

BIND (Baryonic INpainting with Deep learning): A Field-level Emulator for Galaxy Groups and Clusters

MAX E. LEE ¹, SHY GENEL ², ZOLTÁN HAIMAN ^{1,3,4}, GREG L. BRYAN ¹, CHRISTOPHER C. LOVELL ^{5,6} AND
BORYANA HADZHIYSKA ^{5,6}

¹*Department of Astronomy, Columbia University, MC 5246, 538 West 120th Street, New York, NY 10027, USA*

²*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Ave, New York, NY, 10010, USA*

³*Department of Physics, Columbia University, MC 5255, 538 West 120th Street, New York, NY 10027, USA*

⁴*Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400 Austria*

⁵*Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, UK*

⁶*Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge, CB3 0HA, UK*

ABSTRACT

Baryonic feedback is a dominant source of systematic uncertainty for upcoming weak-lensing surveys, but current tools for modeling its effect rely on spherical approximations and density profiles calibrated almost entirely on two-point statistics. We introduce BIND (Baryonic INpainting with Deep learning), a conditional flow-matching model that learns a field-level mapping from dark-matter-only halos to their hydrodynamical counterparts. BIND is trained on halos from the 1024 paired hydrodynamical and dark-matter-only simulations of the CAMELS $50 h^{-1}$ Mpc SB35 suite and samples dark matter, gas, and stellar mass fields over redshift across the full 35-dimensional Λ CDM and IllustrisTNG galaxy formation parameter space. BIND recovers dark matter, gas, and stellar masses at the percent level, reproduces azimuthally averaged profiles to $\lesssim 10\%$ at all radii, and matches halo shape distributions with high fidelity. The learned parameter dependence captures the rank correlations between the generated fields and the subgrid parameters, and the field-level response to individual parameter variations is recovered in both sign and morphology. Halo mass is never supplied as conditioning, yet the baryon fraction, stellar-to-halo mass relation, inter-component scaling relations, and the joint covariance of their residuals are all reproduced. We finally show that, applied halo-by-halo to a $(50 h^{-1} \text{ Mpc})^3$ N -body volume with 512^3 particles, BIND reproduces the projected matter power spectrum suppression to the accuracy ceiling set by pasting in the hydrodynamical halos themselves, in minutes on one GPU. We release the trained BIND models and all generated halos as open-source tools. A companion paper (M. E. Lee et al. 2026a) extends BIND to thermodynamic fields and non-Gaussian weak-lensing statistics.

Keywords: Galaxies (573) — Cosmology (343)

1. INTRODUCTION

Exploiting the full statistical power of future weak lensing surveys requires accurate modeling of the matter distribution on small, highly non-linear scales. Baryonic processes, including feedback from active galactic nuclei (AGN) and supernovae, radiative cooling, and star formation, have previously been a subdominant effect in cosmological analyses, but for upcoming surveys they will become the dominant source of systematic uncertainty. Therefore, we need modeling approaches that

account for these effects (e.g., Y. P. Jing et al. 2006; N. E. Chisari et al. 2019; A. Schneider et al. 2019). In response to this demand, a domain in cosmology and astrophysics has emerged, developing methods for baryonic accounting in cosmological analyses.

The most straightforward and arguably cost-efficient treatment of baryonic effects is to remove small-scale data entirely, a strategy adopted in several weak-lensing analyses, including DES, KiDS, and HSC (e.g. L. F. Secco et al. 2022; X. Li et al. 2023; R. Dalal et al. 2023). Such cuts discard the non-linear scales over which baryons dominate, but these same small scales have also been shown to contain a wealth of cosmological informa-

tion, and their removal leads to a non-negligible loss in cosmological constraining power (e.g., Table V and Section IV B of W. Fang & Z. Haiman 2007).

The alternative extreme is to model baryonic effects directly with cosmological hydrodynamical simulations that self-consistently evolve dark matter and baryons. When appropriately calibrated, this approach serves as an ideal testing ground for baryonic effects. However, these simulations, such as IllustrisTNG, EAGLE, SIMBA, BAHAMAS, FLAMINGO or COLIBRE (e.g. M. Vogelsberger et al. 2014; J. Schaye et al. 2015; A. Pillepich et al. 2018a; R. Davé et al. 2019; J. Schaye et al. 2023, 2026), have limited mass and spatial resolution, and the galaxy-formation physics below the resolution scale is implemented through uncertain subgrid prescriptions that differ from code to code (M. P. van Daalen et al. 2011; M. Vogelsberger et al. 2020; F. Villaescusa-Navarro et al. 2021). In addition, their high computational cost precludes their use in forward-modeling pipelines that require simultaneous sampling of cosmological and galaxy-formation parameters.

To reduce the computational cost of hydrodynamical simulations while avoiding small-scale data cuts, semi-analytic baryonic correction models (BCMs) have been proposed. BCMs modify dark-matter-only (DMO) simulations, which are far cheaper than hydrodynamical simulations, by applying physically motivated empirical prescriptions to halo density profiles (A. Schneider & R. Teyssier 2015; A. Schneider et al. 2019; G. Aricò et al. 2020). In these models, a baryonic halo profile is decomposed into components such as cold bound gas, ejected gas, and relaxed dark matter, each with an analytic form. DMO halos from N-body simulations can then have their particles radially shifted so that the resulting halo profile matches the sum of the baryonic decomposition. Ideally the parameters from these models could be fit to X-ray and CMB observations, but as of yet, they have primarily been instead calibrated to hydrodynamical simulations (however, see M. Kovač et al. 2025, for approaches fitting to data).

BCMs have proven highly effective for cosmological analyses. The A20-BCM introduced in G. Aricò et al. (2020) reproduces the suppression of the 3D matter power spectrum due to baryons at the percent level across several state-of-the-art hydrodynamical simulations. Building on this, the BACCO project constructed a neural-network-based emulator over a multi-dimensional baryonic and cosmological parameter space, enabling fast evaluation of BCM predictions for $P(k)$ (G. Aricò et al. 2021a).

Similar to BCM models, halo-model approaches such as HMcode-2020 introduce a small number of baryonic

feedback parameters into an augmented halo model to match the matter power spectrum to percent-level accuracy (A. J. Mead et al. 2021), but they do not directly specify the underlying field-level redistribution of mass within and around halos. As a result, their impact on higher-order and non-Gaussian statistics must be calibrated a posteriori and may not faithfully represent the true baryon-induced spatial structure (M. E. Lee et al. 2026b).

The more recent component-wise baryonification (BFC) framework of A. Schneider et al. (2025) extended previous BCMs by assigning each DMO particle from an N -body simulation a dark matter, gas, and stellar component, which are displaced individually, as well as thermodynamic properties using hydrostatic and ideal gas equations. The BFC model was validated against the FLAMINGO and IllustrisTNG simulation suites, achieving $\lesssim 2\%$ agreement in the matter power spectrum suppression up to $k = 5 \ h \text{Mpc}^{-1}$ across a range of feedback prescriptions. In a companion paper, M. Kovač et al. (2025) applied the BFC framework to jointly fit kinematic Sunyaev-Zel’dovich (kSZ) observations from the Atacama Cosmology Telescope (ACT) and X-ray gas fractions from eROSITA, finding that current multiwavelength data prefer a stronger feedback model than assumed in most hydrodynamical simulations, with a resulting power spectrum suppression of 20–25% at $k = 5 \ h \text{Mpc}^{-1}$. This is consistent with recent findings by B. Hadzhiyska et al. (2024); L. Bigwood et al. (2024) and J. Siegel et al. (2025), who also find that feedback models in state-of-the-art simulations appear to underpredict the level of baryon suppression inferred from observations.

A complementary strategy is to account for baryons in cosmic shear analyses directly from data rather than through a model. Cross-correlating shear with tracers of the ionized gas, such as the kSZ effect (L. Bigwood et al. 2024) or fast radio burst dispersion measures (C. Leung et al. 2025; A. Wayland et al. 2026), constrains, or in principle nulls, the baryonic contribution to the lensing signal, and joint kSZ and CMB-lensing measurements around galaxies and groups now recover the gas fraction as a function of halo mass directly (B. Hadzhiyska et al. 2025). These approaches are largely independent of the subgrid physics of any one simulation, and they provide the empirical targets that any baryonification model, including the one presented here, must eventually be calibrated against.

BCMs, halo-model approaches, and BFCs have been calibrated and tested almost exclusively on two-point statistics. The first steps beyond this regime, such as joint fits to the power spectrum and bispectrum,

have shown that while BCMs can achieve $\lesssim 3\%$ agreement with the bispectrum, discrepancies persist on small scales and for certain triangle configurations (G. Aricò et al. 2021b). Moreover, when tested on non-Gaussian statistics such as peak counts, BCMs can fail more dramatically. M. E. Lee et al. (2023) found that while the A20-BCM reproduces weak-lensing peak counts at the percent level for peaks with signal-to-noise $S/N < 4$, it systematically underpredicts the number of the highest peaks, particularly for deep, wide-area surveys such as the Vera C. Rubin Observatory LSST (Ž. Ivezić et al. 2019) and *Euclid* (R. Laureijs et al. 2011).

More recently, several field-level and map-level approaches have been developed to go beyond analytic halo profiles. D. Sharma et al. (2025) proposed a model for baryonic effects where DMO simulation modes are multiplied by an isotropic Fourier-space transfer function, $\sqrt{P_{\text{hydro}}(k)/P_{\text{DMO}}(k)}$. This approach is motivated by the observation that the cross-correlation coefficient between matched hydrodynamical and N -body simulations remains close to unity down to small scales and is accurate for two-point statistics across thousands of simulations (D. Sharma et al. 2025). However, because the transfer function is isotropic in k -space, this approach cannot capture anisotropic, halo-centric mass redistribution.

