FLORAH: A generative model for halo assembly histories

Tri Nguyen [1 2 3]   Chirag Modi [3 4]   L. Y. Aaron Yung [5]   Rachel S. Somerville [3]

Abstract

Dark matter accounts for 85% of the matter in our Universe. The mass assembly history (MAH) of dark matter halos plays a leading role in shaping the formation and evolution of galaxies. MAHs are used extensively in semi-analytic models of galaxy formation, yet current analytical methods to generate them are unable to capture their relationship with the halo internal structure and large-scale environment. This paper introduces FLORAH, a machine-learning framework for generating assembly histories of dark matter halos. We train FLORAH on the assembly histories from the MultiDark N-body simulations and demonstrate its ability to recover key properties such as the time evolution of mass and dark matter concentration. By applying the Santa Cruz semi-analytic model on FLORAH-generated assembly histories, we show that FLORAH correctly captures assembly bias, which cannot be reproduced with current analytical methods. FLORAH is the first step towards a machine learning-based framework for planting merger trees; this will allow the exploration of different galaxy formation scenarios with great computational efficiency at unprecedented accuracy.

1. Introduction

In the Λ-Cold Dark Matter (ΛCDM) framework, dark matter (DM) halos form hierarchically via the mergers of smaller DM halos (White & Rees, 1978). The process by which DM and baryonic matter (such as gas and stars) come together to form a gravitationally bound structure is known as the assembly history of a halo or galaxy. In simulations, this is captured in the form of "merger trees", i.e. progenitor and descendant halos are linked across multiple snapshots. The properties of galaxies (such as stellar mass and star formation rate) are thus closely linked to the assembly history of their halos and their formation environment. Understanding this intricate halo-galaxy connection remains one of the key open questions of modern astrophysics.

N-body simulations provide a powerful tool to directly study the formation and evolution of DM halos as they merge with other halos and interact with the large-scale environment (Vogelsberger et al., 2020). Because computational cost grows rapidly with the simulated volume and resolution, these simulations are often run with only DM particles and do not include baryonic physics. Semi-analytic models (SAMs) are then commonly used to populate DM halos with galaxies. SAMs combine simplified prescriptions for baryonic physics (e.g. radiative cooling, star formation, AGN feedback, etc.) within merger trees to calculate observable properties (Somerville & Davé, 2015; Yung et al., 2019).

Reliably resolved merger trees are key to accurately modeling the evolutionary history of galaxies and their properties. Merger trees extracted from N-body simulations are quite accurate, though at a great computational cost. It is not feasible to run multiple simulations with sufficient dynamic range to simultaneously capture halos encompassing dwarf galaxies ($10^5 - 10^{10}$ M$_\odot$) to galaxy clusters ($10^{14} - 10^{15}$ M$_\odot$) up to high redshifts and in large volumes.

To work around these computational challenges, previous works have developed analytic methods that rely on the Extended Press-Schechter (EPS) formalism (Bond et al., 1991; Bower, 1991). EPS trees are constructed by sampling the conditional mass probability $p(M_1|M_0, z_0, z_1)$ that a halo with mass M_0 at redshift z_0 had a mass of M_1 at an earlier redshift $z_1 > z_0$. For halos of any given initial mass M_0 and redshift z_0, the algorithm uses Monte-Carlo methods and assumes the Markov property to construct its past merger histories (Somerville & Kolatt, 1999; Zentner, 2007). However, this approach does not fully capture the subtleties of the accretion process. For instance, the accretion rate depends on the environment in which the halos form – it also

---

[1]Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA [2]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, Massachusetts 02139, USA [3]Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010 [4]Center for Computational Mathematics, Flatiron Institute, 162 5th Avenue, New York, NY 10010 [5]NASA Goddard Space Flight Center, 8800 Greenbelt Road Greenbelt, MD 20771, USA. Correspondence to: Tri Nguyen <tnguy@mit.edu>.

varies over time as the cosmic matter density evolves as a function of time. EPS techniques are unable to capture the relationship among the assembly history, the halo structure, and the environment and thus can disagree with trees extracted from N-body simulations at the factor of few levels, e.g. (Li et al., 2007).

