Uncovering Physical Drivers of Dark Matter Halo Structures with AuxiliaryVariableGuided Generative Models

Arkaprabha Ganguli [*]
Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, IL 60439
aganguli@anl.gov

Anirban Samaddar [*]
Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, IL 60439
asamaddar@anl.gov

Florian Kéruzoré
High Energy Physics Division
Argonne National Laboratory
Lemont, IL 60439, USA
fkeruzore@anl.gov

Nesar Ramachandra
Computational Science Division
Argonne National Laboratory
Lemont, IL 60439
nramachandra@anl.gov

Julie Bessac
Computational Science Center
Data, Analysis and Visualization Group
National Renewable Energy Laboratory
Golden, CO, USA
julie.bessac@nrel.gov

Sandeep Madireddy
Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, IL 60439
smadireddy@anl.gov

Emil Constantinescu
Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, IL 60439
emconsta@anl.gov

Abstract

Deep generative models (DGMs) compress high-dimensional data but often entangle distinct physical factors in their latent spaces. We present an auxiliary-variable-guided framework for disentangling representations of thermal SunyaevZeldovich (tSZ) maps of dark matter halos. We introduce halo mass and concentration as auxiliary variables and apply a lightweight alignment penalty to encourage latent dimensions to reflect these physical quantities. To generate sharp and realistic samples, we extend latent conditional flow matching (LCFM), a state-of-the-art generative model, to enforce disentanglement in the latent space. Our **D**isentangled **L**atent-**CFM** (DL-CFM) model recovers the established mass-concentration scaling relation and identifies latent space outliers that may correspond to unusual halo formation histories. By linking latent coordinates to interpretable astrophysical properties, our method transforms the latent space into a diagnostic tool for cosmological structure. This work demonstrates that auxiliary guidance preserves generative flexibility while yielding physically meaningful, disentangled embeddings, providing a generalizable pathway for uncovering independent factors in complex astronomical datasets.

[*] equal contribution

Machine Learning and the Physical Sciences Workshop, NeurIPS 2025.

1 Introduction

Deep generative models (DGMs)including variational autoencoders (VAEs), normalizing flows, and diffusion models are indispensable for modeling complex, high-dimensional scientific data. However, when applied to scientific datasets, heterogeneous in modality, fidelity, and accuracy, with stochastic measurements and multi-scale structure, DGMs often lack interpretability [Yang et al., 2021]. Domain scientists seek to characterize patterns and associations among physical quantities for prediction, uncertainty quantification (UQ), and mechanistic understanding, but they often have only *partial knowledge* of these physical quantities: some factors are measured and their link to the data distribution is known ("known knowns"), others are measured but their influence is uncertain ("known unknowns"), and many remain unanticipated ("unknown unknowns") Hatfield [2022]. In this setting, DGMs frequently learn *entangled* latent spaces, where a single coordinate influences multiple unrelated aspects of the data, thereby hindering interpretability and downstream applications (e.g., sensitivity analysis, inverse design, hypothesis testing). *Disentangled* representations instead aim for latent coordinates that correspond to independent factors of variation, so that adjusting one coordinate affects only its associated factor [Wang et al., 2023]. Astronomical imaging provides a compelling testbed; we focus on dark matter halos observed via the thermal Sunyaev–Zel'dovich effect (tSZ, Zeldovich and Sunyaev [1969]), identified in maps of the cosmic microwave background (CMB), where images exhibit rich structure tied to physically meaningful, computable auxiliary information, e.g., halo mass and concentration. These auxiliary variables can be computed for each halo map and paired with the image. However, existing unsupervised approaches for disentanglement (e.g., β-VAE [Higgins et al., 2017], FactorVAE [Kim and Mnih, 2018], DIP-VAE [Kumar et al., 2018]) only encourage factorization in the latent space, but do not exploit partially known physical covariates like these mass and concentration. When such covariates are available and scientifically important, a pragmatic alternative is to *guide* disentanglement with auxiliary variables, softly steering selected latent coordinates to align with target factors while allowing the remaining latents to capture residual variability.

Recent work [Ganguli et al., 2025] proposed *AuxVAE* which has instantiated this idea within a VAE by partitioning the latent space into two segments - an auxiliary-informed block and a residual block, and adding lightweight penalties that (i) align each auxiliary-informed dimension with its corresponding physical variable and (ii) discourage cross-correlation with other latents, while leaving the residual block free to model remaining variation. This improves interpretability without requiring full supervision over all factors. However, across rich, high-detail datasets, including the present application to tSZ halo maps, standard VAEs often over-smooth fine structure, under-represent small-scale variability, and lag behind more expressive generative models in sample fidelity and generalization [Yacoby et al., 2020, Bozkurt et al., 2021].

To overcome these limitations, we turn to *conditional flow matching* (CFM), a class of powerful flow-based DGMs that learn sample-generation by regressing probability flow (or transport) vector fields between simple reference distributions and data distributions [Lipman et al., 2022, Tong et al., 2023]. CFM inherits the benefits of sharpness and diversity from continuous normalizing flows, while enjoying stable and scalable training via supervised regression of vector fields. Recent approaches in flow matching Guo and Schwing [2025], Samaddar et al. [2025] have adapted the deep latent variable models (such as VAEs) to the flow matching for structured generation, efficient training, and accurate inference. In this paper, we propose *Disentangled Latent-Conditional Flow Matching* (DL-CFM), that marries the interpretability of auxiliary-guided VAEs with the fidelity of CFM. Concretely, we first infer a low-dimensional code z with a VAE encoder and impose the same auxiliary alignment regularizers used in Ganguli et al. [2025] so that selected coordinates of z correspond to known physical factors (e.g. halo mass and concentration). We then train a CFM-based generator conditioned on z to produce high-resolution tSZ maps. In effect, the VAE encoder provides a structured, interpretable bottleneck, and the flow-matching decoder renders those factors into realistic, high-detail images. In summary, this paper makes the following contributions:

- **Disentanglement in flow matching.** We introduce DL-CFM, bringing auxiliary-variable guidance into conditional flow matching via a lightweight VAE encoder with simple alignment/decoupling losses. To our knowledge, this is the first approach to enable disentangled control within CFM without degrading fidelity.

- **Application in tSZ map generation and control.** On simulated tSZ maps of halos, DL-CFM learns accurate data distribution and generates realistic maps of diverse samples with interpretable control along mass and concentration. Guided traversals allow targeted synthesis at specified settings and separate known factors from residual morphology.

- **Scientific validation and diagnostics.** The learned latents recover the expected mass-concentration trend and surface outliers (e.g., disturbed systems or artifacts), enabling sensitivity analyses and anomaly discovery with a compact, interpretable representation, as well as fast generation of realistic mock datasets from minimal inputs.

Next, we describe our approach in Section 2 and present experimental results in Section 3.