Several other approaches to baryonic modeling have been explored that leverage machine learning. At the halo level, U. Chadayammuri et al. (2023) used a U-Net to map DMO cluster fields to gas density, temperature, and X-ray observables, reproducing observed mass-observable scaling relations for massive clusters. Similar halo-based approaches have used machine learning to infer baryonic properties such as gas mass and temperature from dark-matter halo properties in simulations such as EAGLE (e.g. B. Moews et al. 2021; C. C. Lovell et al. 2022, and the references therein). Other generative models such as diffusion models (J. Ho et al. 2020; Y. Song et al. 2021) have also been explored for cosmological fields, for example, Cosmo-FOLD, introduced by A. Mishra et al. (2026), uses a latent diffusion model to generate 3D dark-matter and gas-density fields conditioned on lower-resolution or partial information, achieving $\lesssim 10\%$ accuracy in power spectra up to $k \lesssim 5 h \text{ Mpc}^{-1}$.

Generative models are particularly attractive because they sample a learned posterior of baryonic effects rather than return a single deterministic prediction. This means they can represent the intrinsic physical scatter, and this has now been demonstrated across a range of astrophysical inverse problems. Score-based priors recover posterior samples of lensed source galaxies (A. Adam

et al. 2022) and of the weak lensing convergence field from noisy shear (B. Remy et al. 2023), they recover the initial conditions of the Universe from the evolved density field (R. Legin et al. 2023), diffusion models reconstruct dark matter fields from galaxies in CAMELS (F. Villaescusa-Navarro et al. 2021; Y. Ni et al. 2023; S. Genel et al. 2026) – a suite of thousands of cosmological hydrodynamical simulations spanning various galaxy formation parameter spaces – while marginalizing over cosmology and feedback (V. Ono et al. 2024), they reconstruct cluster gas and dark matter maps from mock SZ and X-ray images (A. Hsu et al. 2024), and they have been used directly for cosmological parameter inference (N. Mudur et al. 2024).

Closest in spirit to the present work, C. Cuesta-Lazaro & S. Mishra-Sharma (2024) generate halo point clouds at the field level with a diffusion model conditioned on cosmology, and S. Pandey et al. (2025) paint galaxies onto dark-matter-only simulations with a transformer-based model trained on CAMELS, reproducing galaxy clustering and one-point statistics at a fraction of the cost of a hydrodynamical run. These neural network and generative methods are an important step toward field-level emulation of baryonic effects, but they either operate on full volumes without explicit halo-level conditioning, or target the galaxy population rather than the continuous fields that set the lensing signal, making it nontrivial to connect them to the halo-level observables associated with these baryonic effects.

Given the rapid progress and successes of the methods described above, several opportunities remain to expand the current state of baryonic modeling.

First, BCMs and BFCs, while halo-centric, rely on fixed analytic profile prescriptions and are not generative, so they cannot sample from a posterior over baryonic field configurations at fixed dark matter and model parameters. Transfer-function approaches are field-level, isotropic in k -space, and similarly non-generative. Map-level ML approaches operate on full volumes rather than individual halos. The advantage of halo-centric approaches is that the source of baryonic effects lies in processes within halos, and by applying full-volume corrections, one loses the halo-centric resolution needed to capture detailed mass redistribution within and around groups and clusters. Further, because most particles in simulations reside outside halos, correcting full volumes is computationally expensive and does not necessarily affect the scales we truly care about.

Second, higher-order and non-Gaussian weak-lensing statistics, such as peak and minimum counts, Minkowski functionals, and field-level summaries, have been shown to add substantial cosmological information beyond the

power spectrum, but existing baryonic correction tools have not been systematically validated for these observables (M. E. Lee et al. 2023, 2026b). In fact, M. E. Lee et al. (2026b) shows explicitly that different lensing statistics are sensitive to different halo masses and radial ranges, implying that models fit to one lensing statistic may fail to capture the profile changes required by another. The best way to ensure all statistics are captured well is with a field-level approach that accurately captures the anisotropic redistribution and full field inside and around halos.

Third, the effect of galaxy-formation model parameters on matter redistribution and its propagation to weak lensing statistics is poorly understood. The 35-dimensional IllustrisTNG galaxy formation parameter space (30 astrophysical + 5 Λ CDM cosmological parameters) contains complex, nonlinear degeneracies between both the astrophysical parameters themselves, and between the subgrid and cosmological parameters together (for example, between supernova and AGN feedback), which analytic prescriptions are potentially ill-equipped to capture (M. E. Lee et al. 2024).

No existing method simultaneously operates at the field level, conditioned on galaxy formation model parameters, in a halo-centric manner while also being generative, enabling proper posterior sampling over baryonic field realizations at fixed inputs. Table 1 summarizes where existing approaches stand on these axes.

In this work, we introduce BIND (Baryonic INpainting with Deep learning), a conditional flow matching model that learns the mapping from dark-matter-only halos to their hydrodynamical counterparts at the field level. BIND is trained on 151,685 halos from matched hydrodynamical and dark-matter-only simulations across 8 redshifts in the CAMELS IllustrisTNG suite of $50 h^{-1}$ Mpc simulations (S. Genel et al. 2026) and is conditioned on redshift and the full set of 35 cosmological and astrophysical parameters from Λ CDM and the IllustrisTNG galaxy formation model.

On a halo-by-halo basis, BIND takes as input a projected two-dimensional DMO surface density field centered on a halo from an N-body simulation and outputs simultaneous predictions for the dark matter, gas, and stellar mass distributions expected in a hydrodynamical simulation with a given set of astrophysical and cosmological parameters. Because the model is generative, it naturally provides samples from the posterior distribution of baryonic fields given a DMO realization and a parameter vector, representing stochastic draws for the same halo. In a companion paper (M. E. Lee et al. 2026a), BIND is extended to output field-level thermodynamic quantities for gauging feedback at the

halo level, such as temperature, pressure, and entropy maps.

By construction, BIND is halo-centric, field-level, and parameter-conditioned. This lets it capture anisotropic mass redistribution in and around halos, learn how different mass components vary across the galaxy-formation parameter space, and preserve higher-order information in the output fields, enabling direct application to downstream statistics. Because the only spatial input to the model is the projected dark-matter field, the generated fields do not depend on any halo mass or radius definition; a mass threshold enters only in choosing which halos to process. This sidesteps one of the persistent ambiguities of halo-based baryonification models, whose profiles are anchored to a particular mass and radius convention. By applying BIND patch-by-patch to the halos of an arbitrary N -body simulation, we show that one can baryonify large cosmological volumes across the IllustrisTNG parameter space in minutes on a single GPU, orders of magnitude faster than re-running a hydrodynamical simulation. This makes BIND suitable for field-level analysis pipelines.

In this paper, we introduce BIND and provide a comprehensive suite of validations exploring its ability to generate accurate density fields with correct parameter dependencies. We are primarily interested in answering:

1. Can a conditional flow matching model learn the mapping from dark-matter-only to hydrodynamical fields at the halo level, conditioned on the joint Λ CDM + IllustrisTNG galaxy-formation parameter space?
2. Does BIND learn the appropriate astrophysical parameter dependencies for the generated hydrodynamical fields?
3. Is BIND able to learn the correlations between field-level observables and how they change as a function of astrophysical parameters?
4. Can BIND be deployed as a practical baryonification engine for two-point statistics such as the matter power spectrum? In a companion paper, we explore the efficacy of BIND on non-Gaussian statistics.

The paper is organized as follows. In § 2 we introduce the training and testing suites ingested by BIND. We then discuss the flow-matching model and the data-processing pipeline in § 3. We tackle the first and second questions above and validate the model both with fixed parameters and across the parameter space in § 4 and § 5, respectively. We then explore the third question and

Table 1. Capabilities of currently implemented baryonic-modeling approaches across the gaps identified in § 1.

Approach	Halo-centric	Field-level	Generative	Param.-cond.
BCM / halo model ^a	yes	no (profiles)	no	BCM params
BFC ^b	yes	particle-level	no	BFC params
Isotropic transfer function ^c	no	yes	no	suite params
Volume-level generative ^d	no	yes	yes	no
Halo image-to-image ^e	yes	yes	no	fixed physics
Galaxy painting ^f	no (galaxies)	no (catalog)	yes	6 (cosmo + astro)
BIND (this work)	yes	yes (2D patches)	yes	35 (cosmo + TNG)

NOTE—^aA. Schneider & R. Teyssier (2015); G. Aricò et al. (2020); A. J. Mead et al. (2021); ^bA. Schneider et al. (2025); ^cD. Sharma et al. (2025); ^dA. Mishra et al. (2026); ^eU. Chadayammuri et al. (2023); ^fS. Pandey et al. (2025).

joint distributions of generated fields in § 6 and consider the final question above with the application of BIND to a full N-body simulation in § 7. We close with a list of caveats in § 8, and conclusions in § 9.

2. SIMULATIONS

All training and testing data in this work come from the Cosmology and Astrophysics with Machine-Learning Simulations (CAMELS; F. Villaescusa-Navarro et al. 2021, 2023) project. CAMELS is a suite of hydrodynamical and N -body simulations designed to train and test machine-learning models on cosmological data. Beyond machine learning, the suite has been used extensively to characterize how subgrid feedback reshapes the baryon distribution, including its effect on the warm-hot circumgalactic medium (I. Medlock et al. 2024b), on X-ray observables of the circumgalactic gas (E. T. Lau et al. 2024), on fast radio burst dispersion measures (I. Medlock et al. 2024a, 2025), and on the dark matter distribution itself (M. Gebhardt et al. 2026). In the updated $50 h^{-1}$ Mpc suite of S. Genel et al. (2026), each simulation follows 512^3 dark-matter particles and, in the hydrodynamical runs, an equal number of gas resolution elements in a periodic comoving box of side length $L = 50 h^{-1}$ Mpc. The hydrodynamical runs employ the IllustrisTNG galaxy-formation model (R. Weinberger et al. 2017; A. Pillepich et al. 2018b), implemented in the AREPO moving-mesh code (V. Springel 2010). For a description of the flagship IllustrisTNG runs and their public data products, see D. Nelson et al. (2019).