In this paper, we introduce FLORAH, a flow-based generative recurrent model, to generate assembly histories of halos. As a first step, we focus on generating the mass assembly histories (MAHs) and DM concentration histories only on the main progenitor branches (MPBs) of merger trees. The MPB tracks the most massive progenitors of a halo and thus is the most important for understanding the assembly history. MPBs can be naturally modeled as a time-ordered sequence; hence we use a recurrent neural network to learn their representative features (mass and concentration). Because we are interested in learning the full distribution of possible assembly histories for a halo, we then use a normalizing flow to perform sampling based on these representative features. Once trained, FLORAH can match the MAHs and DM concentration histories of MPBs from the N-body simulations. In addition, we apply the Santa Cruz SAM (SC-SAM) (Somerville et al., 2015; Gabrielpillai et al., 2022) to populate FLORAH-generated MPBs with galaxies and demonstrate that FLORAH can correctly capture the assembly bias of galaxies, which existing EPS-based methods cannot reproduce.

2. Methodology

2.1. Simulation and data preprocessing

We use the VSMDPL box from the MultiDark simulations (Klypin et al., 2016). VSMDPL is a dark-matter only, N-body simulation run with the GADGET-2 smoothed-particle hydrodynamics code (Springel et al., 2001; Springel, 2005). The simulation adopts cosmological parameters that are broadly consistent with the Planck 2013 results (Planck Collaboration, 2014): $\Omega_m = 0.307, \Omega_\Lambda = 0.693, h = 0.678, \sigma_8 = 0.823, n_s = 0.960$. VSMDPL has a box size of $160\,\mathrm{Mpc}\,h^{-1}$, a DM particle mass $M_{\mathrm{DM}} = 6.2 \times 10^6$ $\mathrm{M}_\odot\,h^{-1}$, a number of the particles of $N = 3840^3$, and an adaptive softening length $\epsilon = 1.0 - 2.0\,\mathrm{kpc}\,h^{-1}$.

We use the publicly available merger trees from VSMDPL, which have been constructed using the ROCKSTAR halo finder and the CONSISTENTTREE algorithm (Behroozi et al., 2013a;b). For each tree, we extract only the main progenitor branches (MPBs) and exclude all MPBs with initial halos (also known as "root" halos) with fewer than 500 DM particles; halos below this limit will have progenitors that are not well-resolved. To improve the generalization of our model, we augment the dataset by creating multiple "sub-branches" for each MPB in the following way. For each branch, we first randomly choose an initial snapshot from

the first 40 snapshots. Then, we sample the branch every $2 - 6$ snapshots up to a maximum redshift of $z_{\mathrm{train}}^{\max} = 10$ (snapshot 108). Randomizing the time steps helps prevent the model from becoming overly reliant on one particular set of time steps and improves generalization. We repeat the augmentation steps 10 times for each MPB.

The input features to the recurrent neural network at each stage are the logarithm of the virial mass $\log_{10} M_{\mathrm{vir}}$, the Navarro-Frenk-White DM concentration c_{vir} (Navarro et al., 1996), and scale factors a of the halo and the next progenitor. Including c_{vir} helps the model learn assembly histories more accurately, possibly because c_{vir} has been found to capture the environmental dependency in many assembly bias studies (Wechsler et al., 2006). We will expand to more halo features (e.g. spin, shape) and environment features (e.g. local density) in future work. The scale factors of the halo and its next progenitor serve as the time features in our framework. For target features, we use the logarithm of the *accreted* mass, defined as:

$$\Delta \log_{10} M_{\mathrm{vir}}^{(i+1)} = \log_{10}(M_{\mathrm{vir}}^{(i+1)}/M_{\mathrm{vir}}^{(i)}), \qquad (1)$$

for the $(i + 1)$-th halo, and the concentration c_{vir} of the next progenitor. Using accreted masses $\Delta \log_{10} M_{\mathrm{vir}}^{(i+1)}$, instead of progenitor masses $\log_{10} M_{\mathrm{vir}}^{(i+1)}$, as targets improve the model performance. During generation, progenitor masses can be derived from masses and accreted masses using Eq. 1.

