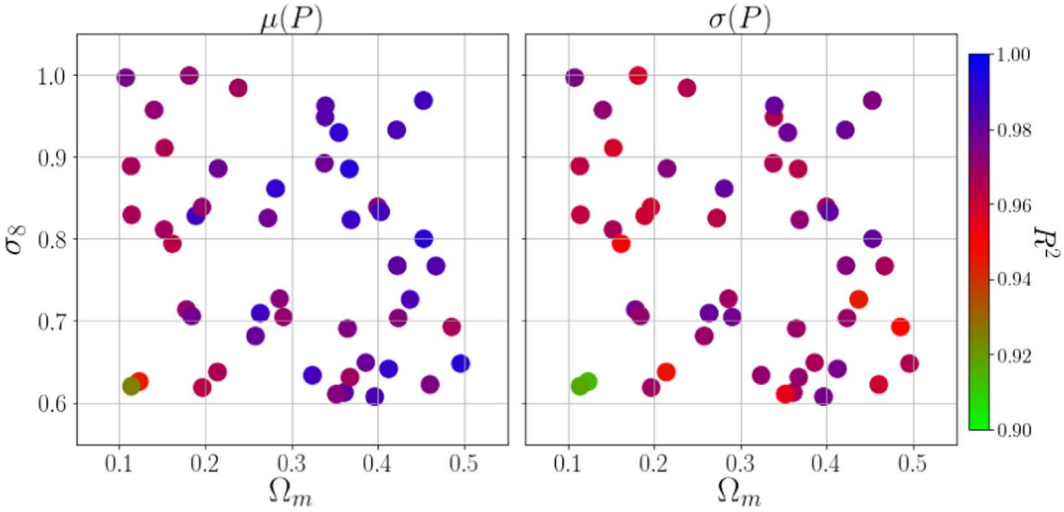

Figure 7. Visual summary of the coefficient of determination R^2 between CAMELS and HIFLOW as a function of cosmology for all the testing set (50 simulations). The R^2 values for the average power (μ(P)) and standard deviation power (σ(P)) are shown in the left and right panels, respectively as indicated by subtitles. The R^2 values are higher for $\mu(P)$ than σ(P), but nevertheless, in all cases, there is strong correlation between the CAMELS and HIFLOW with $R^2 > 90\%$.

We see that, in all cases, there is an impressive agreement between the HIFLOW and CAMELS in terms of the average and standard deviation powers ($\mu(P)$ and σ(P)) for various sets of cosmology, achieving $R^2 > 90\%$. In Figure 6, we show the ratio between the conditional HIFLOW and CAMELS of the mean (left) and standard deviation (right) power spectrum over all the testing set. The solid lines show the average, and the shaded area shows the standard deviation of the ratios, whereas dashed lines show several reference lines for the perfect match and two times more or less than the true powers as produced by CAMELS. In all cases and over all the prior range, the HIFLOW is able to reproduce CAMELS within a factor of $\lesssim 2$, depending on the wavenumber. As seen before, the larger discrepancies on small scales is expected since the model does not explicitly include the expected structure of maps as an inductive bias. We finally present a visual summary of R^2 for the testing set in Figure 7 as a function of cosmology. The R^2 values for the average power (μ(P)) and standard deviation power (σ(P)) are shown in the left and right panels, respectively as quoted in subtitles. The R^2 values are higher (darker) for the $\mu(P)$ than σ(P). This is expected since it is generally easier for models to reproduce the average behavior than higher-order statistics (e.g., the variance). Nevertheless, in all cases, there is strong correlation between the CAMELS and HIFLOW with $R^2 > 90\%$. This indicates that HIFLOW is able to explain at least 90% of the total variance of the CAMELS's power spectra data, which is quite promising due to the small size of the training data set, and model simplicity. It is worth nothing that we have retrained the conditional HIFLOW several times, and the results have always been similar. Hence, we do not expect an additional source of variance due to the random initialization of the network parameters.