2.1. The SB35 Sobol Sequence Set

We train BIND on the SB35 suite (S. Genel et al. 2026), consisting of 1024 paired IllustrisTNG hydrodynamical and N -body simulations spanning a 35-dimensional parameter space of cosmological and galaxy

formation parameters sampled with a Sobol sequence (I. M. Sobol’ 1967). The parameter space includes five cosmological parameters and 30 galaxy-formation subgrid parameters governing stellar feedback, AGN feedback, star formation, and interstellar medium processes in the IllustrisTNG model (for a full breakdown of the parameters, see S. Genel et al. 2026, particularly Table 1). The Sobol sequence ensures space-filling coverage of this high-dimensional parameter space, with each simulation using a unique combination of parameter values and random initial phases. From every simulation we use eight snapshots spanning $z = 0$ –4; however, the evaluations of Sections 4 and 7 are performed at $z = 0$, and § 5.4 validates the model across the full redshift range.

Each hydrodynamical simulation and its gravity-only N -body twin is evolved from identical initial conditions. Therefore, by holding the initial conditions fixed and aside from intrinsic randomness due to the butterfly effect (S. Genel et al. 2019), differences between the N -body and hydrodynamical outputs arise solely from baryonic physics and provide a clean training signal for the mapping of pixelated mass maps $\mathbf{M}_{\text{DMO}} \rightarrow \mathbf{M}_{\text{hydro}}$.

2.2. Evaluation Suites

We use three independent simulation sets for evaluation.

The **Sobol sequence (SB35) test set** consists of the 102 SB35 simulations withheld from training. Each simulation contains a different set of cosmological and astrophysical parameters and initial conditions.

The **One-Parameter (1P) set** contains simulations in which a single parameter is varied at a time while all others are held at their fiducial values, and all runs share the same initial conditions.

The **Cosmic Variance (CV) set** consists of 27 simulations run at the fiducial IllustrisTNG parameter values but with different random initial conditions.

3. MODEL AND TRAINING

Our goal is to learn the conditional distribution,

$$p(\mathbf{M}_{\text{hydro}} \mid \mathbf{M}_{\text{DMO}}, \boldsymbol{\theta}_{\text{TNG}}, z). \quad (1)$$

Each channel of $\mathbf{M}_{\text{hydro}}$ is a halo-centered projected mass map on an $H \times W$ pixel grid, for $c \in \{\text{dm, gas, stars}\}$, where the value in a given pixel is the total mass in that pixel's column through the box,

$$M_{c,ij} = \sum_k m_{c,ijk}, \quad (2)$$

where the sum runs over all particles of species c whose projected position falls in pixel (i, j) , along the full $50 h^{-1}$ Mpc line of sight. Throughout, bold symbols denote pixelated two-dimensional maps, so that $\mathbf{M}_{\text{hydro}} \in \mathbb{R}^{3 \times H \times W}$ with one channel per species. $\mathbf{M}_{\text{DMO}} \in \mathbb{R}^{1 \times H \times W}$ is the conditioning input computed in the same way and is the dark-matter-only patch centered on a halo. Note that we have chosen the fields as masses here, though this choice is somewhat arbitrary. In our companion paper (M. E. Lee et al. 2026a), we include gas pressure, temperature, and entropy fields such that $\mathbf{M}_{\text{hydro}} \in \mathbb{R}^{6 \times H \times W}$, but for clarity in this work, we limit the channel outputs to just the three mass fields.

The parameter vector $\boldsymbol{\theta}_{\text{TNG}} \in \mathbb{R}^{35}$ encodes the full set of cosmological and astrophysical parameters of the simulation (Section 2.1); however, in general, this vector can be any conditioning parameter vector that is associated with the halo. One future plan discussed in § 8 is to train BIND using physical inputs measured from each simulation, rather than the TNG parameter vector. For this first iteration of BIND, we stick to the parameter set handed down by the simulation. The scalar z denotes the snapshot redshift, supplied to the network as the scale factor $a = 1/(1+z)$. BIND is trained on halos from multiple snapshots, and its redshift conditioning is validated in Section 5.4.

The BIND framework is built on a conditional flow matching network that learns the mapping from dark-matter-only halo fields and IllustrisTNG model parameters to the baryonic mass fields of the paired hydrodynamical simulation. In the following subsections, we detail the data-processing pipeline, review the flow-matching formalism, describe the network architecture, and specify the training procedure.

3.1. Data Processing

Converting raw simulation particle data into training samples for BIND involves four main steps: halo iden-

tification and centering, random rotation, line-of-sight projection, and field normalization.

We treat the N -body $50 h^{-1}$ Mpc CAMELS suite of simulations as the base set from which halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ are identified⁷ at 8 redshifts between $z = 0$ and $z = 4$. We find halos with the friends-of-friends algorithm (M. Davis et al. 1985), and we take their centers and bound substructure from the SUBFIND catalogs (V. Springel et al. 2001). For each halo, we exploit the periodic boundary conditions of the simulation box and re-center the volume on the particle with the deepest gravitational potential (GROUPPOS in the FoF catalog), placing the halo at $(x, y, z) = (25, 25, 25) h^{-1}$ Mpc.

We place a stationary observer at $(x, y, z) = (25, 25, 0) h^{-1}$ Mpc and apply a random rotation about the halo center. The periodic boundary conditions ensure that, with appropriate tiling of the simulation box, a $50 h^{-1}$ Mpc cube remains volume-filling after the rotation. We store the three Euler angles defining each rotation and apply them to the paired hydrodynamical simulation to ensure exact spatial correspondence between the N -body and baryonic projections. We perform this procedure 10 times per halo, drawing independent rotation matrices each time to augment the training set.

We project each rotated volume grid along the z -axis using cloud-in-cell (CIC) interpolation (R. W. Hockney & J. W. Eastwood 1988). We retain the central 128×128 pixel patch as the primary N -body conditioning field $\mathbf{M}_{\text{DMO}}$. For the hydrodynamical simulations, we use the same procedure to project the dark-matter, gas, and stellar density fields. All fields are then defined on a regular $H \times W = 128 \times 128$ grid with $0.049 h^{-1}$ Mpc resolution per pixel extending $6.25 h^{-1}$ Mpc.

We highlight a few of the design choices described above. First, we do this in 2D, which lets us explore the architecture and test our procedures at reduced computational cost. 3D fields increase the memory and computational requirements by orders of magnitude, so this first iteration of BIND remains 2D. This choice is a well-motivated first step, though, as baryonification applied directly at the map level reproduces weak lensing two-point and higher-order statistics to within a few percent (D. Anbajagane et al. 2024; A. J. Zhou et al. 2025). Extending the procedure to 3D fields is a goal saved for future development of BIND. The other important caveat is the projection depth. Each training sample is $6.25 \times 6.25 h^{-1}$ Mpc and 128^2 pixels, computed as the

⁷ M_{200c} is the mass enclosed within the radius R_{200c} , inside which the mean density is $200\rho_c$, with ρ_c the critical density of the universe.

projected mass over the full $50 h^{-1}$ Mpc line of sight. While the majority of the mass within a halo will be from material associated with that halo, there will, of course, be foreground and background matter not associated with the halo included in this projection. The projection may also include multiple halos along the line of sight, which could bias our conditioning vector. Also note that we center our training data only on halo centers; subhalos are not used as explicit training data, but because each massive halo has subhalos in the field, they are implicitly included in the training. We made this projection choice to facilitate pasting the baryonified halos back into N -body simulations, which we discuss in § 7. However, we have also explored the effect of this projection and trained an equivalent model using projections of the $6.25 h^{-1}$ Mpc cube centered on each halo (see Appendix C).

Because projected surface density fields span several orders of magnitude in intensity and contain many near-zero pixels, all map channels, $\mathbf{M}_c$, are compressed with

$$\mathbf{M}'_c = \log_{10}(1 + \mathbf{M}_c), \quad (3)$$

which maps zero-density pixels to zero while suppressing the extreme dynamic range. Each transformed channel $\mathbf{M}'_c$ is then standardized to zero mean and unit variance,

$$\tilde{\mathbf{M}}_c = \frac{\mathbf{M}'_c - \mu_c}{\sigma_c + \epsilon}, \quad (4)$$

where μ_c and σ_c are estimated from a random subsample of 10,000 training maps, and $\epsilon = 10^{-8}$ prevents division by zero. This normalization is applied identically to the target baryonic fields and $\mathbf{M}_{\text{DMO}}$, but with their own computed μ_c and σ_c per field.

The parameter vector $\boldsymbol{\theta}_{\text{TNG}}$ is normalized component-wise and in accordance with the parameter sampling to generate the Sobol sequence in S. Genel et al. (2026). This requires many parameters (see Table 1 of S. Genel et al. 2026 for the list of log-space parameters) to first be log-transformed before linear rescaling to $[0, 1]$:

$$\tilde{\theta}_j = \frac{\theta_j^{(\log)} - \theta_{j,\min}^{(\log)}}{\theta_{j,\max}^{(\log)} - \theta_{j,\min}^{(\log)} + \epsilon}, \quad (5)$$

where

$$\theta_j^{(\log)} = \begin{cases} \log_{10}(\theta_j), & \text{if parameter } j \text{ is log-scaled,} \\ \theta_j, & \text{otherwise.} \end{cases} \quad (6)$$

The bounds $\theta_{j,\min}$ and $\theta_{j,\max}$ are taken from the SB35 parameter tables (S. Genel et al. 2026), so the normalization range is well-defined for any simulation in the suite, not only those in the training subset.