2.2. Neural network architecture

We model each sub-branch as a sequence of N halos, with the root halo denoted with the index zero. The input and target feature vectors are:

$$\vec{x} = \{x^{(i)} \in \mathbb{R}^{f_{\mathrm{in}}}\} = \{\log_{10} M_{\mathrm{vir}}^{(i)}, c_{\mathrm{vir}}^{(i)}, a^{(i)}, a^{(i+1)}\}, \quad (2)$$

$$\vec{y} = \{y^{(i+1)} \in \mathbb{R}^{f_{\mathrm{out}}}\} = \{\Delta \log_{10} M_{\mathrm{vir}}^{(i+1)}, c_{\mathrm{vir}}^{(i+1)}\}, \quad (3)$$

where $f_{\mathrm{in}} = 4$, $f_{\mathrm{out}} = 2$, and $i = 0, ..., N - 2$. Because we do not include any "end-of-sequence" token in our framework, the feature vector includes only the first $N - 1$ halos. During the generation process, we can choose to terminate the assembly history at a maximum redshift or minimum progenitor mass. Our goal is to estimate the truth conditional distribution of $y^{(i+1)}$, denoted as $p(y^{(i+1)}|\{x^{(\leq i)}\})$.

We use a recurrent neural network $g_\varphi : \mathbb{R}^{N_{\mathrm{in}}} \to \mathbb{R}^H$ with parameters φ to extract H summary features from the input features of each halo. The summary features are then:

$$\vec{z} = \{z^{(i)} \in \mathbb{R}^H\} = \{g_\varphi(x^{(i)}, h^{(i)})\} \qquad (4)$$

$$h^{(i)} = z^{(i-1)} \text{ if } i > 0 \text{ else } 0 \qquad (5)$$

The hidden state $h^{(i)}$ is dependent on the input features of all the previous halos $x^{(\leq i)}$, which allows the network
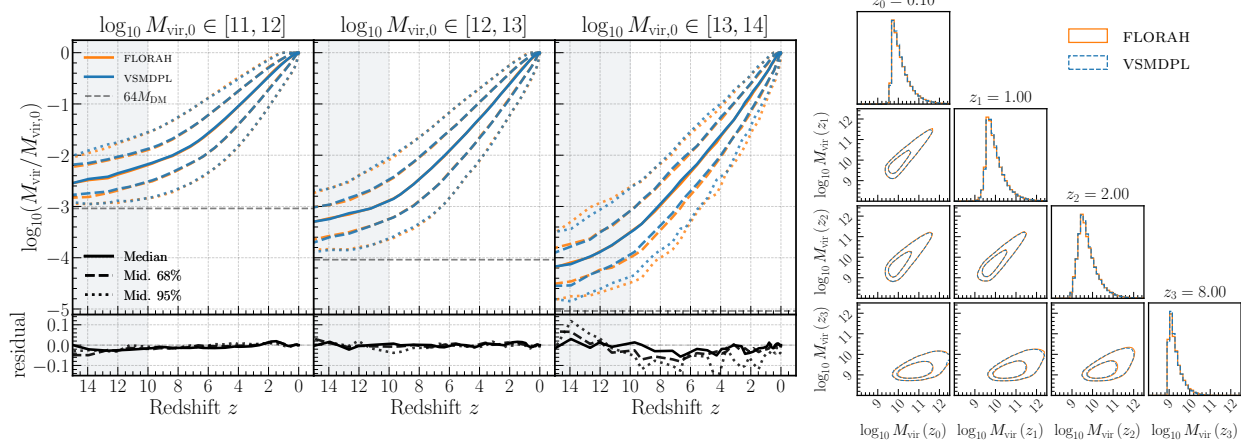
Figure 1. Left: The top panel shows the median, middle-68%, and middle-95% percentile containment regions of the MAHs of MPBs in VSMDPL (blue) and generated by FLORAH (orange) in three mass bins (in $M_\odot$ unit). The shaded gray box ($z > 10$) denotes the "extrapolation region" beyond the maximum training redshift. The residuals are shown in the bottom panel. Right: The joint distribution of $\log_{10} M_{\rm vir}$ across a few chosen redshifts. The contour lines show the 68% and the 95% intervals.

to "memorize" the entire assembly history. Our recurrent network consists of 2 Gated Recurrent Unit (GRU) layers, each with $H = 64$ hidden channels.