2 Methodology

We use auxiliary physical variables to shape the latent space of a deep generative model so that selected coordinates align with known factors (e.g., halo mass and concentration) while preserving generative flexibility. Concretely, we propose `DL-CFM`, which adapts the AuxVAE loss function to the state-of-the-art latent conditional flow matching model for high-quality sample generation and physics-aware latent space disentanglement. This section specifies the setting and the *main loss terms* needed to reproduce our method; extended definitions and derivations are deferred to the Appendix.

2.1 Notations

Let $x \in \mathbb{R}^p$ be an observation (a tSZ halo image) and $u \in \mathbb{R}^d$ the auxiliary variables (here, halo mass and concentration). The VAE introduces a latent $z \in \mathbb{R}^{dz}$ and partitions it as

$$z = \big(z_{\text{aux}},\, z_{\text{rec}}\big), \qquad z_{\text{aux}} \in \mathbb{R}^d \text{ (auxiliary-guided)}, \quad z_{\text{rec}} \in \mathbb{R}^{dz-d} \text{ (reconstruction-focused)}.$$

To align z_{aux} with u while leaving z_{rec} free to capture remaining unknown unknowns, we use an *auxiliary-informed prior*

$$p(z \mid u) = \mathcal{N}(\mu_0(u), \Sigma_0)\,, \quad \mu_0(u) = \big(u_1, \ldots, u_d, 0, \ldots, 0\big), \quad \Sigma_0 = \operatorname{diag}\big(\tau^2 \mathbf{I}_d,\, \mathbf{I}_{dz-d}\big)\,, \tag{1}$$

where $\tau^2 \ll 1$ is a small variance, generally taken as inverse of the batch-size, that softly tethers the guided coordinates to u (we normalize u to $[0,1]$).

2.2 Auxiliary-variable guided disentangled Latent CFM

Recent approaches in flow matching Guo and Schwing [2025], Samaddar et al. [2025] have combined deep latent variable models with flow-based generative models to ensure efficient training and inference. We propose `DL-CFM` that extends these state-of-the-art flow matching approaches to enforce disentanglement in the latent space using the auxiliary variables in our data.

We aim to learn a time-dependent vector field $v_\theta(x_t, z, t)$ that evolves samples from a simple source distribution, x_0, to the high-dimensional halo dataset conditioned on the disentangled latent variable z. To enforce the disentanglement, we propose the loss function,

$$\begin{aligned}\mathcal{L}_{\texttt{DL-CFM}} = {} & \underbrace{\mathbb{E}_{p(t), q_\phi(z|x), p_t(x_t|x_0,x)}\, ||v_\theta(x_t, z, t) - u_t(x_t|x_0, x_1)||_2^2}_{\text{conditional flow matching loss}} + \beta\, \underbrace{\mathrm{KL}(q_\phi(z \mid x) \,\|\, p(z \mid u))}_{\text{conditional prior match}} \\ & + \lambda_1 \sum_{j=1}^{d} \Big(\underbrace{\mathsf{Align}(u_j, \mu_{\phi,\text{aux},j})}_{\text{explicitness}} + \underbrace{\mathsf{Decorr}\big(u_j, \mu_{\phi,\text{aux},-j}\big)}_{\text{intra-independence}} \Big) + \lambda_2\, \underbrace{\mathsf{Decorr}\big(u, \mu_{\phi,\text{rec}}\big)}_{\text{inter-independence}}\,.\end{aligned} \tag{2}$$

Here $\mu_\phi(\cdot)$ denotes the encoder mean; $\mu_{\phi,\text{aux}}$ and $\mu_{\phi,\text{rec}}$ are its restrictions to the auxiliary-guided and reconstruction-focused coordinates, and $\mu_{\phi,\text{aux},-j}$ excludes the j^{th} guided coordinate. The three regularizers enforce: (i) *explicitness* (guided coordinate j tracks u_j in a one-to-one manner), (ii) *intra-independence* (no cross-correlation among guided latents), and (iii) *inter-independence* (reconstruction-focused latents are decorrelated from u). We instantiate Align and Decorr as lightweight correlation-based penalties computed from minibatch statistics of μ_ϕ. We leave the details of training and sampling for `DL-CFM` in App. B. Equation (2) introduces no extra networks and only a few scalar hyperparameters $(\beta, \lambda_1, \lambda_2, \tau^2)$.

3 Experimental results

We evaluate (i) *generation quality* and (ii) *disentanglement effects* on synthetic thermal Sunyaev–Zel'dovich (tSZ) halo images. Our model of interest is `DL-CFM`, which couples an auxiliary-guided VAE bottleneck to a conditional flow matching model. We compare `DL-CFM` to a state-of-the-art baseline ICFM Tong et al. [2023] model in terms of generation quality.

Methods	# Params.	Sinkhorn (↓)	Energy (↓)	Gaussian (↓)	Laplacian (↓)
ICFM	34.42M	**4564.03** ± 37.883	84.073 ± 0.797	**0.00813** ± 0.00020	0.00693 ± 0.00011
DL-CFM	38.06M	4819.211 ± 32.718	**83.148** ± 0.767	**0.00813** ± 0.00014	**0.00678** ± 0.00012

Table 1: Table shows the generation quality for ICFM and DL-CFM in terms of different distance metrics (mean ± Sd). In terms of most metrics, the two approaches show similar generation quality.

3.1 Data and experimental setup

Simulations and halos. We use cosmological hydrodynamic simulations from de Souza Vitório et al. [2025] run with the Conservative Reproducing Kernel, Hybrid/Hardware Accelerated Cosmology Code (CRK-HACC; Frontiere et al. [2023]); details are in App. A. Halos are identified with a friends-of-friends finder, and halo centers are set to the most bound dark matter particle. For each halo, we compute two physical properties that are expected to capture key intracluster medium (ICM) morphology Kéruzoré et al. [2024]: mass M_{200c} and concentration c_{200c}. We select halos with $M_{200c} > 10^{13.5}\, h^{-1} M_{\odot}$. Details on the halo catalog and associated properties are detailed in Appendix A.1. We represent each halo with a Compton-y mapthe line-of-sight integral of the electron pressureof the thermal SunyaevZeldovich (tSZ) signal. Each image is paired with its (M_{200c}, c_{200c}) values. Our experiments test whether DL-CFM matches the sample fidelity of CFM *while enabling controlled generation* along the mass and concentration axes.

3.2 Generation quality

Table 1 shows different distance metrics (mean ± sd) calculated between the training samples and the generated samples from ICFM and DL-CFM model trained on the tSZ halo data set. We observe that, in terms of most distance metrics, both approaches show similar generation quality. DL-CFM performs marginally better in terms of the Energy metric. Both methods show a large Sinkhorn distance, with ICFM showing a lower distance than DL-CFM.