We split the 1024 SB35 simulations into 90% for training (922 simulations) and 10% for testing (102 simulations). Within the training set, we reserve 20% (184 simulations) for validation, and use the remaining 80% (738 simulations) for training. After applying 10 random rotations per halo, the training set contains approximately 360,000 projected halo maps spanning the full 35-dimensional parameter space.

3.2. Flow Matching

The overall goal in our conditional flow matching is to learn a velocity field $\mathbf{v}_\phi$ that transports samples from an isotropic Gaussian prior, $\mathbf{M}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, to the target baryonic field $\mathbf{M}_{\text{hydro}}$ conditioned on the paired dark-matter-only halo $\mathbf{M}_{\text{DMO}}$, a parameter vector $\boldsymbol{\theta}_{\text{TNG}}$ and the scale factor a . In this section, we now describe how we build the velocity model.

The trajectory of the field from its Gaussian initial state to the final field follows a path such that intermediate states on the time interval, $t \sim \mathcal{U}(0, 1)$, are simply a linear interpolation of the initial and final fields,

$$\mathbf{M}_t = (1 - t) \mathbf{M}_0 + t \mathbf{M}_{\text{hydro}}. \quad (7)$$

This defines a straight-line probability path from noise at $t = 0$ to data at $t = 1$ (Y. Lipman et al. 2023), an objective developed concurrently as stochastic interpolants (M. S. Albergo & E. Vanden-Eijnden 2023) and as rectified flow (X. Liu et al. 2023). We use this formulation rather than the diffusion one (J. Ho et al. 2020; Y. Song et al. 2021) because the interpolation in Eq. 7 fixes the probability path in advance and makes the regression target in Eq. 8 constant along each training trajectory. However, we have also explored similar architectures using diffusion models and found that, for this problem, which involves highly sparse stellar fields, the flow-matching model yields the highest-fidelity reconstructions.

The target velocity along this path is the difference between the target and noise fields with respect to the time, which by construction is 1,

$$\mathbf{v}^* = \mathbf{M}_{\text{hydro}} - \mathbf{M}_0. \quad (8)$$

We parameterize the velocity field with a neural network $\mathbf{v}_\phi(\mathbf{M}_t, t, \mathbf{M}_{\text{DMO}}, \boldsymbol{\theta}_{\text{TNG}}, a)$, with neural network parameters ϕ , and we write $\mathbf{v}_\phi$ for brevity, suppressing the scale-factor argument in the equations below. The network is trained to minimize the conditional flow matching objective

$$\mathcal{L}_{\text{FM}}(\phi) = \mathbb{E}_{t, \mathbf{M}_0, \mathbf{M}_{\text{hydro}}} \left[\|\mathbf{v}_\phi - (\mathbf{M}_{\text{hydro}} - \mathbf{M}_0)\|_2^2 \right], \quad (9)$$

Table 2. Measured computational cost for training and generating with BIND.

Task	Hardware	Wall time
Model Training	8×H100	19 h (~ 150 GPU-h)
Load model + normalization	A100 (MIG 2/7)	29 s
Generate one halo	A100 (MIG 2/7)	0.9 s
Baryonify a $50 h^{-1}$ Mpc box	A100 (MIG 2/7)	~ 40 s
Halo-pasting step alone (CPU)	1 core	0.05 s
100-draw posterior, one halo	A100 (MIG 2/7)	88 s

where the conditioning inputs $(\mathbf{M}_{\text{DMO}}, \boldsymbol{\theta}_{\text{TNG}}, a)$ are held fixed for each training sample.

At inference, we generate baryonic fields by integrating the learned probability-flow ODE from $t = 0$ to $t = 1$,

$$\frac{d\mathbf{M}}{dt} = \mathbf{v}_\phi(\mathbf{M}, t, \mathbf{M}_{\text{DMO}}, \boldsymbol{\theta}_{\text{TNG}}, a), \quad (10)$$

starting from $\mathbf{M}(0) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. We integrate Eq. (10) with a first-order Euler scheme using $N_{\text{steps}} = 20$ uniform time steps $\Delta t = 1/N_{\text{steps}}$,

$$\mathbf{M}_{k+1} = \mathbf{M}_k + \Delta t \mathbf{v}_\phi(\mathbf{M}_k, t_k, \mathbf{M}_{\text{DMO}}, \boldsymbol{\theta}_{\text{TNG}}, a). \quad (11)$$

Because the probability path in Eq. (7) is a straight line, the learned velocity field is nearly constant along each trajectory, and the Euler integrator remains accurate even with this modest number of function evaluations (Y. Lipman et al. 2023). In Table 2 we show that generating a single 128×128 halo patch with 20 steps takes approximately one second.

3.3. Architecture

We parameterize the velocity field $\mathbf{v}_\phi$ with a conditional U-Net (O. Ronneberger et al. 2015). Appendix A gives the full specification of widths, depths, and attention placement; here we briefly describe the model architecture.

BIND is conditioned on two different kinds of information, the first of which is spatial. The dark-matter-only patch $\mathbf{M}_{\text{DMO}}$ is a map where each of its pixels carries conditional information about where the resulting baryon fields should end up. We preserve the alignment between pixels in this spatial conditioning and the hydro conditioning by simply concatenating it channel-wise with the noisy field at each integration time t ,

$$\mathbf{X}_t^{\text{in}} = [\mathbf{M}_t \mid \mathbf{M}_{\text{DMO}}] \in \mathbb{R}^{(3+1) \times H \times W}. \quad (12)$$

Because both noise and the dark-matter-only conditioning are stacked pixel by pixel, the convolutions see the

local dark matter morphology at every encoding step, and the generated gas, stellar, and dark matter channels stay tied to the structure of the halo they are being painted onto.

The second kind of conditioning has no spatial structure and is simply a vector of numbers including the flow time t , the parameter vector $\boldsymbol{\theta}_{\text{TNG}}$, and the scale factor a . These numbers are global scalars that apply to the whole patch. To incorporate them, rather than concatenating them, which risks the values being overwhelmed by spatial conditioning, we inject them via feature-wise linear modulation (FiLM; E. Perez et al. 2018). Each of t , a , and $\boldsymbol{\theta}_{\text{TNG}}$ is embedded separately to the same dimension, so that each carries the same conditioning weight, and the embeddings are summed into a single conditioning vector $\mathbf{e}$ (Appendix A). The flow time and the redshift are encoded with the standard sinusoidal features before their embeddings, as both are single scalars whose effect on the output is smooth but strongly nonlinear, while the parameter vector is passed through its own multilayer perceptron.

$$\boldsymbol{\theta}_{\text{TNG}} \mapsto (1 + \gamma(\mathbf{e})) \odot \text{Norm}(\boldsymbol{\theta}_{\text{TNG}}) + \beta(\mathbf{e}), \quad (13)$$

where γ and β are learned linear projections of $\mathbf{e}$ and Norm is group normalization (Y. Wu & K. He 2018). This conditioning vector then acts at every depth in both the encoder and the decoder, modulating the features themselves.

The network architecture is a U-Net that is not purely convolutional. Rather than relying on network depth and additional convolutions to carry spatial information across the map, we insert multi-head self-attention layers (A. Vaswani et al. 2017) at the two coarsest resolutions of the U-Net, 32^2 and 16^2 , where every position in the feature map can interact directly with every other position. We introduced this feature to enhance small-scale fidelity within halos of the resulting generations and found that, without it, our results degrade. We therefore traded the generalizability of a purely convo-

lutional network for accuracy within the generated halos we care about.

3.4. Training

We train with the AdamW optimizer (I. Loshchilov & F. Hutter 2019) for up to 200 epochs, using a batch size of 64 maps per GPU across 8 H100 GPUs in mixed `bfloat16` precision (which enhances both speed and memory efficiency), implemented in PyTorch (A. Paszke et al. 2019) with PyTorch Lightning’s Distributed Data Parallel strategy. The learning rate warms up linearly and then anneals to zero on a cosine schedule, and we keep an exponential moving average of the network weights, which we use at inference. Checkpoints are saved every 10 epochs. The schedule, the averaging, and the complete set of hyperparameters are given in Appendix B.

In Table 2, we show that training a single BIND model in the above fashion takes ~ 150 GPU-hours, and once trained, it can produce baryonic halos in under a second. For baryonifying a $50 h^{-1}$ Mpc box with ~ 45 halos, the cost scales roughly linearly with generation time and takes about 40 seconds.

4. FIDELITY AT FIXED PARAMETERS

This section addresses the first question posed in the introduction. *Can BIND reproduce baryonic fields of individual halos at fixed parameters?* We start with qualitative field-level comparisons and move to more quantitative comparisons of integrated masses, radial profiles, and projected shapes. The response of those fields across the parameter space is the subject of Section 5. Table 3 indexes the full validation suite of this paper.

4.1. Qualitative Field-Level Comparisons

We begin with a qualitative examination of BIND’s outputs for a representative cluster-mass halo ($M_{200c} \sim 10^{14} M_{\odot} h^{-1}$) drawn from the CV test set. The N -body dark-matter field, centered on the halo as described in Section 3.1, is provided as the conditioning input together with the fiducial IllustrisTNG parameter vector θ_{fid} .

Fig. 1 illustrates both the generative trajectory and the final field-level comparison between BIND and the true hydrodynamical realization of the halo. Reading left to right, the first three columns show snapshots of the flow at $t = 0$ (isotropic Gaussian noise), $t = 0.5$ (an intermediate state), and $t = 1$ (the fully denoised output). The fourth and fifth columns show the ground-truth hydrodynamical fields and the residuals between the BINDed and true fields, respectively. By $t = 0.5$,

the network has already established the large-scale mass distribution, and the remaining integration steps primarily sharpen the morphology and populate small-scale features.