To estimate $p(y^{(i+1)}|\{x^{(\leq i)}\})$, we use a normalizing flow (Rezende & Mohamed, 2015; Papamakarios et al., 2017; 2019) that is conditioned on the summary features $z^{(i)}$. The flow thus estimates a conditional probability distribution $\hat{p}_\phi(y^{(i+1)}|g_\varphi(x^{(i)}, h^{(i)}))$ with learnable parameters ϕ. Our flow model consists of 4 Masked Autoregressive Flow (MAF) transformations. Each MAF includes a 4-layer Masked Autoencoder for Distribution Estimation (MADE) with a hidden dimension of 128 (Germain et al., 2015).

During training, we optimize the parameters $\{\varphi, \phi\}$ of the GRU and flow simultaneously using the negative log-density

$$\mathcal{L} = - \sum_{i=0}^{N-2} \log \hat{p}_\phi(y^{(i+1)}|g_\varphi(x^{(i)}, h^{(i)})) \qquad (6)$$

as the optimization loss. We use the AdamW optimizer (Loshchilov & Hutter, 2019) with default parameters and train until convergence. At the end of each *epoch* (defined as one full iteration over the training set), we evaluate the loss on the validation samples and reduce the learning rate by a factor of 10 if no improvement is seen after 20 epochs. Training is terminated if the validation loss has not improved after 40 epochs, which typically takes $\sim 200 - 300$ epochs or ~ 8 hours on a single NVIDIA Tesla V100 GPU.

To generate the MPB with a trained model, we start with a root feature $x^{(0)}$, an initial hidden state $h^{(0)}$ set to zero, and a list of scale factors $\{a^{(i)}\}$. At each time step i, we pass a halo feature $x^{(i)}$ and hidden state $h^{(i)}$ through the GRU layers, extract the summary features $z^{(i)} = g_\varphi(x^{(i)}, h^{(i)})$, and use the flow to sample the first progenitor halo $\hat{y}^{(i+1)} \sim$

$\hat{p}_\phi(\hat{y}^{(i+1)}|z^{(i)})$. We then convert the accreted mass to progenitor mass using Eq. 1. We use this sampled progenitor mass as the input for the next time step with the feature vector $x^{(i+1)} = (\log_{10} M_{\rm vir}^{(i+1)}, c_{\rm vir}^{(i+1)}, a^{(i+1)}, a^{(i+2)})$ and hidden state $h^{(i+1)} = z^{(i)}$. We repeat this procedure until a minimum halo mass or a maximum redshift is reached.

3. Results

For the training and validation datasets, we extract $306,014$ and $34,400$ MPBs, respectively, from an $(80\,{\rm Mpc}\,{\rm h}^{-1})^3$ sub-volume of VSMDPL and apply the preprocessing steps in Section 2.1. For our test dataset, we extract $387,031$ MPBs from a $(80\,{\rm Mpc}\,{\rm h}^{-1})^{-3}$ VSMDPL sub-volume different from the one used for training/validation. For the generation process, we take the initial halo at $z = 0$ of each MPB in the test dataset and use the mass and concentration as the initial input $x^{(0)} = (\log_{10} M_{\rm vir}^{(0)}, c_{\rm vir}^{(0)})$ We sample the scale factors every $2 - 6$ snapshots starting from $z = 0$ to obtain up to a redshift of $z = 15$, beyond the maximum training redshift ($z_{\rm train}^{\max} = 10$), to explore the model's capability for extrapolation. Additionally, we terminate generation when the progenitor mass falls below $100\,M_{\rm DM}$; below this limit, ROCKSTAR does not reliably identify halos, and $M_{\rm vir}$ and $c_{\rm vir}$ can have large uncertainties. For a fair comparison with VSMDPL, we also remove these unresolved halos in VSMDPL. We generate $387,031$ FLORAH MPBs.