To test whether the differences seen at the power spectrum level might impact inference and parameter recovery, we train a

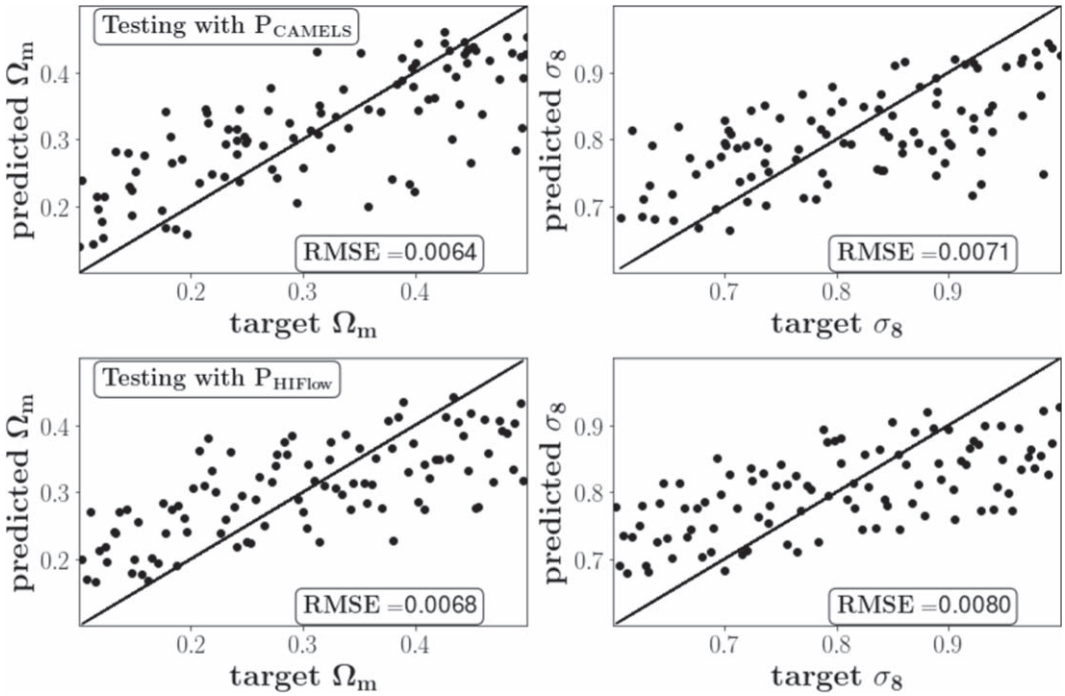

Figure 8. Correlation between the target and predicted cosmology using a simple multilayer perceptron trained on HI power spectra from CAMELS. Results of testing with the power spectra from CAMELS (P_{CAMELS}) and HIFLOW (P_{HIFlow}) are shown in the top and bottom panels, respectively. Solid black lines represent the identity line (target vs. target). While trained on CAMELS, the scatter predicted by testing with HIFLOW is similar to the target scatter produced by testing with CAMELS as quoted by the RMSE values. This shows that the differences seen in the power spectra between CAMELS and HIFLOW (see Figures 5, 6, 7) have a minimal impact on parameter recovery.

simple multilayer perceptron on HI power spectra from CAMELS and attempt to quantify the accuracy using testing samples from CAMELS versus HIFLOW as shown in Figure 8. The best-performing multilayer perceptron includes 4 layers with 10 neurons each using a learning rate of 0.001. The results of testing with power spectra from CAMELS and HIFLOW are shown in the top and bottom panels in Figure 8, respectively. The solid lines show the identify line (target versus target), and the distance to the identity lines is quantified with the root mean square error (RMSE) as quoted in the panels. We see that RMSE values and the scatter are relatively the same whether testing with data from CAMELS or HIFLOW. This indicates that the differences seen between CAMELS and HIFLOW power spectra (see Figures 5, 6, 7) would have a minimal impact on parameters recovery and inference.