All three baryonic components are reproduced with good qualitative agreement. The errors associated with the channels appear to increase with respect to the distance a field is from its conditioning DMO input. For example, the generated dark-matter field closely traces the N -body input, and the residuals are approximately noise-like, with no coherent large-scale structure. The gas channel residuals retain more coherent spatial structure, reflecting the sensitivity of the intracluster gas to pressure forces and AGN and stellar feedback processes that are not fully encoded in the dark-matter morphology alone and for which the parameters were not able to fully communicate. The stellar channel, being highly sparse and concentrated in a small number of bright central pixels, shows the largest pixel-level scatter between the generated and true fields. For this row, we apply a small 1-pixel Gaussian smoothing to highlight differences in the residual, but even then the differences in the stellar channel are saturated and hard to discern. This is expected, as the precise location and amplitude of individual stellar clumps are stochastic for a given dark matter conditioning field and parameter set, and a generative model is not expected to reproduce a particular realization exactly, but rather the correct statistical ensemble. Beyond the stochasticity of the stellar channel is its sparsity. Individual pixel shifts relative to the truth cause large differences, whereas the DMO channel, which has a much smoother density distribution around density peaks, is less sensitive to this effect.

To highlight this point, we show in Fig. 2 a single halo generated several times with BIND using different initial Gaussian noise realizations. The mean over 100 realizations is shown in the sixth column and highlights that, across draws, the field is much smoother than the truth, which is itself just a single draw from a distribution. The final column of Fig. 2 shows the standard deviation of masses over 100 generations for a single halo. For the gas and dark matter channel, this amounts to noise from different placements of satellite galaxies within the patch. However, for the stellar channel, the spread is much larger, particularly around the edges of halos and subhalos. The typical shapes, density profiles, and masses for each individual draw roughly match the truth (see § 4); however, the stellar channel is itself stochastic and has a large spread.

These qualitative observations motivate the quantitative analyses that follow. The next sections systematically characterize BIND’s accuracy through increas-

Table 3. Summary of the validation suite.

Test	Sets	Figure(s)	Headline result
Integrated masses	CV, 1P, SB35	Figs. 3–4	$< 1\%$ DM/total, $\sim 1\%$ gas, $\sim 10\%$ stars
Radial profiles	CV, 1P, SB35	Fig. 5	$\lesssim 10\%$ at all radii to $3.125 h^{-1}$ Mpc
Projected shapes	CV, 1P, SB35	Fig. 6	q medians and widths matched
Mass–parameter correlations	SB35	Fig. 8	Spearman matched, 15 strongest params
Profile–parameter correlations	SB35	Fig. 9	matched at all radii
1P field-level response	1P	Fig. 10	sign and morphology reproduced
Baryon fraction	CV, 1P, SB35	Fig. 12	median and scatter, no mass conditioning
Scaling relations, residual corner	CV	Figs. 13–14	slopes, normalizations, joint scatter
Power spectrum suppression	CV, SB35	Fig. 16	tracks the hydro-replace ceiling

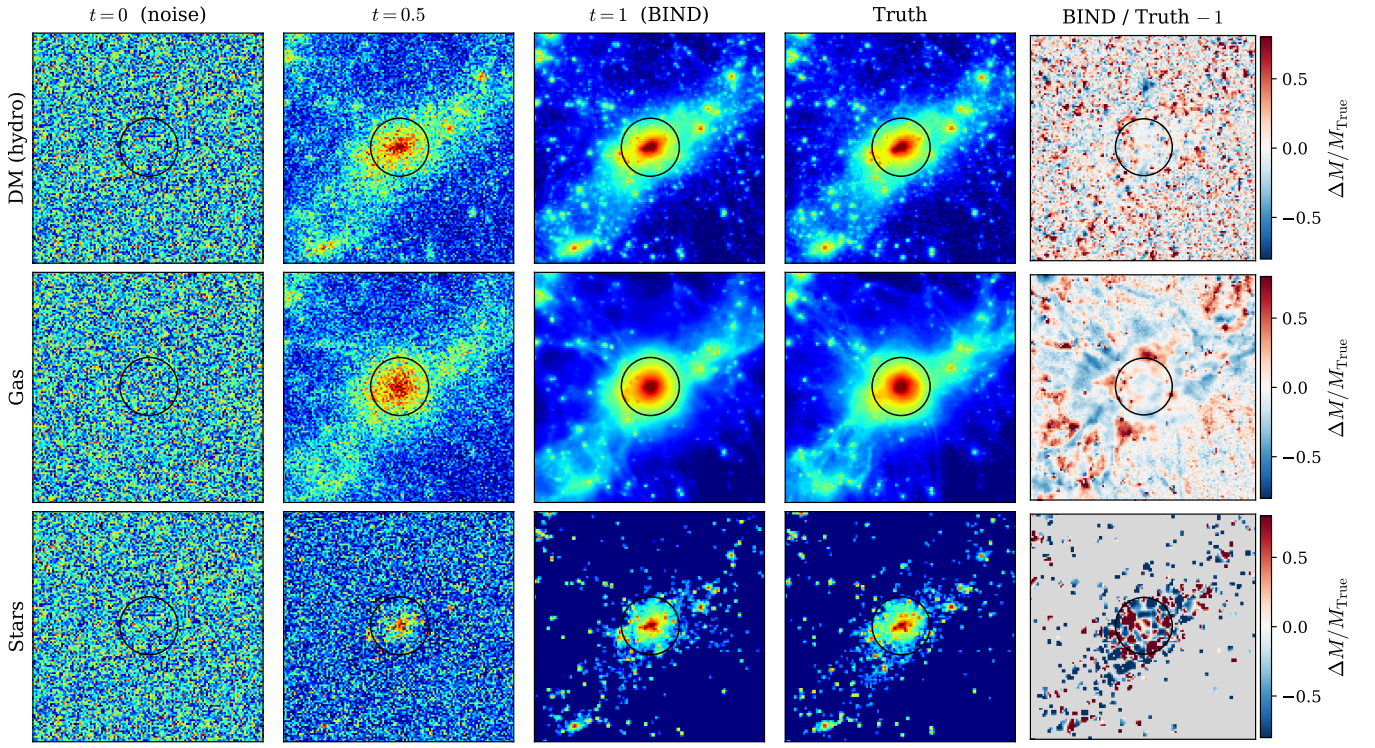


Figure 1. Flow matching trajectory and field-level comparison for a representative cluster-mass halo ($M_{200c} \sim 10^{14} M_{\odot} h^{-1}$) drawn from the CV test set, conditioned on the fiducial IllustrisTNG parameter vector θ_{fid} . *Columns 1–3:* Snapshots of projected surface density distributions of the generative trajectory at $t = 0$ (Gaussian noise), $t = 0.5$ (intermediate state), and $t = 1$ (final generated field). *Column 4:* Ground-truth hydrodynamical fields from the paired IllustrisTNG simulation. *Column 5:* Signed residual $(M_{\text{BIND}} - M_{\text{hydro}})/M_{\text{hydro}}$. *Rows:* Projected dark-matter, gas, and stellar masses, respectively. The residual panel for the stellar-mass comparison includes 1-pixel-level Gaussian smoothing. The generated fields reproduce the large-scale morphology of all three components. Residuals in the dark-matter channel are noise-like, while the gas channel retains coherent structure in the residuals, and the stellar channel shows significant pixel-level scatter owing to its sparsity.

ingly difficult tests, starting with cumulative mass comparisons (Section 4.2), moving to radially resolved surface density profiles (Section 4.3), then looking at projected shapes of halos in each channel (Section 4.4), before turning to the learned dependence on cosmological and astrophysical parameters (Section 5).

4.2. Integrated Mass

We begin with the total integrated mass, which measures whether BIND allocated the correct matter proportions to each channel. For every halo with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ in the full test set (CV, 1P and Sobol test set), we sum the projected pixel values within R_{200c} of

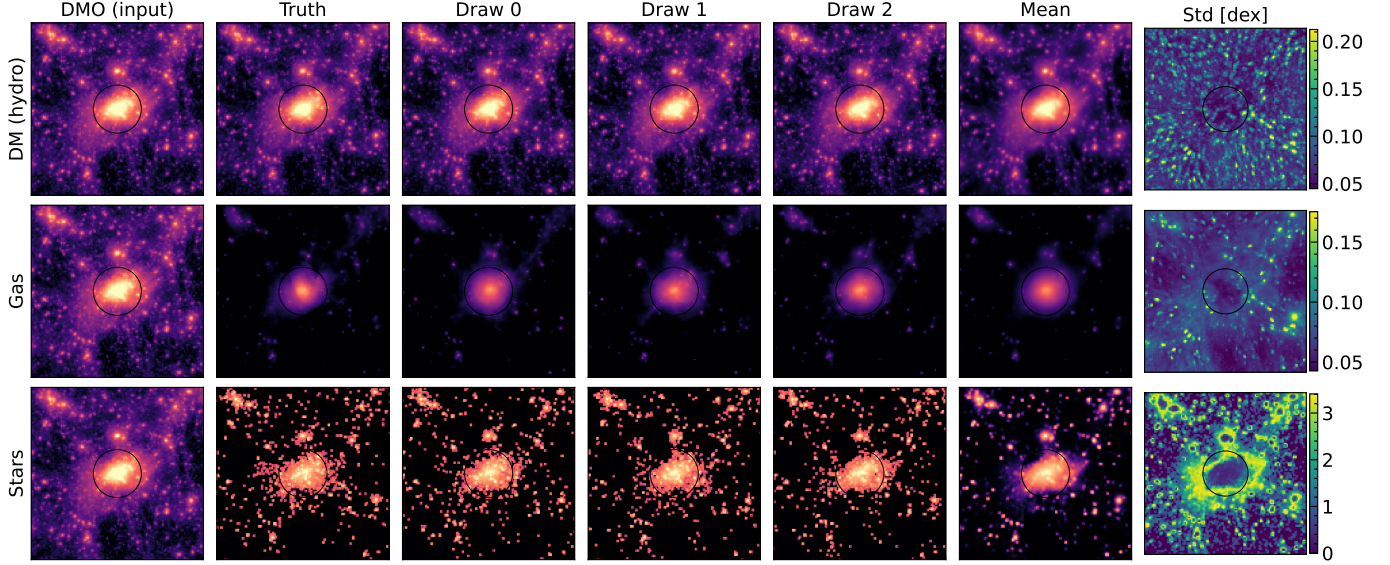


Figure 2. We show various generative draws for a halo. The *first column* shows the DMO conditioning for the halo, the *second column* shows the truth, and the *third-fifth columns* show individual draws with BIND, where the only change was the initial Gaussian field used in sampling. The *sixth column* shows the mean over 100 draws, and the *seventh column* shows the standard deviation over the 100 draws, in units of dex. The rows show the DM (hydro), gas, and stellar channels, respectively.