3.3 Disentanglement effects

We test whether the guided latents isolate the intended auxiliary information and support controlled generation. **(a) Latentauxiliary alignment:** We scatter $\{\log M_{200c},\, c_{200c}\}$ against the first five latent coordinates: the first two are auxiliary-guided (z_{aux}) and the next three are representative reconstruction-focused (z_{rec}). DL-CFM shows near one-to-one, monotonic relationships on the guided axes and weak correlations elsewhere, consistent with the training objective; the generated massconcentration trend matches the simulation catalog (Fig. 1). **(b) Controlled traversals.** We traverse each guided dimension in z_{aux} while holding z_{rec} fixed (Fig. 2). Samples vary systematically and interpretably along mass and concentration, enabling targeted generation of tSZ halos at desired (M_{200c}, c_{200c}) without sacrificing fidelity.

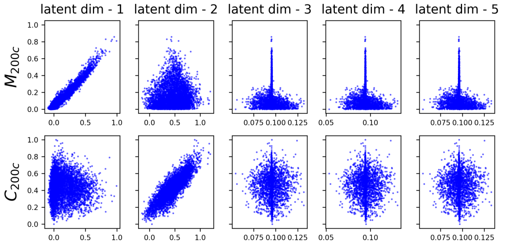

Figure 1: Alignment of guided latents with mass and concentration.

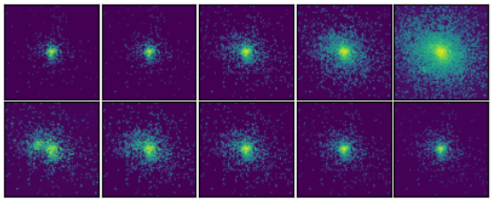

Figure 2: Traversals along guided latents (z_{aux}) with z_{rec} fixed. Rows: mass, concentration.

(c) Conditional distributional shifts. We illustrate controlled generation by fixing the auxiliary-guided latents to a low-mass, low-concentration setting, $(z_1, z_2) = (0.001, 0.001)$ (with z_1 and z_2 aligned to mass and concentration), and sampling only the reconstruction-focused latents z_{rec}. Figure 3 shows generated tSZ halos from two regimes: the *center* of the z_{rec} distribution (top) and its *tails* (bottom). Center samples appear relaxed and single-peaked, whereas tail samples exhibit complex, multi-peaked morphologies indicative of disturbed systems or active merger status. This conditional shift demonstrates that z_{rec} captures residual structure beyond (M_{200c}, c_{200c}) and enables targeted sampling for sensitivity analyses and discoverye.g., identifying cases where center-based concentration estimates under-represent complex, multi-core systems. Additional settings (high-mass/low-concentration and low-mass/high-concentration) are shown in Appendix C.

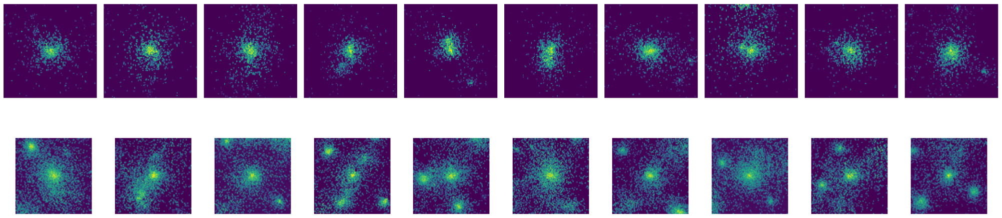

Figure 3: Generating samples from the center (top) and tail (bottom) of the reconstruction-focused latents z_{rec}, with the first two auxiliary-guided coordinates fixed at $(z_1, z_2) = (0.001, 0.001)$.

4 Discussion

`DL-CFM` couples auxiliary-guided latents with conditional flow matching to synthesize tSZ halo maps with high fidelity and interpretable control. Simple alignment and decorrelation penalties expose known factors while allowing residual latents to capture remaining unknown variability; these latents generate diverse structures under fixed auxiliary settings. Overall, auxiliary-guided flows offer a compact route to uniting interpretability with state-of-the-art generative quality. Applying the method to real data will require careful selection and calibration of auxiliary variables and explicit treatment of instrumental or systematic effects.

Acknowledgments and Disclosure of Funding

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) and the Scientific Discovery through Advanced Computing (SciDAC) Program Nuclear Physics partnership titled “Femtoscale Imaging of Nuclei using Exascale Platforms” and the SciDAC-RAPIDS institute, FASTMath Institute programs under Contract No. DE-AC02-06CH11357. Work at Argonne National Laboratory is supported by UChicago Argonne LLC, Operator of Argonne National Laboratory. Argonne, a U.S. Department of Energy Office of Science Laboratory, is operated under Contract No. DE-AC02-06CH11357. The training is carried out on Swing, a GPU system at the Laboratory Computing Resource Center (LCRC) of Argonne National Laboratory.

References

Shuo Yang, Tianyu Guo, Yunhe Wang, and Chang Xu. Adversarial robustness through disentangled representations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4):3145–3153, May 2021. doi: 10.1609/aaai.v35i4.16424. URL `https://ojs.aaai.org/index.php/AAAI/article/view/16424`.

Peter Hatfield. Quantification of Unknown Unknowns in Astronomy and Physics. *arXiv e-prints*, art. arXiv:2207.13993, July 2022. doi: 10.48550/arXiv.2207.13993.

Xin Wang, Hong Chen, Si’ao Tang, Zihao Wu, and Wenwu Zhu. Disentangled representation learning, 2023.

Ya. B. Zeldovich and R. A. Sunyaev. The Interaction of Matter and Radiation in a Hot-Model Universe. *Astrophys. Space Sci.*, 4:301–316, 1969. doi: 10.1007/BF00661821.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017. URL `https://openreview.net/forum?id=Sy2fzU9gl`.

Hyunjik Kim and Andriy Mnih. Disentangling by factorising. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2649–2658. PMLR, 10–15 Jul 2018. URL `https://proceedings.mlr.press/v80/kim18b.html`.

Abhishek Kumar, Prasanna Sattigeri, and Avinash Balakrishnan. Variational inference of disentangled latent concepts from unlabeled observations. In *6th International Conference on Learn-*

ing Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net, 2018. URL https://openreview.net/forum?id=H1kG7GZAW.

Arkaprabha Ganguli, Nesar Ramachandra, Julie Bessac, and Emil Constantinescu. Enhancing interpretability in generative modeling: statistically disentangled latent spaces guided by generative factors in scientific datasets. *Machine Learning*, 114(9):197, 2025.

Yaniv Yacoby, Weiwei Pan, and Finale Doshi-Velez. Failure modes of variational autoencoders and their effects on downstream tasks. *arXiv preprint arXiv:2007.07124*, 2020.

Alican Bozkurt, Babak Esmaeili, Jean-Baptiste Tristan, Dana Brooks, Jennifer Dy, and Jan-Willem van de Meent. Rate-regularization and generalization in variational autoencoders. In *International Conference on Artificial Intelligence and Statistics*, pages 3880–3888. PMLR, 2021.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

Pengsheng Guo and Alexander G Schwing. Variational rectified flow matching. *arXiv preprint arXiv:2502.09616*, 2025.