6. Inferring Cosmology with HIFLOW

In the previous section, we have performed a comparison between maps generated from HIFLOW versus CAMELS, and found that the HIFLOW, for a given cosmology, is able to generate maps with power spectra that are in a good agreement with those expected from CAMELS in terms of the mean and standard deviation. We now focus on how HIFLOW can be used to perform parameter inference at the field level by inverting the flow.

The advantage of NF lies in their ability to learn a tractable likelihood that can be converted to posterior distribution. In our case, since all priors are uniform, the learned likelihood is essentially equivalent to the posterior. To estimate the posterior for a given observed set of parameters as seen in Figure 9, we adopt the following approach. We first evaluate the probability for a random set of observed parameters (Ω_m, σ_8) given their observed HI map from CAMELS (red cross symbols). We then create a uniform grid of 10,000[17] set of parameters over the whole prior range, and evaluate their probabilities given the same observed HI map. We directly use the 10,000 probabilities to estimate the 1σ, 2σ, and 3σ levels as seen in the blue contours. To visualize the probability distribution over the whole prior range, we use the Smooth bivariate spline approximation to interpolate between probabilities to create the light–dark blue background that corresponds to the high–low probability regions given the observed cosmology (red cross) as seen in Figure 9. In all cases, the red cross symbols from different parts in the prior range are always within the 1σ–3σ of the posterior, indicating that HIFLOW is able to recover the correct cosmologies for the selected observed HI maps. The contours are quite tight, and hence HIFLOW has the ability to exclude large part of the parameter space, and narrow down the broad range of possible scenarios. The well-known negative correlation between Ω_m and σ_8 is also seen in all panels. This figure illustrates an example of how to perform inference using HIFLOW for an observed HIMAP.

7. Marginalizing over Astrophysics with HIFLOW

In Figure 4, we have shown that HI column density distribution is insensitive to the variations in the astrophysical parameters. We now turn our attention to answering the question whether performing inference with HIFLOW would be affected by the variations in these astrophysical parameters. In other words, could the HIFLOW conditioned only on

[17] We have checked the results using 1000 and 10,000 grid points and found similar posterior.