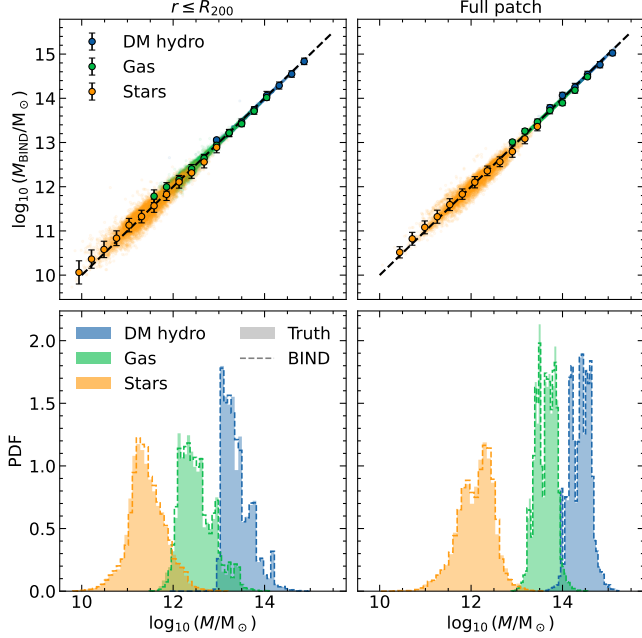


Figure 3. Per-halo integrated-mass comparison between the BINDed and true hydrodynamical halos for the stellar, gas, and dark matter channels, integrated within each halo’s R_{200c} (left column) and over the full $6.25 \times 6.25 h^{-1}$ Mpc patch (right column). Points show individual halos from the SB35 test, CV, and 1P suites. Black outlined points show binned medians with 1σ error bars. Lower panels compare the mass distributions of the true (solid) and generated (outlined) populations. The dashed line corresponds to perfect prediction.

the halo center and within the full 128×128 patch for each matter component. We then compute the relative error,

$$\frac{\Delta M}{M_{\text{hydro}}} = \frac{M_{\text{BIND}}}{M_{\text{hydro}}} - 1, \quad (14)$$

where M_{BIND} is the generated mass and M_{hydro} is the true mass from the hydrodynamical simulation.

In Fig. 3, we show the match between the stellar, gas, and dark matter channels integrated in both the radial (left panels) and full patch (right panels) apertures. All channels for each of the halos from the SB35 test set, CV set, and 1P set are shown as points in the upper panels, with binned medians and 1σ error bars shown as the black-outlined points. The lower panels show each channel’s mass distribution for the truth (solid) vs. the generated (dashed line). The agreement at this total-mass level for both radial and full patches is spectacular and shows that BIND can capture each channel’s mass distribution separately. We also see that the scatter in the predictions vs. the true values is smaller for the channels that have a higher correlation with the dark matter conditioning field. The DM-mass predictions are the most precise with respect to the true halos, while the stellar predictions have the largest scatter. Note that this plot does not show sensitivities to different parameters; instead, it combines all halos from all test simulations. In doing so, the presentation inherently shows the distribution of errors across parameter space under all parameter conditioning.

To expand this into a more parameter-sensitive comparison, Fig. 4 shows this relative difference for each

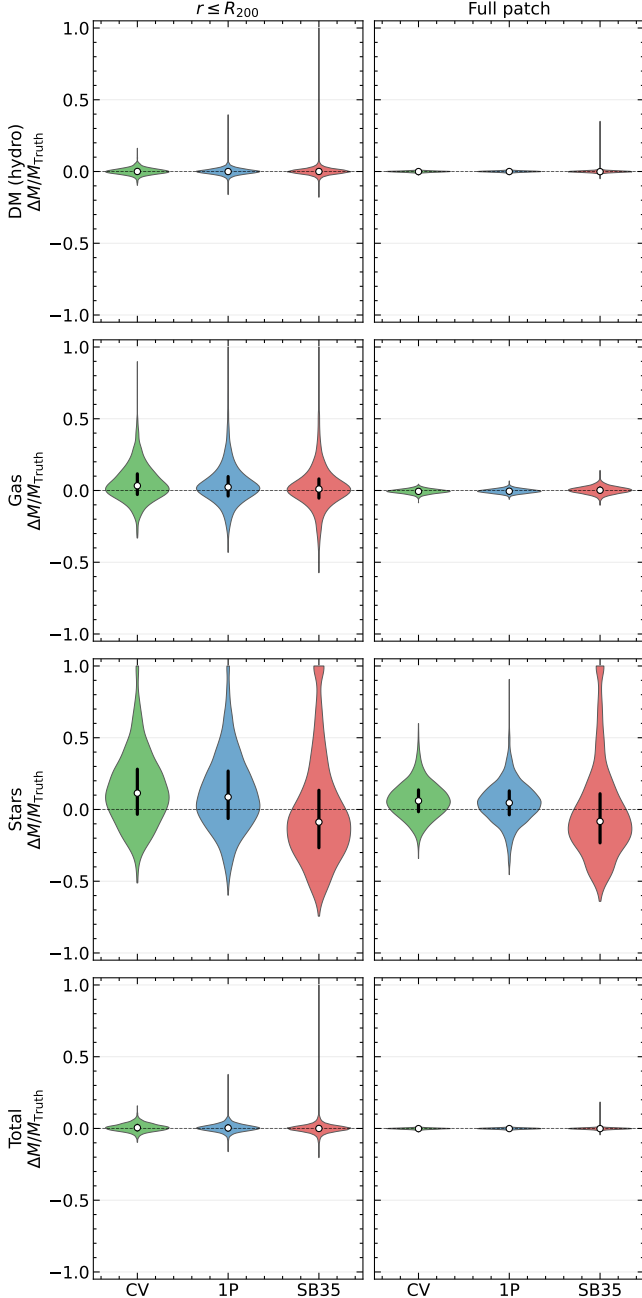


Figure 4. Relative error in integrated mass, $M_{\text{BIND}}/M_{\text{hydro}} - 1$, for each matter component (rows: dark matter, gas, stars, and total) and two apertures (columns: within R_{200c} and within the full 128×128 patch). We show results separately for the CV (green), 1P (blue), and SB35 test suites (red). Lines are centered on the median difference, and the extent marks the 25th and 75th percentiles. The violin shows the full distribution, which is mostly Gaussian for each residual, though the stellar channel deviates and includes a population of outliers. BIND recovers integrated masses across all components and test sets, with median errors below 1% for dark matter and total mass, $\sim 1\%$ for gas, and $\sim 10\%$ for stars. We show the bias and scatter values in Table 4.

Table 4. Bias and scatter of the integrated-mass residual $\Delta M/M_{\text{Truth}} \equiv (M_{\text{BIND}} - M_{\text{Truth}})/M_{\text{Truth}}$ summarized from Fig. 4. Bias is the per-halo median, and scatter is reported as half the distance between the 16th and 84th percentiles. Residuals are clipped to $[-1, 1]$ before summarizing, matching the violin display.

Channel	CV	1P	SB35
	bias / scat. [%]	bias / scat. [%]	bias / scat. [%]
<i>$r \leq R_{200c}$</i>			
DM (hydro)	+0.07 / 1.94	-0.02 / 1.88	+0.02 / 1.76
Gas	+3.42 / 10.60	+2.24 / 10.95	+1.15 / 10.62
Stars	+10.51 / 23.39	+8.38 / 24.66	-8.95 / 32.42
Total	+0.48 / 2.45	+0.30 / 2.33	-0.00 / 2.24
<i>Full patch</i>			
DM (hydro)	-0.04 / 0.45	-0.03 / 0.46	-0.09 / 0.54
Gas	-0.66 / 1.52	-0.52 / 1.45	+0.16 / 2.42
Stars	+6.20 / 11.10	+4.17 / 12.44	-8.26 / 27.26
Total	-0.09 / 0.45	-0.08 / 0.45	-0.12 / 0.53

matter component, computed independently for each test set, and Table 4 shows the percent of bias and scatter for each of the channels and suites. Across all three sets, the predicted integrated masses remain unbiased and have small scatter. The median relative error for the hydro dark-matter and total-mass channels is sub-percent, while gas and stellar masses are recovered at the $\sim 1\%$ and $\sim 10\%$ level, respectively. Enlarging the aperture from R_{200c} to the full patch (right column) produces little change in the bias but noticeably reduces the scatter, consistent with pixel-level noise averaging down over larger areas. As we will show in § 4.3, the majority of the stellar mass is determined by the central pixels’ predictions, which host the brightest central galaxies of our halos. So it stands to reason that our 10% stellar-mass bias within R_{200c} persists even as we increase the aperture to the full patch. The gas channel, however, shows that for each suite, the integrated masses are sub-percent accurate.

This serves as the most fundamental test of BIND. For virtually all halos across a parameter space entirely unseen during training, the model has learned to correctly partition the mass of the N -body conditioning halo among the dark matter, gas, and stellar components across a variety of cosmologies and astrophysical parameter variations. This implies that the network has learned the baryon fraction and stellar-to-halo mass relation in some implicit sense, which we investigate further in Section 6.