Anirban Samaddar, Yixuan Sun, Viktor Nilsson, and Sandeep Madireddy. Efficient flow matching using latent variables, 2025. URL https://arxiv.org/abs/2505.04486.

Isabele Lais de Souza Vitório, Michael Buehlmann, Eve Kovacs, Patricia Larsen, Nicholas Frontiere, and Katrin Heitmann. Exploring the Core-galaxy Connection. *The Open Journal of Astrophysics*, 8:82, June 2025. doi: 10.33232/001c.141464.

Nicholas Frontiere, J. D. Emberson, Michael Buehlmann, Joseph Adamo, Salman Habib, Katrin Heitmann, and Claude-André Faucher-Giguère. Simulating Hydrodynamics in Cosmology with CRK-HACC. ApJS, 264(2):34, February 2023. doi: 10.3847/1538-4365/aca58d.

Florian Kéruzoré, L. E. Bleem, N. Frontiere, N. Krishnan, M. Buehlmann, J. D. Emberson, S. Habib, and P. Larsen. The picasso gas model: Painting intracluster gas on gravity-only simulations. *The Open Journal of Astrophysics*, 7:116, December 2024. doi: 10.33232/001c.127486.

Salman Habib, Adrian Pope, Hal Finkel, Nicholas Frontiere, Katrin Heitmann, David Daniel, Patricia Fasel, Vitali Morozov, George Zagaris, Tom Peterka, Venkatram Vishwanath, Zarija Lukić, Saba Sehrish, and Wei-keng Liao. HACC: Simulating sky surveys on state-of-the-art supercomputing architectures. New A, 42:49–65, January 2016. doi: 10.1016/j.newast.2015.06.003.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. Planck 2018 results. VI. Cosmological parameters. A&A, 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

Hillary L. Child, Salman Habib, Katrin Heitmann, Nicholas Frontiere, Hal Finkel, Adrian Pope, and Vitali Morozov. Halo Profiles and the Concentration-Mass Relation for a ΛCDM Universe. ApJ, 859(1):55, May 2018. doi: 10.3847/1538-4357/aabf95.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022. URL https://arxiv.org/abs/1312.6114.

A Data

To train the proposed DL-CFM model, we use synthetic images produced from a cosmological hydrodynamic simulation. The set of simulations we use is described in detail in de Souza Vitório et al. [2025]; in this section, we briefly summarize their main properties and the generation of the images.

A.1 Simulations

The simulation set was generated using the Hybrid/hardware Accelerated Cosmology Code (HACC, Habib et al. [2016]); more specifically, we focus on simulations produced using the CRK-HACC hydrodynamic solver described in Frontiere et al. [2023]. Initial conditions are generated shortly after the Big Bang, at redshift $z = 200$, using the best-fit cosmology derived from the *Planck* analysis of the cosmic microwave background (Planck Collaboration et al. [2020]). The simulation spans a volume of $(576\ h^{-1}\mathrm{Mpc})^3$ (comoving) and simulates the evolution of $2 \times (2304)^3$ particles (half dark matter, half gas) all the way to present day, at $z = 0$. We specifically focus on the "non-radiative" (or "adiabatic") version of the hydrodynamic simulation (de Souza Vitório et al. [2025]), in which all baryonic matter is modeled as gas interacting through hydrodynamic equations, and no sub-resolution physics are included.

A.2 Halo catalog

In this work, we focus on the very last stage of the simulation (present day, $z = 0$). A friends-of-friends (FoF) algorithm is run on the dark matter particles of the simulation to identify halos. For each halo found by the FoF finder, the center of the halo is defined as the position of the most bound dark matter particle, and the entire matter distribution (both gas and dark matter) around this position are used to compute halo properties.

In this work, we focus on a small subset of halo properties, expected to contain most of the information about a halo and its gas distribution Kéruzoré et al. [2024]:

1. Halo mass M_{200c}, computed as the mass enclosed within a characteristic radius, R_{200c}, within which the average halo matter density is 200 times greater than the critical density of the Universe;
2. Halo concentration c_{200c}, quantifying how much of the halo matter is contained in the center of the halo as opposed to its outskirts;
3. Distance between the halo potential peak and center of mass, $\Delta x / R_{200c}$, known to be an indicator of the disturbed state of halos (specifically of their merging status, see *e.g.* Child et al. [2018];
4. Halo ellipticity, e, and prolaticity, p, quantifying the overall triaxial shape of the halo.

A.3 Image generation

The thermal Sunyaez-Zel'dovich is a spectral distortion of the Cosmic Microwave Background due to its interaction with free electrons, in particular in the hot plasma forming the intracluster medium (ICM) of massive halos. The amplitude of this distortion is given by the Compton-y parameter, proportional to the line-of-sight (LoS) integral of the electron pressure P_e in the ICM:

$$y = \frac{\sigma_\mathrm{T}}{m_\mathrm{e} c^2} \int_\mathrm{LoS} P_\mathrm{e}\, \mathrm{d}l, \tag{3}$$

where σ_T is the Thompson scattering cross-section, and $m_\mathrm{e} c^2$ is the electron rest-frame energy.

The images used in this work are maps of the Compton-y parameter in massive ($M_{200c} > 10^{13.5}\, h^{-1} M_\odot$) dark matter halos. For each halo, the gas density and temperature is projected on a 3D grid using the cloud-in-cell algorithm. The grid is $(64)^3$ cells, with a box size of $(4 \times R_{200c})^3$. The gas pressure is computed as the product of density and pressure, and converted to electron pressure using the ratio of the mean molecular weight of the fully ionized gas and of electrons. The resulting 3D pressure distribution is then integrated along three orthogonal lines of sight using eq. 3, resulting in three (64×64) Compton-y images per halo.

B DL-CFM details and derivations

B.1 Flow matching details

Flow matching attempts to transport samples x_0 from a simple source distribution with density p_0 to a complex data distribution with density p_1. This is done through a time-dependent vector field $u_t : [0, 1] \times \mathbb{R}^p \to \mathbb{R}^p$ which defines an ordinary differential equation,

$$\frac{d\phi_t(x)}{dt} = u_t(\phi_t(x)); \phi_0(x) = x_0 \tag{4}$$

where $\phi_t(x) = x_t$ is the solution of the ODE or *flow* with the initial condition in Eq 4. The primary objective of flow matching models is to train a neural network $v_\theta(., t)$ to learn the ground-truth vector field u_t. However, the true vector field is unknown for most real-world datasets. Therefore, the common approach is to fix the conditional vector field $u_t(.|x_0, x)$ conditioned on the sample pairs $(x_0, x) \sim q(x_0, x)$, where q is a joint distribution with marginals p_0 and p_1. Following the ICFM Tong et al. [2023], we fix $u_t(.|x_0, x) = x - x_0$ and $q = p_0 \times p_1$ and $p_0 = N(0, I)$ as the source distribution. This leads to the conditional flow matching objective,

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{p(t),q(x_0,x),p_t(x_t|x_0,x)} \left[||v_\theta(x_t, t) - u_t(x_t|x_0, x_1)||_2^2 \right] \tag{5}$$

where $p_t(.|x_0, x)$ is the probability path which we fix to be $N(.; tx + (1-t)x_0, \sigma^2 I_p)$. Following Theorem 2.1 in Tong et al. [2023], one can show that the fixed vector field $u_t(.|x_0, x) = x - x_0$ induces this probability path. With $p(t) = unif(0, 1)$, Eq. 5 is a tractable objective which can be minimized with respect to the neural network parameters θ.