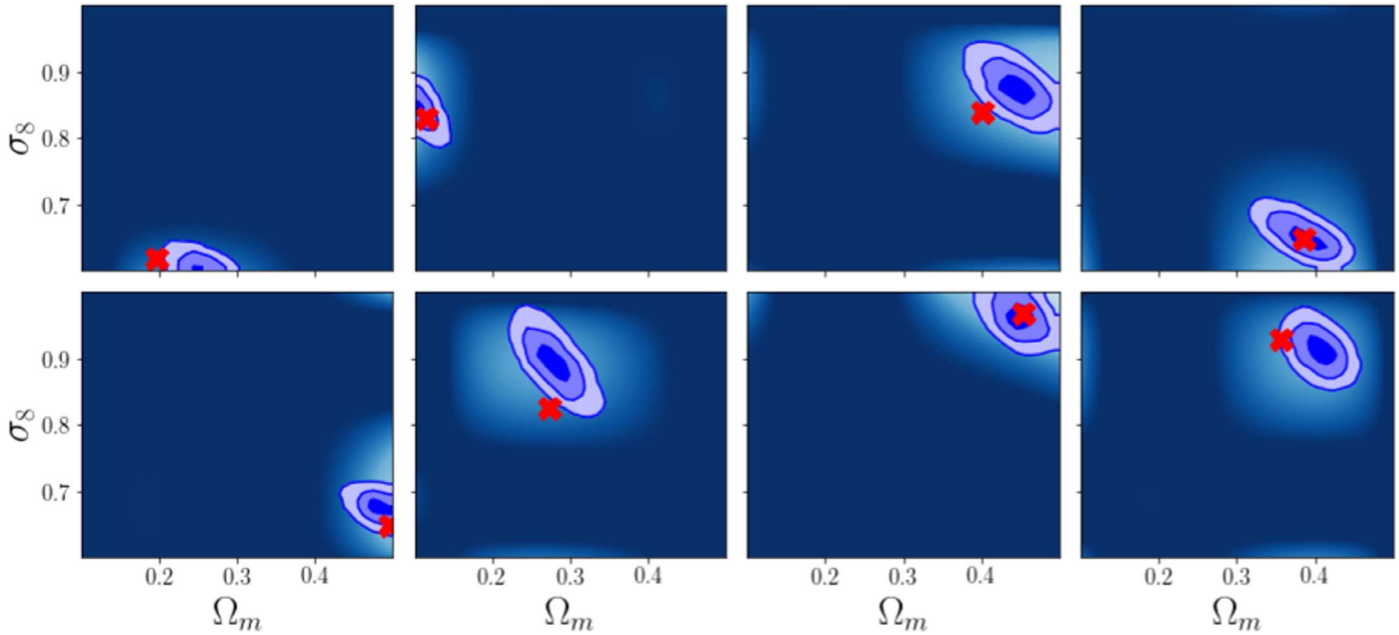

Figure 9. Several examples of posterior distributions from HIFLOW for a random set of observed parameters (red cross). The contours represent the 1σ–3σ levels of the posterior. The blue regions of the posterior indicate low probability values, and the lighter blue shows the high probability regions, where the contours reside. At different parts of the prior range, the HIFLOW is able to recover the observed HI maps within 3σ level. By inverting the flow, this is an illustrative example of how HIFLOW can be used to perform efficient and powerful parameter inference.

cosmology (Ω_m, σ_8) marginalize over the astrophysics (the other four stellar and AGN feedback parameters)?

To answer this question, we now condition the HIFLOW on all six parameters. We then test whether the correct cosmology can be recovered while varying the other four astrophysical parameters. We select the two sets with the four parameters ($A_{\rm SN1}$, $A_{\rm AGN1}$, $A_{\rm SN2}$, A_{AG2}) that correspond to low (0.62, 0.62, 0.65, 0.65) and high (3.62, 3.62, 1.85, 1.85) feedback strengths, as an example. We then repeat the same inference exercise for the same observed HI maps in Figure 9, and show the results in Figure 10. The green and orange contours are obtained using the conditional HIFLOW on all six parameters by setting the feedback parameters to low and high strength levels as defined above, respectively. The blue contours are the same as in Figure 9 that are obtained using the conditional HIFLOW only on cosmology, and the red cross shows the cosmological parameters for the selected observed HI maps. For all the selected observed parameters at different regions in the prior range, the conditional HIFLOW on cosmology produces posteriors that show high accuracy in parameter recovery, because all observed parameters (red cross) are within the 3σ level. On the other hand, conditioning on all six parameters clearly fails to recover the observed parameters for different feedback strengths (green, orange). This is expected because the column density distribution is insensitive to the variations in the feedback parameters (see Figure 4). The contours by conditioning only on cosmology (blue) are smaller than others (green, orange). This shows the ability of the model to exclude the larger part of the parameter space, and narrow the range of possible models. For all contours, the anticorrelation between Ω_m and σ_8 exists. This figure illustrates the ability of the conditional HIFLOW on cosmology to successfully marginalize over the astrophysical parameters at the field level, regardless of the strengths of the stellar and AGN feedback.

8. Concluding Remarks

We have presented HIFLOW, an efficient and fast generative model of HI maps at $z \sim 6$ from the CAMELS simulations. This new tool is designed using MAF, which is a class of NF. The MAF used here is a stack of 10 masked autoencoders for density estimation, following closely the initial implementation by Papamakarios et al. (2017). We have trained HIFLOW on 64×64 HI maps generated from the ILLUSTRISTNG LH set at $z \sim 6$ to learn the conditional density $p(\mathrm{HI}, \Omega_m, \sigma_8)$.