4.3. Radial Profiles

Rather than integrating within a fixed aperture, we now examine BIND’s accuracy as a function of pro-

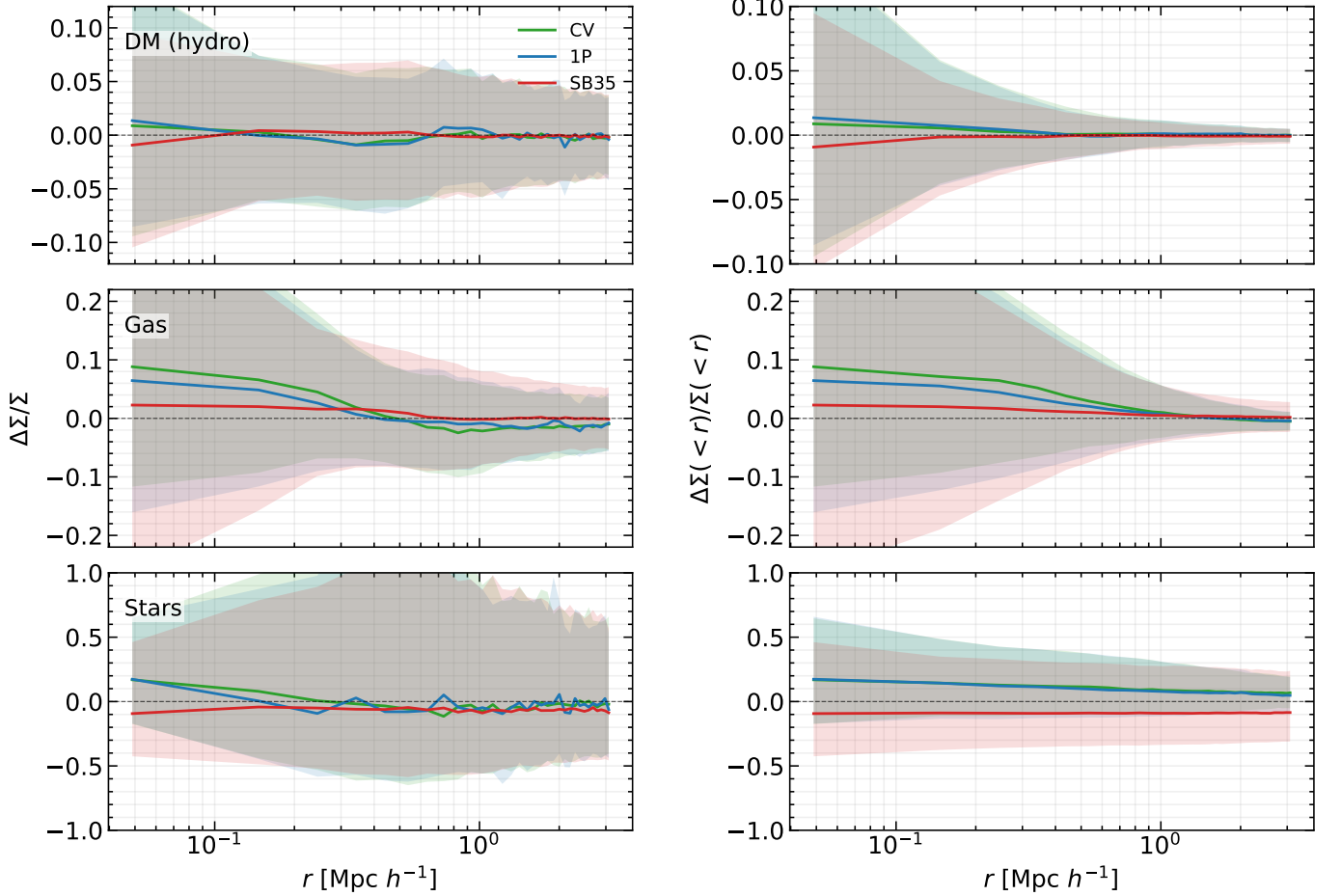


Figure 5. *Left column:* mean fractional error in the azimuthally averaged projected surface density profile, $\langle \Sigma_{\text{BIND}} / \Sigma_{\text{hydro}} - 1 \rangle$, and *right column:* cumulative surface density profile, $\langle \Sigma_{\text{BIND}}(< r) / \Sigma_{\text{hydro}}(< r) - 1 \rangle$ as a function of projected radius for each matter component and test set (CV: green; 1P: blue; SB35 test: red). Profiles are computed in 19 logarithmically spaced annuli out to $r = 3.125 h^{-1}$ Mpc. Shaded bands indicate the halo-to-halo scatter (16th–84th percentile) instead of the error on the median to reflect the intrinsic halo-to-halo scatter, and the y -axis is scaled differently for each quantity in order to show its full range. BIND achieves $\lesssim 10\%$ accuracy at all radii across all three test sets, with the largest deviations confined to the innermost bins where AGN feedback and steep stellar cusps are hardest to reproduce.

jected radius. For each halo, we compute the azimuthally averaged projected surface density profile and cumulative surface density profile in 19 logarithmically spaced annuli extending to the edge of the patch ($r = 3.125 h^{-1}$ Mpc), and, similar to the integrated mass, compute the mean fractional error

$$\langle \Delta \Sigma(r) \rangle = \left\langle \frac{\Sigma_{\text{BIND}}(r)}{\Sigma_{\text{hydro}}(r)} - 1 \right\rangle, \quad (15)$$

averaged over all halos in each test set.

Fig. 5 shows this fractional error for both the marginal (left column) and cumulative (right column) profiles for each matter component and test set. BIND achieves $\lesssim 10\%$ accuracy across all radii and all three evaluation sets. The largest deviations are confined to the innermost radial bins, which span $r < 97.7 h^{-1}$ kpc and contain only two pixels at our $48.8 h^{-1}$ kpc resolution,

where the stellar surface density profile is steepest, and the model finds it most difficult to learn. Comparing the left and right columns, the gas distributions, while overpredicted in the cores at the 10% level for the CV and 1P sets, converge on the correct cumulative masses as the profile approaches the virial radius. Conversely, the stars, which are overpredicted in the central pixels but seemingly well predicted outside of this region, are the sole cause of the stellar mass bias that we observed in Fig. 4.

4.4. Projected Halo Shapes

The previous tests explore integrated pixel quantities that are axisymmetric or cumulative. In this section, we explore whether BIND also reproduces the *shapes* of the baryonic fields. This is a stronger test of BIND as it requires the model to learn to partition mass across

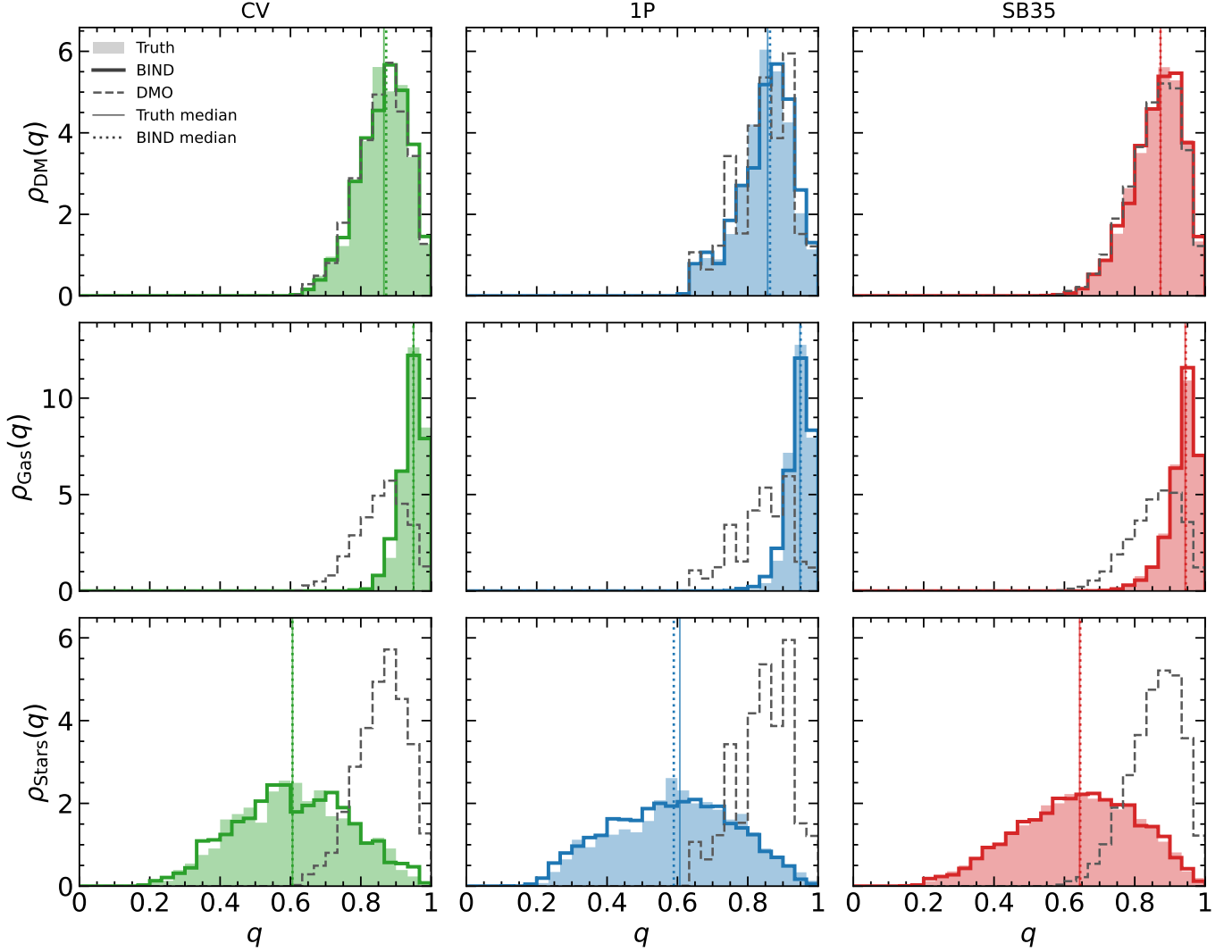


Figure 6. Distributions of the projected axis ratio $q = \sqrt{\lambda_{\min}/\lambda_{\max}}$ measured within R_{200c} . Rows show the dark-matter, gas, and stellar channels, while columns show the CV (green), 1P (blue), and SB35 test (red) suites. The filled histograms are measured from the true CAMELS simulations, and the outlines are measured from BINDed halos. We also show the DMO conditioning field in gray dashed, with its single q distribution repeated in every row as the no-learning baseline. Thin solid and dotted vertical lines mark the truth and BIND medians. In the dark-matter row, the DMO baseline already lies on top of the truth, so the agreement largely comes from the conditioning field rather than learning. In the gas rows, the DMO field is more elliptical; in the stellar fields, it is more round. BIND matches the true medians to ≤ 0.007 (dark matter), ≤ 0.002 (gas), and ≤ 0.017 (stars).