Recent approaches in Samaddar et al. [2025] have extended the CFM model to incorporate data-driven latent structures. The authors model the data as a latent mixture model governed by the latent variable z, $p_1 = \int q(z)q(.|z)dz$. The latent variables are learned from the data using a pretrained latent variable model. In this work, we learn the latent variables from the data using a lightweight encoder model. Instead of pretraining, we train the encoder along with the learned vector field parameters θ, maximizing the `DL-CFM` loss in Eq. 2.

B.2 ELBO with conditional prior and closed-form KL

With a Gaussian encoder $q_\phi(z \mid x) = \mathcal{N}\big(\mu_\phi(x), \Sigma_\phi(x)\big)$ and the prior in (1), the per-sample KL term admits the closed form

$$\text{KL}(q_\phi(z \mid x) \,\|\, p(z \mid u)) = \tfrac{1}{2}\left[\log \frac{|\Sigma_0|}{|\Sigma_\phi|} - d_Z + \big(\mu_\phi - \mu_0\big)^\top \Sigma_0^{-1}\big(\mu_\phi - \mu_0\big) + \text{tr}\big(\Sigma_0^{-1}\Sigma_\phi\big)\right]. \tag{6}$$

The reconstruction term is Monte Carloestimated via samples from $q_\phi(z \mid x)$.

B.3 Why regulate the expected variational posterior

Let $q_\phi(z) = \int q_\phi(z \mid x)p(x)\,dx$ and $p_\theta(z) = \int p_\theta(z \mid x)p(x)\,dx$. By convexity of KL [Kumar et al., 2018],

$$\text{KL}\big(q_\phi(z) \,\|\, p_\theta(z)\big) = \text{KL}\Big(\mathbb{E}_{p(x)} q_\phi(z \mid x) \,\Big\|\, \mathbb{E}_{p(x)} p_\theta(z \mid x)\Big) \;\le\; \mathbb{E}_{p(x)}\text{KL}\big(q_\phi(z \mid x) \,\|\, p_\theta(z \mid x)\big). \tag{7}$$

Maximizing the standard ELBO decreases the RHS of (7) but may still leave residual dependencies in $q_\phi(z)$. Our correlation regularizers directly target these dependencies at the *population* level using minibatch estimates.

B.4 Correlation-based regularizers

Let $v \in \mathbb{R}^{m_v}$ and $w \in \mathbb{R}^{m_w}$. Define

$$\Sigma_{vw} = \mathbb{E}\big[(v - \mathbb{E}v)(w - \mathbb{E}w)^\top\big], \qquad \text{Corr}(v, w) = \text{diag}(\Sigma_{vv})^{-\frac{1}{2}}\,\Sigma_{vw}\,\text{diag}(\Sigma_{ww})^{-\frac{1}{2}}.$$

To capture nonlinear relations we use polynomial lifts up to degree K (applied elementwise), and aggregate:

$$R_0^K(v, w) = \frac{1}{K\,m_v m_w} \sum_{k \neq k'} \sum_{i=1}^{m_v} \sum_{j=1}^{m_w} \left|\text{Corr}\Big(v^k, w^{k'}\Big)_{ij}\right|, \tag{8}$$

$$R_1^K(v, w) = \frac{1}{K\,m_v m_w} \sum_{k \neq k'} \sum_{i=1}^{m_v} \Big(1 - \left|\text{Corr}\Big(v^k, w^{k'}\Big)_{ii}\right|\Big). \tag{9}$$

Intuition: R_0 penalizes generic cross-dependence (off-diagonals), while R_1 rewards one-to-one alignment (diagonals close to ± 1).

Surrogate with encoder means. Let $\mu_\phi(x) = \mathbb{E}[z \mid x]$. For polynomials k, k', by the law of total variance,

$$\text{Cov}\big(u^k, z_{\text{rec}}^{k'}\big) = \mathbb{E}_{(x,u)}\Big[\underbrace{\text{Cov}\big(u^k, z_{\text{rec}}^{k'} \mid x, u\big)}_{=0}\Big] + \text{Cov}_{(x,u)}\Big(u^k,\, \mathbb{E}[z_{\text{rec}}^{k'} \mid x]\Big) = \text{Cov}_{(x,u)}\big(u^k,\, \mu_{\phi,\text{rec}}^{k'}\big), \tag{10}$$

and analogously for the guided block. Hence, batch correlations of μ_ϕ suffice to estimate dependencies.

B.5 Instantiating Align and Decorr

We use

$$\mathsf{Align}(u_j, \mu_{\phi,\mathrm{aux},j}) \;=\; 1 - R_1^K(u_j, \mu_{\phi,\mathrm{aux},j}), \qquad \mathsf{Decorr}(a, b) \;=\; R_0^K(a, b),$$

which produces the main loss in (2). Both terms are scale-free and computed from standardized batch statistics; $K \in \{1, 2\}$ works well in practice.

B.6 Training and inference algorithms

Algorithm A.1 shows the training steps of `DL-CFM`. Given n images, the regularizers, and the initialized networks, we draw a single sample z_i from the encoder distribution $q_\phi(.|x_i)$ for each x_i. These are concatenated with the noisy samples $x_i^{t_i}$ and the noise level t_i and passed through the vector field network $v_\theta(.,.,.)$. The latent variables, the output of the vector field network, and the true conditional vector field target are used in the disentangled loss in Eq. 2, which is optimized with respect to the parameters θ, ϕ.

Algorithm A.1 `DL-CFM` training

```
Given $n$ sample $(x_1, ..., x_n)$ from $p_1(.)$, regularizers $\beta, \lambda_1, \lambda_2$
Initialize $v_\theta(\cdot, \cdot, \cdot)$ and encoder layer parameters $\phi$
for $k$ steps do
    Sample latent variables $z_i \sim q_\phi(.|x_i)$ for all $i = 1, ..., n$
    Sample $(x_1^0, ..., x_n^0)$ from $\mathcal{N}(0, I)$ and noise levels $(t_1, ..., t_n)$ from $Unif(0, 1)$ and compute $(u_{t_1}(.|x_0, x), ..., u_{t_n}(.|x_0, x))$
    compute $v_\theta(x_i^{t_i}, z_i, t_i)$ where $x_i^{t_i}$ is the corrupted $i$-th data at noise level $t_i$
    Compute $\nabla \mathcal{L}_{\texttt{DL-CFM}}$ and update $\theta, \phi$
end for
return $v_{\hat{\theta}}(\cdot, \cdot, \cdot), q_{\hat{\phi}}(.|x)$
```

Algorithm A.2 shows the inference procedure for `DL-CFM`. Following Samaddar et al. [2025], we reuse the training data to draw samples from `DL-CFM`. Given the budget of K samples, a random batch of K training data is sampled, then passed through the encoder to draw samples from the latent space. For each latent sample z_i, we iteratively solve the ODE in Eq. 4 for h steps. Fig. A.1 shows the schematic of the `DL-CFM` inference procedure. Note that the latent sampling is performed once during the inference and fixed for h ODE denoising steps.