In the left panel of Figure 1, we show the median, middle-68% percentile, and middle-95% percentile containment region of the MAHs of VSMDPL (blue) and FLORAH (orange) MPBs, along with their residuals, in three mass bins. The residual of the containment region is computed by averaging the residuals of the corresponding upper and lower
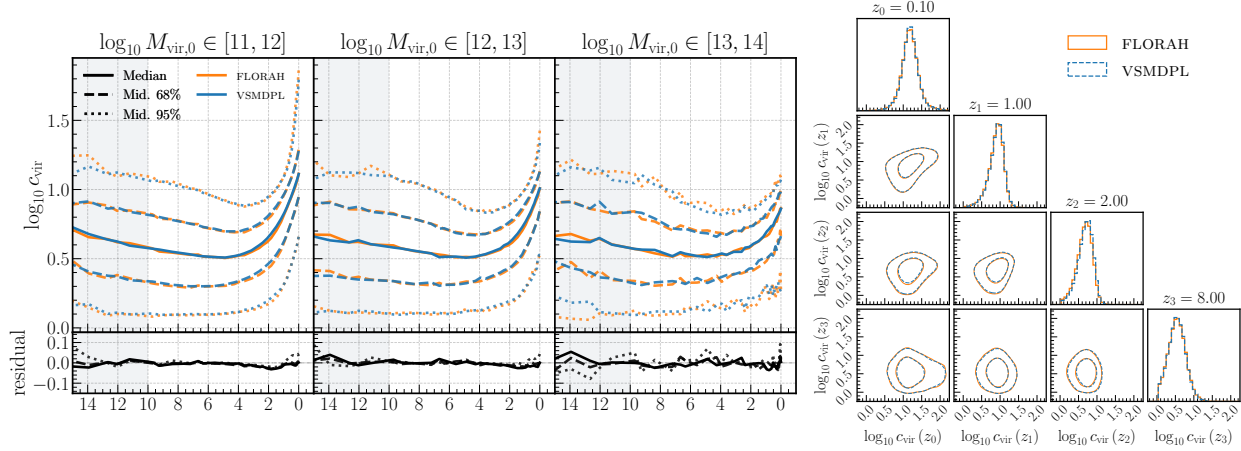
Figure 2. The DM concentration histories in VSMDPL and generated by FLORAH for three mass bins. Panels are the same as Figure 1.
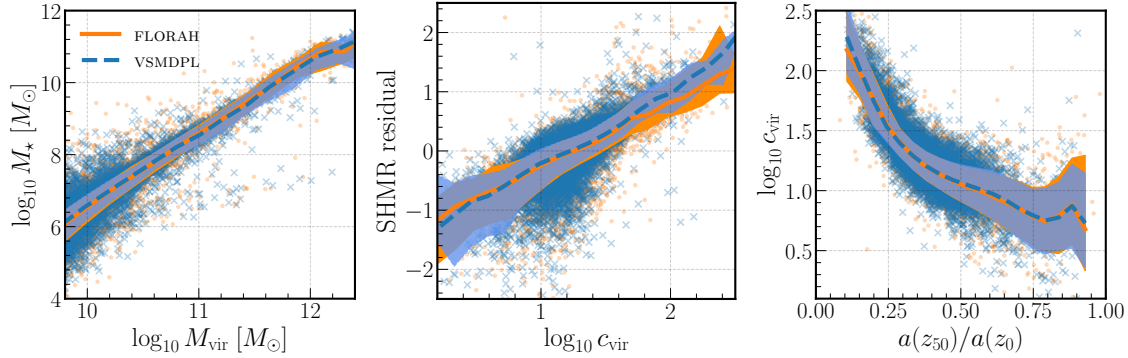

Figure 3. Left: The stellar-to-halo mass relation (SHMR) at $z = 0$ computed by the SC-SAM. Center: The residual of SHMR, defined as the difference between the $M_\star/M_{\rm vir}$ value for each halo and the median value in its corresponding $M_{\rm vir}$ bin, as a function of the DM halo concentration. Right: The relation between the DM concentration $c_{\rm vir}$ and the scale factor $a(z_{50})$ at which a halo forms 50% of its mass.

