HIFlow: Generating Diverse HI Maps Conditioned on Cosmology using Normalizing Flow

SULTAN HASSAN1,2, FRANCISCO VILLAESCUSA-NAVARRO3,1, BENJAMIN WANDELT4,1, DAVID N. SPERGEL1,3, DANIEL ANGLES-ALCÁZAR5,1, SHY GENEL1,6, MILES CRANMER3, GREG L. BRYAN7,1, ROMEELE DAVÉ8,2,9, RACHEL S. SOMERVILLE1, MICHAEL EICKENBERG10, DESIKA NARAYANAN11,12, SHIRLEY HO1,3,13, and SAMBATRA ANDRIANOMENA14,2

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 Université, CNRS, UMR 7095, Institut d’Astrophysique de Paris, 98 bis boulevard Arago, 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

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 the maximum amount of information. We present HIFlow: a fast emulator that is able to generate neutral hydrogen (HI) maps conditioned only on cosmology ($\Omega_m$ and $\sigma_8$), after training on the state-of-the-art simulations from the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project. HIFlow is designed using a class of normalizing flow models, the Masked Autoregressive Flow (MAF), which we demonstrate are capable of generating realistic maps without explicitly using the 2D structure or accounting for any symmetries. HIFlow is able to generate new diverse HI maps in the column density range $N_{\text{HI}} \sim 10^{14} - 10^{21} \text{cm}^{-2}$ at $z \sim 6$, and naturally mimic the cosmic variance effects. Remarkably, HIFlow is able to reproduce the CAMELS average and standard deviation HI power spectrum (Pk) within a factor of $\lesssim 2$, scoring a very high $R^2 > 90\%$. HIFlow will enable the testing of Pk pipelines for HI surveys, and assist in computing other statistical properties beyond Pk that require generating new diverse samples of high dimensional datasets, such as the covariance matrix. This new tool represents a first step towards enabling rapid parameter inference at the field level, maximizing the scientific return of future HI surveys, and opening a new avenue to minimize the loss of information due to data compression.

Keywords: early universe – dark ages, reionization, first stars, methods: statistical

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 (HERA, DeBoer et al. 2017), the Low Frequency Array (LOFAR, van Haarlem et al. 2013), the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST, 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 (SPHEREx, Doré et al. 2014), and Euclid (Racca et al. 2016). 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 work has presented methods to search for the optimal summary statistic, such as the information maximising neural networks (IMNNs, 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, the robust 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: constraining cosmology and astrophysics (Hassan et al. 2020; Villaescusa-Navarro et al. 2021a,b), identifying sources driving cosmic reionization (Hassan et al. 2019), constraining the reionization history (Mangen et al. 2020), learning galaxy properties from 21 cm lightcones (Prelogović et al. 2021), recovering astrophysical parameters (Gillet et al. 2019), painting HI on the matter field from N-body simulations (Wadkar et al. 2021), removing astrophysical effects (Villanueva-Domingo & Villaescusa-Navarro 2021) and providing optimal summary statistics for simulation-based inference (Zhao et al. 2021).

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 (uncertain) 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 Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS, Villaescusa-Navarro et al. 2021c) project, which is the largest data set designed to train machine learning models, provides only 1,000 simulations per subgrid model for exploring its 6-dimensional parameter space. Furthermore, linking these simulations directly to statistical tools, such as EMCEE (Foreman-Mackey et al. 2013) or PYDENTF (Alsing et al. 2019), to perform inference is beyond the reach of current computing capability. A powerful alternative approach is to design an emulator that is able to generate diverse large-scale fields with a minimal cost. This is the goal of this paper.