Algorithm A.2 `DL-CFM` inference

```
Given sample size $K$, trained $v_{\hat{\theta}}(.,.,.)$ and $q_{\hat{\phi}}(.|x)$, number of ODE steps $n_{ode}$
Select $K$ training samples $(x_1^{train}, ..., x_K^{train})$
Sample latent variables $z_i \sim q_{\hat{\phi}}(.|x_i^{train})$ for all $i = 1, ..., K$
Sample $(x_1^0, ..., x_n^0)$ from $\mathcal{N}(0, I)$
$h \leftarrow \frac{1}{n_{ode}}$
for $t = 0, h, ..., 1 - h$ and $i = 1, ..., K$ do
    $x_i^{t+h}$ = ODEstep($v_{\hat{\theta}}(x_i^t, z_i, t), x_i^t$)
end for
return Samples $(x_1, ..., x_K)$
```

B.7 Implementation details

The models trained have the same U-Net architecture from Tong et al. [2023]. For I-CFM, the model takes the input (x_t, t), the variables are projected onto an embedding space and concatenated along the channel dimension, then passed through the U-Net layers to output the learned vector field.

In `DL-CFM`, we use a deep convolutional neural network as the encoder network. The network consists of four convolutional downsampling blocks, where each convolutional layer is followed by a batch normalization and a leaky ReLU activation. The output is then passed through two dilated convolution blocks with batch normalization and leaky ReLU activation. The output from the convolutional encoder is flattened and passed through a linear layer to predict the mean and the log-variance of the latent space. Using the reparameterization trick Kingma and Welling [2022], we

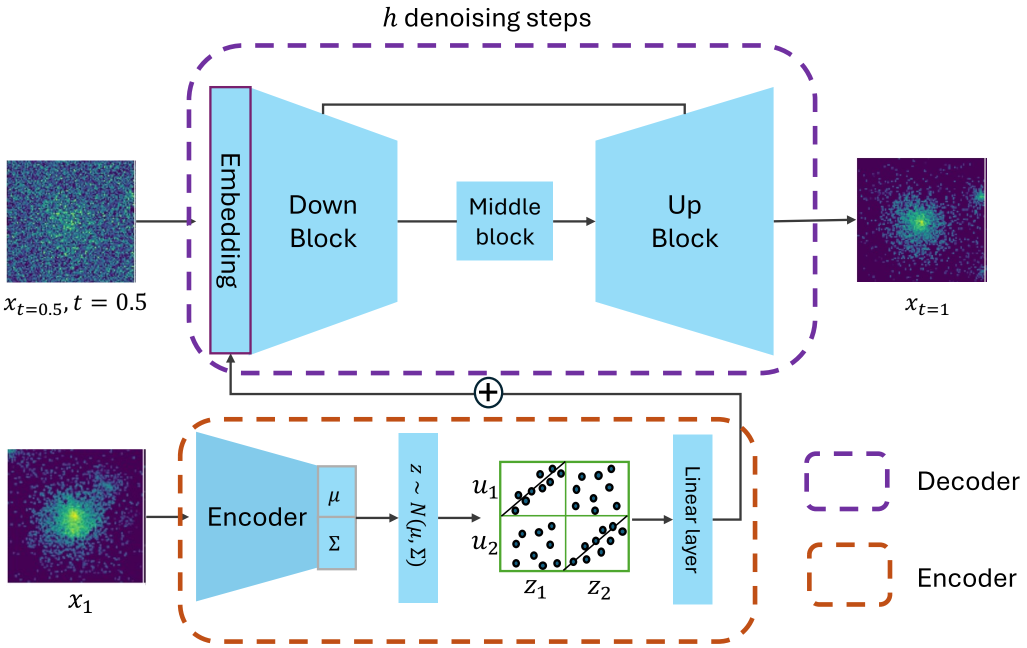

Figure A.1: Schematic of the DL-CFM inference. Given a training sample, we fix the sampled latent from the disentangled latent space. The latent variable is used in the vector field network to evolve the source samples to the data distribution. For demonstration, we show two snapshots of the iterative reverse process at $t = 0.5$ (left) and $t = 1$ (right) using the vector field U-Net.

Hyperparameters	**tSZ data**
Train set size	10,142
# steps	240K
Training batch size	128
Optimizer	Adam
Learning rate	2e-4
Latent dimension	256
Number of model channels	128
Number of residual blocks	2
Channel multiplier	[1, 2, 2, 2]
Number of attention heads	4
Dropout	0.1
$(\beta, \lambda_1, \lambda_2)$	(8e-5, 8e-2, 1e-2)
Probability path σ	0

Table A.1: Hyperparameter settings used for DL-CFM model training on the tSZ dataset.

sample the latent variable and project it to the embedding space of the CFM model using a single trainable MLP layer. These feature embeddings are added (Fig. A.1) to the time embeddings and passed to the U-Net. We use the same U-Net architecture as the ICFM for all experiments. The model is trained using the loss in Eq. 2 to enforce disentanglement in the latent spaces. Other hyperparameters and their fixed values are presented in Table A.1. For both models, inference was performed using the adaptive dopri5 solver.

The code for training and evaluation of DL-CFM can be found in https://anonymous.4open.science/r/Latent_CFM-66CF.

B.8 Computational cost

Both ICFM and DL-CFM models were trained using NVIDIA A100 GPUs. For both models, it took $\sim$ 24 hours to complete 240K training steps.