percentile curves. The MAHs plateau out near the resolution limit $100\,M_{\rm DM}$ (dashed black line, computed from the high end of each mass bin), as expected because we remove all halos below this mass limit. Across all mass bins, FLORAH MAHs exhibit remarkable agreement with the VSMDPL MAHs. Due to the smaller sample size, the MAHs in the most massive bin are noisier. Notably, FLORAH demonstrates the ability to extrapolate well beyond the maximum training redshift ($z_{\rm train}^{\rm max} = 10$, shaded gray box). The right panel of Figure 1 displays the joint distributions of $M_{\rm vir}$ across a few chosen redshifts ($z = 0.1, 1, 2, 8$). The contours (68% and 95%) indicate a good match between the FLORAH MPBs and VSMDPL MPBs, showing that FLORAH can predict the mass correlations across a wide range of redshifts. In Figure 2, we show the $c_{\rm vir}$ histories of VSMDPL and FLORAH MPBs in the three mass bins (left panel) and the joint distributions of $c_{\rm vir}$ across multiple redshifts (right panel). Once again, both the histories and the joint distribution of $c_{\rm vir}$ captured by FLORAH agree well with those of VSMDPL.

We cannot directly observe merger trees, but observable properties of galaxies can be predicted by semi-analytic models (SAMs). SAMs take in merger trees as inputs and solve ordinary differential equations for observables like galaxy luminosities and quasi-observables like stellar mass. To test that FLORAH-generated assembly histories can be used in place of N-body assembly histories in SAMs, we apply the SC-SAM to predict the stellar masses $M_\star$ of galaxies. For a fair comparison with VSMDPL, we input only the MPBs of VSMDPL trees to the SAM. The left panel of Figure 3 shows the stellar-to-halo mass relation (SHMR). Next, we compute the SMHR residual, defined as the difference between the $M_\star/M_{\rm vir}$ value for each halo and its median value in the corresponding $M_{\rm vir}$ bin. The correlation between the SMHR residual and the DM concentration $c_{\rm vir}$ is shown in the middle panel. This demonstrates that the galaxy properties depend on secondary halo characteristics beyond halo mass, which can lead to the phenomenon known as "assembly bias". In addition, we show the relation between $c_{\rm vir}$ and the formation scale factor in the right

panel. In all cases, the relations predicted by FLORAH are consistent with VSMDPL. *A SAM run on FLORAH-generated assembly histories accurately reproduces the correlation between stellar mass, halo formation history, and halo concentration.* Existing EPS-based methods cannot reproduce these correlations.

In summary, we developed FLORAH, a generative model based on recurrent neural networks and normalizing flows to generate the main progenitor branches of merger trees. Trained on merger trees from the VSMDPL simulation, FLORAH can accurately capture the mass and DM concentration histories and reliably extrapolate them beyond the training redshift. We used the SC-SAM to demonstrate that FLORAH MAHs can also recover correlations between galaxy properties and assembly history, such as the correlation between stellar-to-halo mass ratio residual and halo concentration. These correlations cannot be captured with current analytical methods. Currently FLORAH only generates main progenitor branches, and in ongoing work, we are extending it to generate full merger trees. Multiple simulations with different redshift ranges and mass resolutions can also be combined in the training dataset. This will enable the construction of merger trees with a wide dynamic range of mass and temporal resolutions, far beyond the capability of exascale numerical simulations, allowing for the exploration of different galaxy formation scenarios with great computational efficiency at unprecedented accuracy. To develop an emulator for generating merger trees, we will also train FLORAH on simulations with varying cosmological parameters to learn the dependence of assembly histories of halos on cosmology.

Software and Data

This research made use of the `IPython` (Perez & Granger, 2007), `Jupyter` (Kluyver et al., 2016), `Matplotlib` (Hunter, 2007), `NumPy` (Harris et al., 2020), `nflows` (Durkan et al., 2020), `PyTorch` (Paszke et al., 2019), `PyTorch Lightning` (Falcon et al., 2020), `SciPy` (Virtanen et al., 2020), and `ytree` (Smith & Lang, 2019) software packages.

The code version used for this research is available at `https://github.com/trivnguyen/florah`.

Acknowledgements

References

Behroozi, P. S., Wechsler, R. H., and Wu, H.-Y. The ROCK-STAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. , 762(2):109, January 2013a. doi: 10.1088/0004-637X/762/2/109.

Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., Busha, M. T., Klypin, A. A., and Primack, J. R. Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. , 763(1):18, January 2013b. doi: 10.1088/0004-637X/763/1/18.

Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N. Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. , 379:440, October 1991. doi: 10.1086/170520.

Bower, R. G. The evolution of groups of galaxies in the Press-Schechter formalism. , 248:332–352, January 1991. doi: 10.1093/mnras/248.2.332.

Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. nflows: normalizing flows in PyTorch, November 2020. URL https://doi.org/10.5281/zenodo.4296287.

Falcon, W. et al. Pytorchlightning/pytorch-lightning: 0.7.6 release, May 2020. URL https://doi.org/10.5281/zenodo.3828935.

Gabrielpillai, A., Somerville, R. S., Genel, S., Rodriguez-Gomez, V., Pandya, V., Yung, L. Y. A., and Hernquist, L. Galaxy formation in the Santa Cruz semi-analytic model compared with IllustrisTNG – I. Galaxy scaling relations, dispersions, and residuals at z = 0. *Monthly Notices of the Royal Astronomical Society*, 517(4):6091–6111, 08 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac2297. URL https://doi.org/10.1093/mnras/stac2297.

Germain, M., Gregor, K., Murray, I., and Larochelle, H. MADE: masked autoencoder for distribution estimation. In Bach, F. R. and Blei, D. M. (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 881–889, 2015. URL http://proceedings.mlr.press/v37/germain15.html.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.

Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

Kluyver, T. et al. Jupyter notebooks - a publishing format for reproducible computational workflows. In *ELPUB*, 2016.

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., and Heß, S. MultiDark simulations: the story of dark matter halo concentrations and density profiles. *Monthly Notices of the Royal Astronomical Society*, 457(4):4340–4359, 02 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw248. URL https://doi.org/10.1093/mnras/stw248.

Li, Y., Mo, H. J., van den Bosch, F. C., and Lin, W. P. On the assembly history of dark matter haloes. , 379(2):689–701, August 2007. doi: 10.1111/j.1365-2966.2007.11942.x.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Navarro, J. F., Frenk, C. S., and White, S. D. M. The Structure of Cold Dark Matter Halos. , 462:563, May 1996. doi: 10.1086/177173.

Papamakarios, G., Pavlakou, T., and Murray, I. Masked autoregressive flow for density estimation. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 2335–2344, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964. URL https://papers.nips.cc/paper/2017/hash/6c1da886822c67822bcf3679d04369fa-Abstract.html.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 2019.

Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Perez, F. and Granger, B. E. IPython: A System for Interactive Scientific Computing. *Computing in Science and Engineering*, 9(3):21–29, Jan 2007. doi: 10.1109/MCSE.2007.53.

Planck Collaboration. Planck 2013 results. XVI. Cosmological parameters. , 571:A16, November 2014. doi: 10.1051/0004-6361/201321591.

Rezende, D. J. and Mohamed, S. Variational inference with normalizing flows. In Bach, F. R. and Blei, D. M. (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 1530–1538, 2015. URL http://proceedings.mlr.press/v37/rezende15.html.

Smith, B. D. and Lang, M. ytree: A python package for analyzing merger trees. *Journal of Open Source Software*, 4(44):1881, dec 2019. doi: 10.21105/joss.01881. URL https://doi.org/10.21105/joss.01881.

Somerville, R. S. and Davé, R. Physical Models of Galaxy Formation in a Cosmological Framework. , 53:51–113, August 2015. doi: 10.1146/annurev-astro-082812-140951.

Somerville, R. S. and Kolatt, T. S. How to plant a merger tree. *Monthly Notices of the Royal Astronomical Society*, 305(1):1–14, 05 1999. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1999.02154.x. URL https://doi.org/10.1046/j.1365-8711.1999.02154.x.

Somerville, R. S., Popping, G., and Trager, S. C. Star formation in semi-analytic galaxy formation models with multiphase gas. *Monthly Notices of the Royal Astronomical Society*, 453(4):4337–4367, 09 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv1877. URL https://doi.org/10.1093/mnras/stv1877.