Our key findings can be summarized as follows:

1. The unconditional HIFLOW is able to reproduce CAMELS HI maps in the column density range $N_{\rm HI} \sim 10^{14-21}\,{\rm cm}^{-2}$, and power spectrum within a factor of $\leqslant 2$. While the model does not incorporate the two-dimensional structure of maps as an inductive bias, the large-scale power at $k < 1.5\,h\,{\rm Mpc}^{-1}$ is recovered with a high accuracy (see Figure 3).
2. While the dependence on stellar and AGN feedback is weak, the statistical properties of the HI distributions are highly sensitive to the cosmological parameters at high redshift $z \sim 6$ (see Figure 4).
3. The conditional HIFLOW on cosmological parameters (generating maps from parameters) accurately predicts the correct average and standard deviation power spectra as obtained by CAMELS within a factor of $\leqslant 2$, scoring $R^2 > 90\%$. (see Figures 5, 6, and 7). These slight differences between CAMELS and HIFLOW do not impact cosmological inference performed using the HI power spectra (see Figure 8).
4. The HIFLOW is successfully able to perform efficient cosmological inference at the field level while marginalizing over astrophysics, regardless of the strength of stellar and AGN feedback (see Figures 9 and 10).

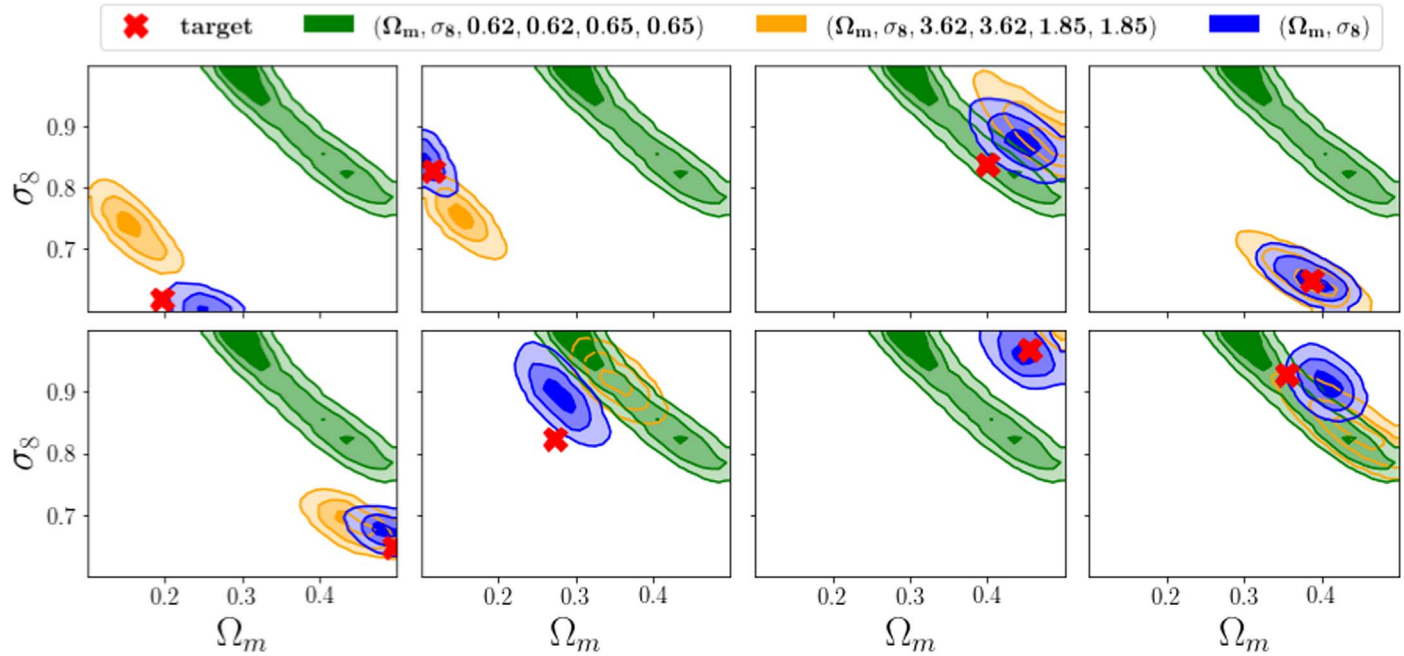

Figure 10. Similar to Figure 9, this figure shows a comparison between conditioning HIFLOW only on the cosmological parameters (blue) vs. on all six parameters (green, orange) as quoted in the legend above panels (Ω_m, σ_8, $A_{\rm SN1}$, $A_{\rm AGN1}$, $A_{\rm SN2}$, A_{AG2}). In all cases, we see that conditioning solely on cosmology is able to recover the observed HIFLOW maps (red cross). As seen in Figure 4, the astrophysical parameters do not impact the HI distribution, and hence they add degeneracies that lead to posteriors far away from the target. This shows that the conditional HIFLOW on cosmology is able to successfully marginalize over the astrophysics at the field level, regardless of the feedback strength.

While trained on ILLUSTRISTNG, the same architecture as used in HIFLOW can be used to train on other state-of-the-art hydrodynamic simulations, such as SIMBA, to generate HI maps or other morphologically similar maps, and perform efficient parameter inference at the field level.

One explanation for the good agreement between HIFLOW and CAMELS is that, at high redshift, the HI distribution is smoother and only sensitive to the cosmological parameters with no influence from the stellar and AGN feedback, as seen in Figure 4. For this reason, a simple NF model is able to learn the HI distribution very well by observing only the one-dimensional representation of the maps (i.e., flattened maps). It is expected that at lower redshifts, when the stellar and AGN feedback is much stronger, a more complex architecture might be needed. It is also worthwhile noting that the HI maps are sensitive to the UV background, and hence we do not expect HIFLOW to agree with models that employ stronger–weaker UV background than what ILLUSTRISTNG implements at $z \sim 6$.