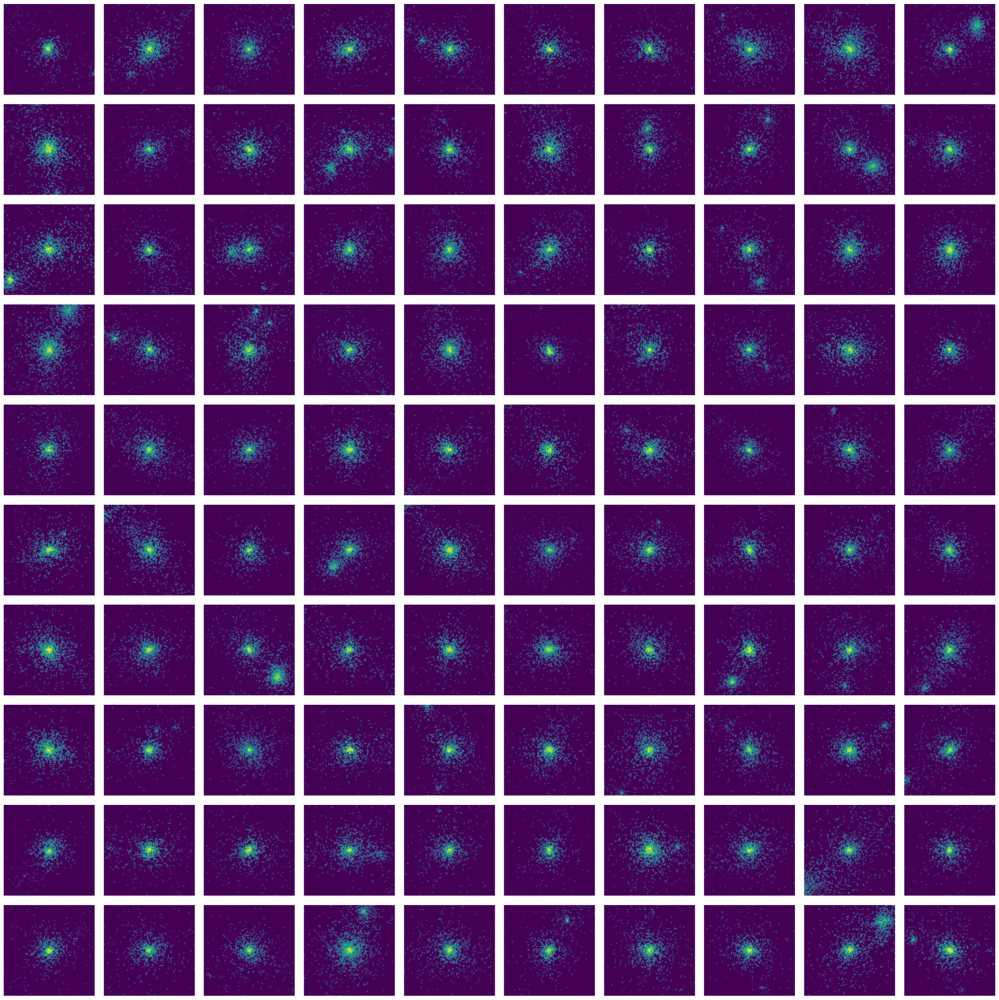

Figure A.2: Generating samples from the *center* of the reconstruction-focused latents z_{rec}, with the first two auxiliary-guided coordinates fixed at $(z_1, z_2) = (0.001, 0.9)$ - **low-mass high-concentration setting**.

C Additional experimental results

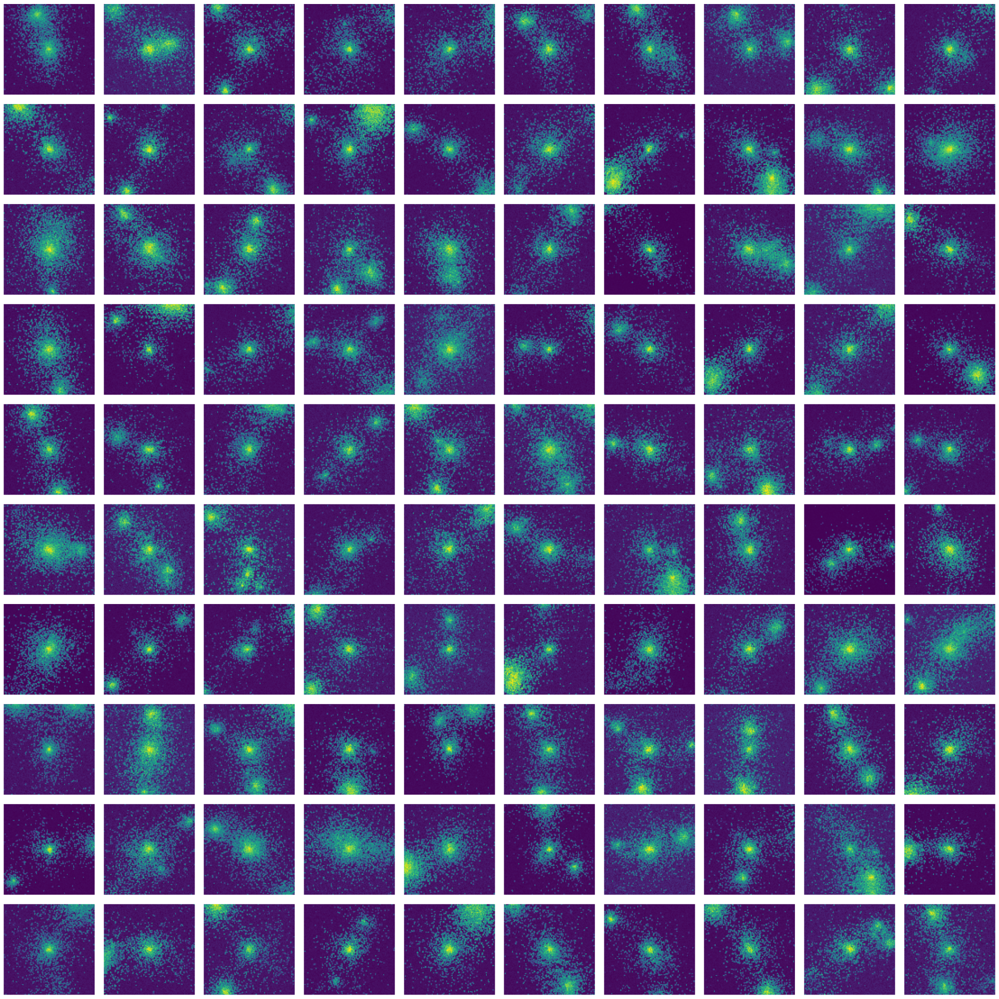

Figure A.3: Generating samples from the *tail* of the reconstruction-focused latents z_{rec}, with the first two auxiliary-guided coordinates fixed at $(z_1, z_2) = (0.001, 0.9)$ - **low-mass high-concentration setting**.