Currently, there are several competing machine-learning techniques to generate new diverse examples of large scale data sets. This includes Generative Adversarial Networks (Goodfellow et al. 2014, GANs), the Vector-Quantized Variational AutoEncoder (VQ-VAE 1-2, van den Oord et al. 2017; Razavi et al. 2019), and Normalizing Flows (NF, Jimenez Rezende & Mohamed 2015; Dinh et al. 2014). 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 modelling Color-Magnitude diagrams (Crammer 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 is 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 emulator 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 Dataset\footnote{https://camels-multifield-dataset.readthedocs.io} (CMD, Villaescusa-Navarro et al. 2021d) 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 real maps from CAMELS in terms of several summary statistics such as the probability distribution and the power spectrum. We next train a conditional HIFlow on the two cosmological parameters, namely $\Omega_m$ and $\sigma_8$, and validate the results at the power spectrum level.

This paper is organized as follows: We briefly discuss the simulations in §2 and present the NF method used in §3. The unconditional and conditional HIFlow results are presented in §4 and §5, respectively. We summarize and make our concluding remarks in §6.

2. SIMULATIONS

We use simulations from the CAMELS project, that 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 datasets. 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 ($\Omega_m$ and $\sigma_8$) and four other parameters that modify the strength of stellar feedback ($A_{SN1}$, $A_{SN2}$) and black hole feedback ($A_{AGN1}$, $A_{AGN2}$) 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 1,000 simulations that explores this six-dimensional parameter space with prior ranges defined as: $\Omega_m \in (0.1, 0.5)$, $\sigma_8 \in (0.6, 1.0)$, $A_{SN1} \in (0.25, 4.0)$, $A_{AGN1} \in (0.25, 4.0)$, $A_{SN2} \in (0.5, 2.0)$, and $A_{AGN2} \in (0.5, 2.0)$, 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). Normalizing flows (NF) are a class of generative models, which allow for exact and efficient density estimation. The core principle of NF is the change-of-variable formula, which constructs a mapping ($f$) between a base distribution ($\pi_u(u)$, usually a Gaussian) and a more complex distribution $p(x)$ (i.e. the observable). Having obtained $f$, a new example of $X$ can be generated using $x = f(u)$, where $u$ is randomly drawn from the base distribution ($u \sim \pi_u(u)$). This transformation ($f$) is required to be invertible as well as differentiable so that the target density $p(x)$ can be exactly evaluated as

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

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

\[ \left| \text{det} \left( \frac{\partial f^{-1}}{\partial x} \right) \right| = \exp \left( - \sum_i \alpha_i \right). \quad (2) \]

We choose to design HIFlow using the Masked Autoregressive Flow (MAF), that has been shown to outperform many successful density estimation methods and generative models, such as the real-valued Non Volume Preserving flow (real NVP, Dinh et al. 2016). 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 $p(x) = \prod_i p(x_i | x_{1:i-1})$, which ensures that future values $i$ are only a func-
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 $25\ h^{-1}\ cMpc \times 25\ h^{-1}\ cMpc$ 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 $\log_{10} N_{HI}/cm^{-2}=14-22$.

tion of the previous values $i-1$, and hence satisfies the autoregressive property. If the flow is modelled using an autoencoder (i.e. series of layers), then masking is required to remove connections between different units in different layers to preserve ordering and the autoregressive property. This approach is called 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 modelled 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 dataset would be propagated back through the network and converted into random numbers to test whether they represent a standard Gaussian.

We generate HI column density maps by projecting gas particles within 5 cMpc/h columns from the ILLUSTRIS-TNG LH set with $64 \times 64$ pixels, resulting in a resolution of $\sim 0.4\ h^{-1}\ cMpc$ at $z \sim 6$. This means we can generate five distinct HI maps along each of the three directions $x, y$ and $z$ per simulation. In total, our dataset contains 15,000 HI maps from the 1,000 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 flatten all maps and transform the data to have a range from -1 to 1. Our best performing MAF 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. Following the terminology by Papamakarios et al. (2017), our design is called MAF MoG (10). We use the hyperbolic tangent as an activation function throughout, Adam (Kingma & Ba 2014) as an optimizer, 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 ($y$). The conditional density would be decomposed as follows: $p(x|y) = \prod_i p(x_i|\mathbf{x}_{1:i-1}, y)$. A visual summary of the conditional HIFlow on cosmological parameters ($\Omega_m$ and $\sigma_8$) is shown in Figure 1. In this analysis, we
design both unconditional and conditional HIFlow and discuss their performance in the next section.