In addition, CAMELS employs a homogeneous ionizing background (e.g., Haardt & Madau 2012), and hence the universe is fully ionized by $z \sim 6$. If instead we used an inhomogeneous background, by modeling radiative transfer (e.g., see Molaro et al. 2019; Hassan et al. 2022) either on the fly or in post processing, then the HI maps would contain the nonlinear morphology of ionized bubbles, which might be challenging to model with the current design of HIFLOW. In this case, advanced architectures, such as the neural spline flows (Durkan et al. 2019), generative flow with invertible 1×1 convolutions (GLOW; Kingma & Dhariwal 2018), or the vector quantized variational autoencoder (Razavi et al. 2019), might be needed. We leave investigating more complex architectures with more complex data sets to future work.

HIFLOW enables many applications including testing power spectral pipelines of HI surveys and assisting in computing statistical properties that require many field samples, such as the covariance matrix of HI maps or their summary statistics. HIFLOW is an initial step toward studying the non-Gaussian nature of HI maps, performing a more efficient parameter inference, powerful HI forecasting for future large-scale and intensity mapping surveys, and thereby maximizing the scientific return of observations by the next generation of facilities, such as Roman and SKA. Natural next steps would include improving the model design by incorporating the data structure as an inductive bias, and learning simultaneously different emission line intensity maps (e.g., C II, CO, Lyα) to enable a joint analysis and provide accurate predictions for future multiwavelength and multimessenger surveys.

The authors acknowledge very helpful and extensive discussions with George Papamakarios (DeepMind), which have improved the paper significantly. The authors also acknowledge comments provided by Dylan Nelson, Kaze Wong, Adrian Price-Whelan, Dan Foreman-Mackey, and Wolfgang Kerzendorf. S.H., F.V.N., B.W., D.N.S., D.A.A., S.G., B.L.B., and R.S.S. acknowledge support provided by Simons Foundation. S.H. also acknowledges support for Program number HST-HF2-51507 provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, incorporated, under NASA contract NAS5-26555. Analysis is performed at the Iron Cluster in the Flatiron Institute and the Popeye-Simons System at the San Diego Supercomputing Centre. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant No. ACI-1548562, and computational resources (Bridges) provided

through the allocation AST190009. F.V.N. was supported by funding from the WFIRST program through NNG26PJ30C and NNN12AA01C. D.A.A. was supported in part by NSF grants AST-2009687 and AST-2108944. G.L.B. was supported in part by NSF grants OAC-1835509, AST-2108470. S.A. acknowledges financial support from the South African Radio Astronomy Observatory (SARAO).