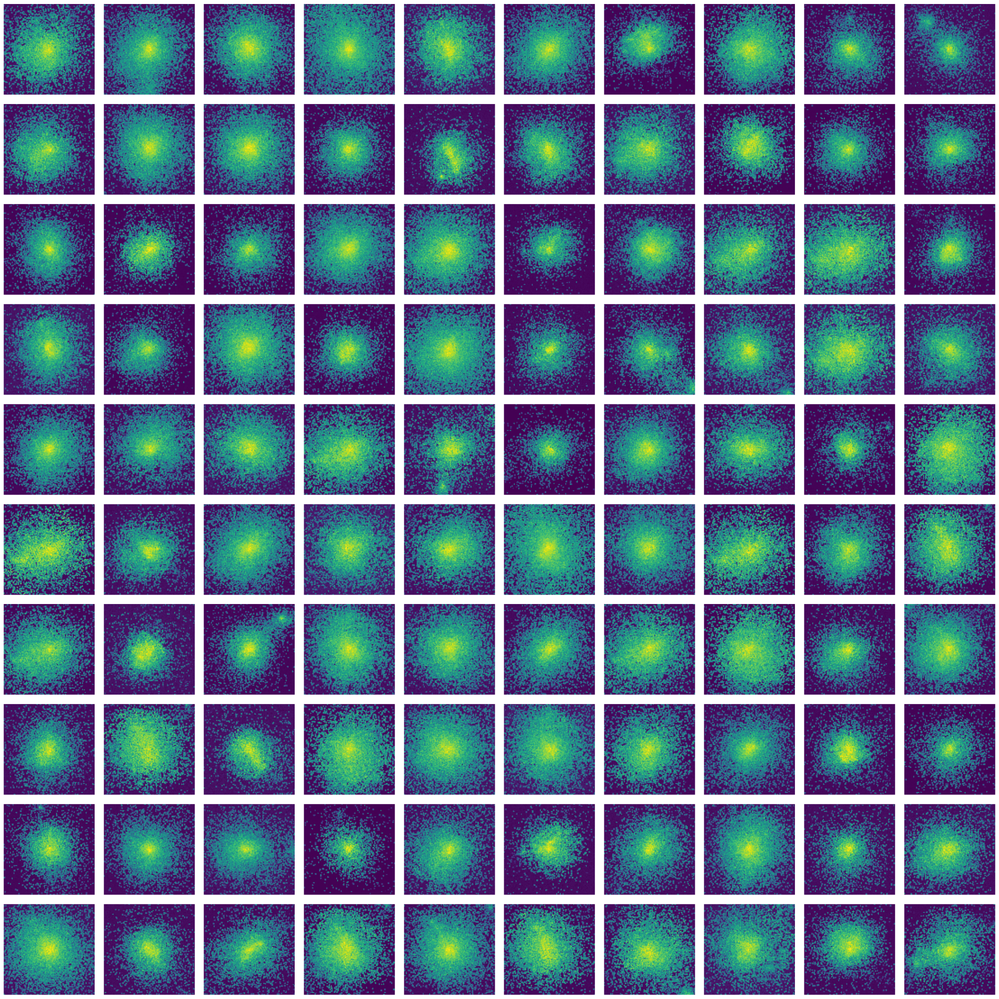

Figure A.4: Generating samples from the *center* of the reconstruction-focused latents z_{rec}, with the first two auxiliary-guided coordinates fixed at $(z_1, z_2) = (0.9, 0.001)$ - **high-mass low-concentration setting**.

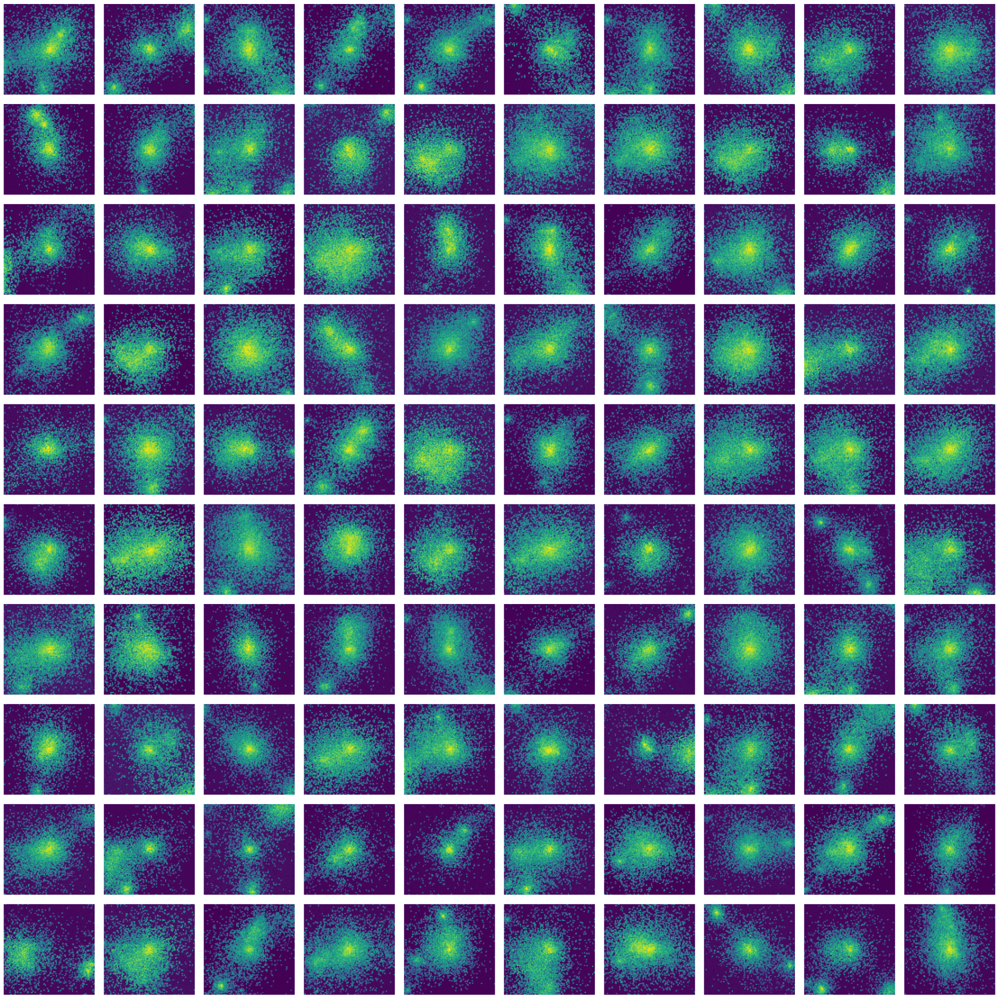

Figure A.5: Generating samples from the *tail* of the reconstruction-focused latents z_{rec}, with the first two auxiliary-guided coordinates fixed at $(z_1, z_2) = (0.9, 0.001)$ - **high-mass low-concentration setting**.

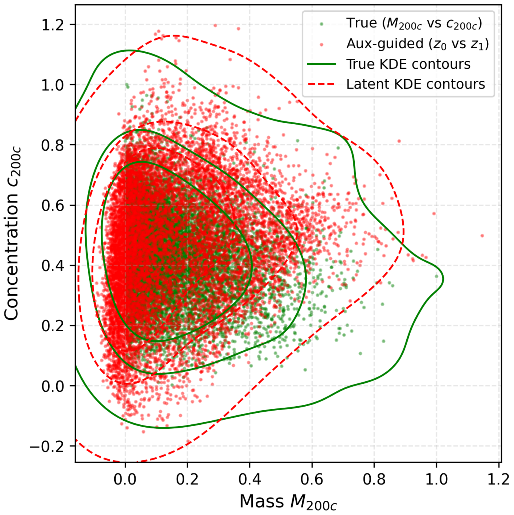

Figure A.6: The auxiliary-guided latents maintain the inter-dependency between halo mass and concentration, yielding consistent $M_{200c}c_{200c}$ dependency structure.