4. UNCONDITIONAL HIFLOW

We now test the performance of HIFlow, without conditioning on parameters. We first show a visual comparison 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 use the 2D information due to flattening the maps before training. However, it is still promising to see that, without explicitly taking advantages of the 2D 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 1-dimensional probability distribution functions (PDF) 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 real 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 $\mu$(PDF) and $\mu$(P) in the top panels and the corresponding standard deviation ($\sigma$(PDF) and $\sigma$(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 $N_{HI} \sim 10^{14-21} \text{cm}^{-2}$, and the power spectrum within a factor of $\sim 2$. The HIFlow is able to recover the large scale power at $< 1.5 \text{h/Mpc}$ in wavenumber with a high accuracy. The disagreement on small scales (around a wavenumber of $3 \text{h/Mpc}$) is expected since the model only sees flattened maps during training, and hence it would be more challenging to model the highly non-linear small scales without additional information such as the 2D 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 ($\mu$(PDF)) and show the results in Figure 4. We find that the stellar ($A_{SN1,SN2}$) and AGN ($A_{AGN1,SN2}$) feedback parameters have no impact on the PDF of the HI maps at $z \sim 6$. On the other hand, the HI distribution is quite sensitive to the variation in the cosmological parameters ($\Omega_m$, $\sigma_8$), and hence we choose to condition our HIFlow only on the cosmological parameters. We next retrain HIFlow, using the same architecture, to learn the following conditional density $p(\text{HI}, \Omega_m, \sigma_8)$ as described earlier in §3.

Figure 5 shows a comparison between the conditional HIFlow (blue) and CAMELS (red) for randomly selected values of $\Omega_m$ and $\sigma_8$ 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. 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). The $R^2$ is defined as follows:

$$R^2 = 1 - \frac{\sum (\text{CAMELS}_i - \text{HIFlow}_i)^2}{\sum (\text{CAMELS}_i - \text{CAMELS})^2}.$$  

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 $\sigma$(P)) for various set 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 overall the testing set. The solid lines show the average and shaded area show the standard deviation of the ratios, whereas dashed lines show several reference lines for the perfect match and two times more/less than the true powers as produced by CAMELS. In all cases and overall the prior range, the HIFlow is remarkably 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 has no knowledge about the structure of maps due to flattening the input.

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 ($\mu$(P)) and standard deviation power ($\sigma$(P)) are shown in the left and right panel, respectively as quoted in subtitle. The $R^2$ values are higher (darker) for the $\mu$(P) than $\sigma$(P). This is expected since it is generally easier for models to reproduce the
**Figure 3.** Comparison between CAMELS (red) and the unconditional HIFlow (blue), in terms of the probability density distribution (PDF, left) and power spectra (P, right). Top and bottom panels show the average $\mu$(PDF) and $\mu$(P) and standard deviation $\sigma$(PDF) and $\sigma$(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 ($\mu$(PDF)), the HI maps are mostly affected only by the cosmological parameters ($\Omega_m$ and $\sigma_8$) and not the astrophysical parameters ($A_{SN1}$, $A_{SN2}$, $A_{AGN1}$, and $A_{AGN2}$), hence we choose to condition HIFlow solely on cosmology.
Figure 5. Comparison between the conditional HIFLOW (blue) and CAMELS (red) for randomly selected values of $\Omega_m$ and $\sigma_8$ from the testing set. The mean and standard deviation power $\mu(P)$ and $\sigma(P)$ are computed over the 15 maps. In all case, the HIFLOW is remarkably able to reproduce CAMELS within a factor of $\leq 2$, 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 ($\mu(P)$) and standard deviation power ($\sigma(P)$) are shown in the left and right panel, respectively as indicated by subtitles. The $R^2$ values are higher for $\mu(P)$ than $\sigma(P)$, but nevertheless, in all cases, there is strong correlation between the CAMELS and HIFLOW with $R^2 > 90%$. 
average behaviour than higher-order statistics (e.g. the variance). Nevertheless, in all cases, there is strong correlation between the CAMELS and HIflow variance). Nevertheless, in all cases, there is strong correlation between the CAMELS and HIflow variance. This indicates that HIflow is able to capture at least 90% of the variance of the power spectrum of CAMELS, which is remarkable given the limitation of the network design.