Springel, V. The cosmological simulation code gadget-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, 12 2005. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2005.09655.x. URL https://doi.org/10.1111/j.1365-2966.2005.09655.x.

Springel, V., Yoshida, N., and White, S. D. Gadget: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy*, 6(2):79–117, 2001. ISSN 1384-1076. doi: https://doi.org/10.1016/S1384-1076(01)00042-2. URL https://www.sciencedirect.com/science/article/pii/S1384107601000422.

Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 2020. doi: https://doi.org/10.1038/s41592-019-0686-2.

Vogelsberger, M., Marinacci, F., Torrey, P., and Puchwein, E. Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66, January 2020. doi: 10.1038/s42254-019-0127-2.

Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., and Allgood, B. The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation. , 652(1):71–84, November 2006. doi: 10.1086/507120.

White, S. D. M. and Rees, M. J. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. , 183:341–358, May 1978. doi: 10.1093/mnras/183.3.341.

Yung, A., Somerville, R., Finkelstein, S., Popping, G., and Davé, R. Semi-analytic forecasts for JWST - I. UV luminosity functions at z = 4-10. *MNRAS*, 483(3):2983–3006, March 2019. doi: 10.1093/mnras/sty3241.

Zentner, A. R. The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth. *International Journal of Modern Physics D*, 16(5):763–815, January 2007. doi: 10.1142/S0218271807010511.

We show a schematic illustration of FLORAH at time step i in Figure 4. During training, instead of sampling $\hat{p}_\phi$ for the output feature $\hat{y}^{(i+1)}$, we optimize the loss $-\log_{10}\hat{p}_\phi$ (Equation 6).
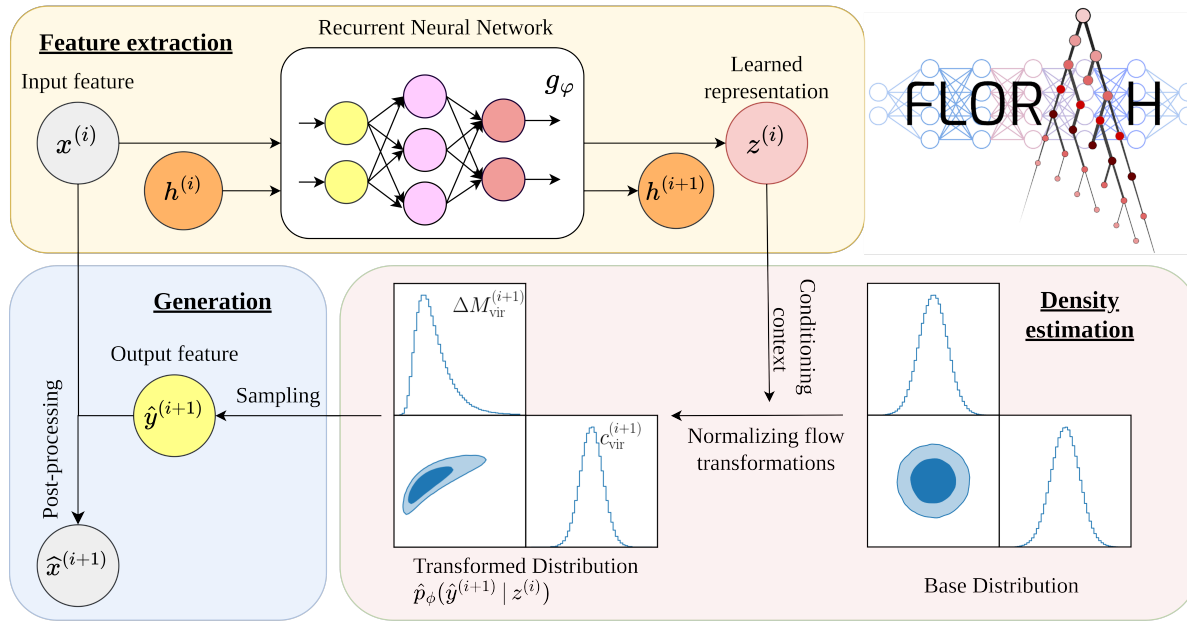

Figure 4. A schematic illustration of FLORAH.