ORCID iDs

Sultan Hassan https://orcid.org/0000-0002-1050-7572
Francisco Villaescusa-Navarro https://orcid.org/0000-0002-4816-0455
Benjamin Wandelt https://orcid.org/0000-0002-5854-8269
David N. Spergel https://orcid.org/0000-0002-5151-0006
Daniel Anglés-Alcázar https://orcid.org/0000-0001-5769-4945
Shy Genel https://orcid.org/0000-0002-3185-1540
Miles Cranmer https://orcid.org/0000-0002-6458-3423
Greg L. Bryan https://orcid.org/0000-0003-2630-9228
Romeel Davé https://orcid.org/0000-0003-2842-9434
Desika Narayanan https://orcid.org/0000-0002-7064-4309
Sambatra Andrianomena https://orcid.org/0000-0001-5957-0719

References

Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, MNRAS, 488, 4440
Charnock, T., Lavaux, G., & Wandelt, B. D. 2018, PhRvD, 97, 083004
Cranmer, M. D., Galvez, R., Anderson, L., Spergel, D. N., & Ho, S. 2019, arXiv:1908.08045
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001
Dinh, L., Krueger, D., & Bengio, Y. 2014, arXiv:1410.8516
Dinh, L., Sohl-Dickstein, J., & Bengio, S. 2016, arXiv:1605.08803
Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv:1412.4872
Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, arXiv:1906.04032
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Germain, M., Gregor, K., Murray, I., & Larochelle, H. 2015, arXiv:1502.03509
Gillet, N., Mesinger, A., Greig, B., Liu, A., & Ucci, G. 2019, MNRAS, 484, 282
Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014, arXiv:1406.2661
Haardt, F., & Madau, P. 2012, ApJ, 746, 125
Hassan, S., Andrianomena, S., & Doughty, C. 2020, MNRAS, 494, 5761
Hassan, S., Davé, R., McQuinn, M., et al. 2022, ApJ, 931, 62
Hassan, S., Liu, A., Kohn, S., & La Plante, P. 2019, MNRAS, 483, 2524
Jimenez Rezende, D., & Mohamed, S. 2015, arXiv:1505.05770
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Kingma, D. P., & Dhariwal, P. 2018, arXiv:1807.03039
Kingma, D. P., Salimans, T., Jozefowicz, R., et al. 2016, arXiv:1606.04934
Kingma, D. P., & Welling, M. 2013, arXiv:1312.6114
Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2021, MNRAS, 504, 5543
Mallat, S. 2011, arXiv:1101.2286
Mangena, T., Hassan, S., & Santos, M. G. 2020, MNRAS, 494, 600
Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, ExA, 36, 235
Molaro, M., Davé, R., Hassan, S., Santos, M. G., & Finlator, K. 2019, MNRAS, 489, 5594
Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv:1705.07057
Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077
Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. 2022, MNRAS, 509, 3852
Racca, G. D., Laureijs, R., Stagnaro, L., et al. 2016, Proc. SPIE, 9904, 99040O
Razavi, A., van den Oord, A., & Vinyals, O. 2019, arXiv:1906.00446
Rouhiainen, A., Giri, U., & Münchmeyer, M. 2021, arXiv:2105.12024
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
Uria, B., Côté, M.-A., Gregor, K., Murray, I., & Larochelle, H. 2016, arXiv:1605.02226
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al. 2021a, arXiv:2109.09747
Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al. 2021c, ApJ, 915, 71
Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., et al. 2021b, arXiv:2109.10360
Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., et al. 2022, ApJS, 259, 61
Villanueva-Domingo, P., & Villaescusa-Navarro, F. 2021, ApJ, 907, 44
Wadekar, D., Villaescusa-Navarro, F., Ho, S., & Perreault-Levasseur, L. 2021, ApJ, 916, 42
Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291
Zhao, X., Mao, Y., Cheng, C., & Wandelt, B. D. 2022, ApJ, 926, 151