6. CONCLUDING REMARKS

We have presented HIflow, an efficient and fast emulator of HI maps at $z \sim 6$ from the CAMELS simulations. This new tool is designed using Masked Autoregressive Flows (MAF), which is a class of normalizing flows. 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 64x64 HI maps generated from the IllustrisTNG LH set at $z \sim 6$ to learn the conditional density $p(HI, \Omega_m, \sigma_8)$.

Our key findings can be summarized as follows:

- The unconditional HIflow is able to reproduce CAMELS HI maps in the column density range $N_{HI} \sim 10^{14-21} \text{cm}^{-2}$, and power spectrum within a factor of $\leq 2$. While the model does not explicitly use the 2D information due to training on flattened maps, the large scale power at $k < 1.5 \text{h/Mpc}$ is recovered with a high accuracy (see Figure 3).

- 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).

- 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 $\leq 2$, scoring $R^2 > 90\%$. (see Figure 5, Figure 6, and Figure 7).

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.

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 normalizing flow model is able to learn the HI distribution very well by observing only 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 modelling radiative transfer (e.g. see Hassan et al. 2021; Molaro et al. 2019) either on-the-fly or in post-processing, then the HI maps would contain the non-linear 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 (VQ-VAE, 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. HIFlow is an initial step towards studying the non-Gaussian nature of HI maps, performing rapid parameter inference at the field level, efficient 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.

ACKNOWLEDGEMENTS

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, and Adrian Price-Whelan. SH, FVN, BW, DNS, DAA, SG, BLB, and RSS acknowledge support provided by Simons Foundation. FVN was supported by funding from the WFIRST program through NNG26PJ30C and NNN12AA01C. DAA was supported in part by NSF grants AST-2009687 and AST-2108944. GLB was supported in part by NSF grants OAC-1835509, AST-2108470. SA acknowledges financial support from the South African Radio Astronomy Observatory (SARAO).
REFERENCES

Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, MNRAS, 488, 4440, doi: 10.1093/mnras/stz1960
Charnock, T., Lavaux, G., & Wandelt, B. D. 2018, PhRvD, 97, 083004, doi: 10.1103/PhysRevD.97.083004
Cranmer, M. D., Galvez, R., Anderson, L., Spergel, D. N., & Ho, S. 2019, arXiv e-prints, arXiv:1908.08045.
https://arxiv.org/abs/1908.08045
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827, doi: 10.1093/mnras/stz937
DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001, doi: 10.1088/1538-3873/129/974/045001
Dinh, L., Krueger, D., & Bengio, Y. 2014, arXiv e-prints, arXiv:1410.8516. https://arxiv.org/abs/1410.8516
Dinh, L., Sohl-Dickstein, J., & Bengio, S. 2016, arXiv e-prints, arXiv:1605.08803.
https://arxiv.org/abs/1605.08803
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067
Germain, M., Gregor, K., Murray, I., & Larochelle, H. 2015, arXiv e-prints, arXiv:1502.03509.
https://arxiv.org/abs/1502.03509
Gillet, N., Mesinger, A., Greig, B., Liu, A., & Ucci, G. 2019, MNRAS, 484, 282, doi: 10.1093/mnras/stz010
Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014, arXiv e-prints, arXiv:1406.2661.
https://arxiv.org/abs/1406.2661
Haardt, F., & Madau, P. 2012, ApJ, 746, 125, doi: 10.1088/0004-637X/746/2/125
Hassan, S., Andrianomena, S., & Doughty, C. 2020, MNRAS, 494, 5761, doi: 10.1093/mnras/staa1151
Hassan, S., Davé, R., McQuinn, M., et al. 2021, arXiv e-prints, arXiv:2109.03840.
https://arxiv.org/abs/2109.03840
Hassan, S., Liu, A., Kohn, S., & La Plante, P. 2019, MNRAS, 483, 2524, doi: 10.1093/mnras/sty2828
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111, doi: 10.3847/1538-4357/ab042c
Jimenez Rezende, D., & Mohamed, S. 2015, arXiv e-prints, arXiv:1505.05770.
https://arxiv.org/abs/1505.05770
Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980. https://arxiv.org/abs/1412.6980
Kingma, D. P., & Dhariwal, P. 2018, arXiv e-prints, arXiv:1807.03039.
https://arxiv.org/abs/1807.03039
Kingma, D. P., Salimans, T., Jozefowicz, R., et al. 2016, arXiv e-prints, arXiv:1606.04934.
https://arxiv.org/abs/1606.04934
Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2021, MNRAS, 504, 5543, doi: 10.1093/mnras/stab1214
Mallat, S. 2011, arXiv e-prints, arXiv:1101.2286.
https://arxiv.org/abs/1101.2286
Mangena, T., Hassan, S., & Santos, M. G. 2020, MNRAS, 494, 600, doi: 10.1093/mnras/staa750
Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, Experimental Astronomy, 36, 235, doi: 10.1007/s10686-013-9334-5
Molaro, M., Davé, R., Hassan, S., Santos, M. G., & Finlator, K. 2019, MNRAS, 489, 5594, doi: 10.1093/mnras/stz2171
Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv e-prints, arXiv:1705.07057.
https://arxiv.org/abs/1705.07057
Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077, doi: 10.1093/mnras/stx2656
Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. 2021, arXiv e-prints, arXiv:2107.00018.
https://arxiv.org/abs/2107.00018
Racca, G. D., Laureijs, R., Stagnaro, L., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9904, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, ed. H. A. MacEwen, G. G. Fazio, M. Lystrup, N. Batalha, N. Siegler, & E. C. Tong, 99040O, doi: 10.1117/12.2230762
Razavi, A., van den Oord, A., & Vinyals, O. 2019, arXiv e-prints, arXiv:1906.00446.
https://arxiv.org/abs/1906.00446
Rouhiainen, A., Giri, U., & Münchmeyer, M. 2021, arXiv e-prints, arXiv:2105.12024.
https://arxiv.org/abs/2105.12024
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757.
https://arxiv.org/abs/1503.03757
Uria, B., Côté, M.-A., Gregor, K., Murray, I., & Larochelle, H. 2016, arXiv e-prints, arXiv:1605.02226.
https://arxiv.org/abs/1605.02226
van den Oord, A., Vinyals, O., & Kavukcuoglu, K. 2017, arXiv e-prints, arXiv:1711.00937.
https://arxiv.org/abs/1711.00937
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2, doi: 10.1051/0004-6361/201220873
Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S.,
et al. 2021a, arXiv e-prints, arXiv:2109.09747.
https://arxiv.org/abs/2109.09747

Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D.,
et al. 2021b, arXiv e-prints, arXiv:2109.10360.
https://arxiv.org/abs/2109.10360

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S.,
et al. 2021c, ApJ, 915, 71, doi: 10.3847/1538-4357/abf7ba

Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D.,
et al. 2021d, arXiv e-prints, arXiv:2109.10915.
https://arxiv.org/abs/2109.10915

Villanueva-Domingo, P., & Villaescusa-Navarro, F. 2021, ApJ, 907, 44, doi:10.3847/1538-4357/abd245

Wadekar, D., Villaescusa-Navarro, F., Ho, S., & Perreault-Levasseur, L. 2021, ApJ, 916, 42,
doi: 10.3847/1538-4357/ac033a

Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291, doi: 10.1093/mnras/stw2944

Zhao, X., Mao, Y., Cheng, C., & Wandelt, B. D. 2021, arXiv e-prints, arXiv:2105.03344.
https://arxiv.org/abs/2105.03344