the field asymmetrically. It is also interesting to explore shape measures here, which are directly relevant to weak-lensing intrinsic alignment studies (e.g., M. A. Troxel & M. Ishak 2015; B. Joachimi et al. 2015; M. E. Lee et al. 2026c).

For each halo in the test and generated catalogs, we measure the projected 2D shape by computing the intensity-weighted inertia tensor within R_{200c} following the common approach in galaxy ellipticity measure-

ments and intrinsic alignment studies,

$$Q_{ij} = \frac{\sum_{\text{pix}} (r_i - \bar{r}_i)(r_j - \bar{r}_j) \sigma}{\sum_{\text{pix}} \sigma}, \quad (16)$$

where σ is the projected surface density (mass per pixel) and the sum runs over all pixels within R_{200c} . The eigenvalues $\lambda_{\max} \geq \lambda_{\min}$ of Q_{ij} define the projected axis ratio and ellipticity,

$$q = \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}} \epsilon = \frac{1 - q}{1 + q}, \quad (17)$$

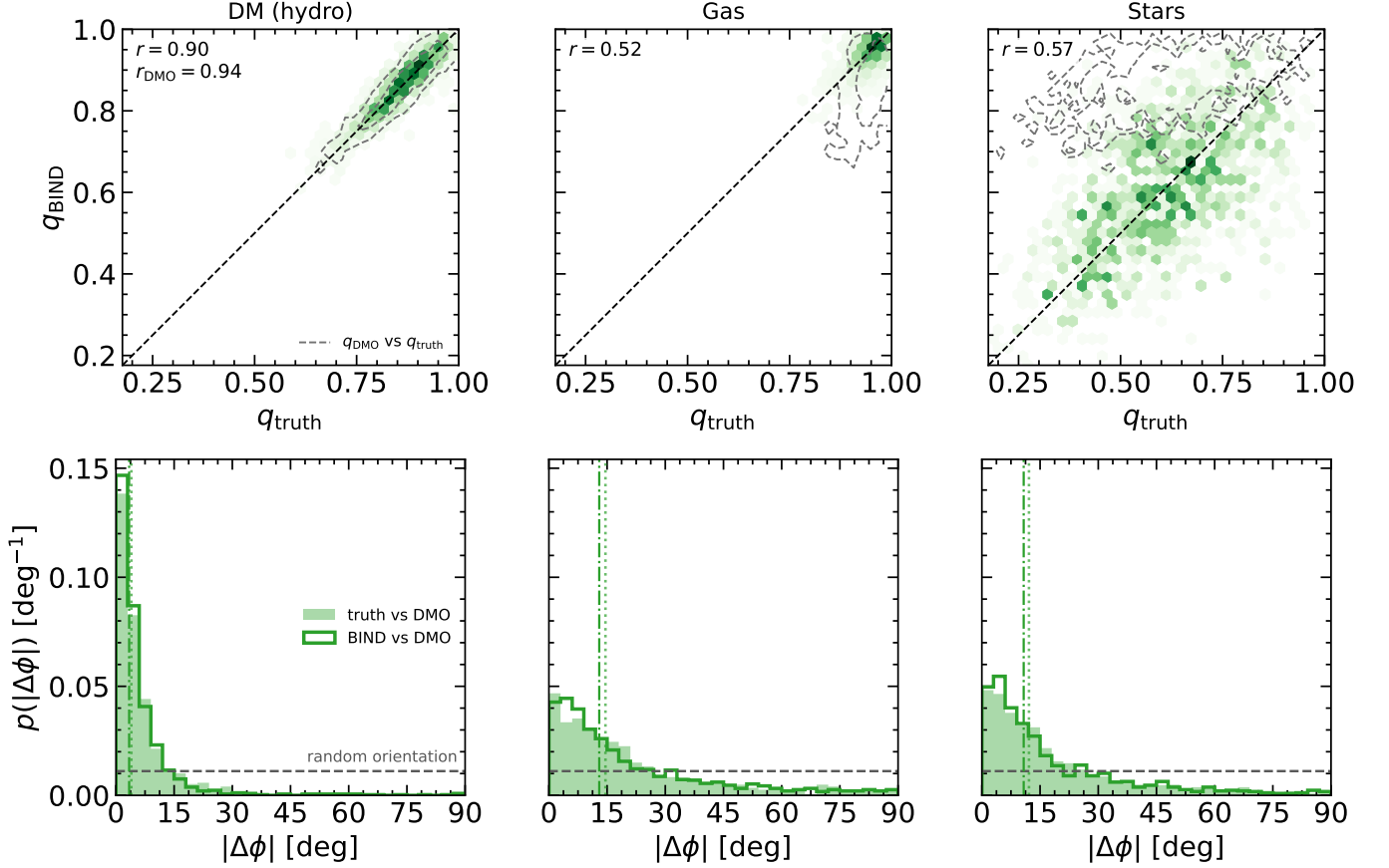


Figure 7. Per-halo shape fidelity measured in the CV suite. *Top row:* generated versus true axis ratio per channel, shown as a 2D density with the 1:1 line, and a black dashed contour showing the DMO baseline q_{DMO} versus q_{truth} . The Pearson correlation is shown for the comparison. Note that the DMO baseline hugs the 1:1 line for dark matter but sits below it for gas and far above it for stars, implying that the conditioning field is more elliptical than the gas and more round than the stars. *Bottom row:* distribution of the major-axis misalignment angle $|\Delta\phi|$ between the generated and DMO axes (solid line) compared to that between the true and DMO axes (filled histogram). The medians are marked by vertical lines (truth dotted, BIND dash-dotted), and the flat random-orientation expectation (median 45°) is shown for reference.

where $q = 1$ corresponds to a perfectly circular projected profile and $q \rightarrow 0$ to a maximally elongated one.

Fig. 6 shows the resulting q distributions for each channel and test set, with the true histograms in solid and the BIND predictions as unfilled histogram lines. The generated histograms match the truth in median (to ≤ 0.007 for dark matter, ≤ 0.002 for gas, and ≤ 0.017 for stars) and in width across all three channels and evaluation sets. BIND also captures the clear ordering among the channels, which hydrodynamical simulations more generally show (M. Velliscig et al. 2015). The dark-matter maps have median $q \simeq 0.87$, reflecting the mildly elliptical projected shapes of group-to-cluster-mass halos. The gas channel is more strongly concentrated toward circular configurations, consistent with the intra-cluster medium’s tendency to thermalize and fill the halo potential isotropically, and the stellar channel shows the broadest distribution with the largest ellipticities, driven by the compact, asymmetric morphologies of brightest

cluster galaxies. Notably, the SB35 test set, which spans the full 35-dimensional parameter space, shows no appreciable shift in median or spread relative to the CV halos at the fiducial parameters, suggesting that halo shape is determined primarily by the dark-matter structure and is largely insensitive to the subgrid physics parameters over the ranges explored here.

Instead of looking at the distribution of shapes over the entire set of simulations, Fig. 7 correlates the shapes halo-by-halo in the top row. The dark-matter axis ratio is recovered at $r = 0.90$ per halo, but we also find that the DMO conditioning field alone achieves $r = 0.94$. We conclude that the dark-matter shape agreement is inherited from the input, so the network has learned to leave the DM field mostly alone. Comparing the gas and star channels, though, we see that the DMO baseline is structurally wrong. We show the halo-by-halo correlation of axis ratio for the DMO halo vs. the gas and stellar halos in black dashed contours, which show clear biases for

the gas and stellar channels, compared to BIND, which obtains a relatively tight correlation for the predicted stellar and gas fields. Computing the median ellipticity for the DMO fields and the gas and stellar fields following Eq. 17, we find that the dark matter is ~ 2.8 times more elliptical than the gas ($\epsilon_{\text{DMO}}/\epsilon_{\text{gas}}$) and ~ 3 times rounder than the stars ($\epsilon_{\text{star}}/\epsilon_{\text{DMO}}$). In these channels, BIND recovers the per-halo axis ratio at $r = 0.52$ and 0.57 , respectively, well beyond what the conditioning field supplies.

We also show the orientations of the different fields with respect to the dark-matter-only conditioning (major-axis misalignment angle) in the bottom panels of Fig. 7. This is motivated by the literature studying the misalignment angle between the dark-matter-only halo and the baryonic components. Measured against the DMO axis, the generated baryons decorrelate from the dark-matter geometry by the same amount as the true baryons do (median $|\Delta\phi| = 13^\circ$ versus 15° for gas and 11° versus 12° for stars). From this, we conclude that the model genuinely learns the orientation of the individual components and their decorrelation from the DMO conditioning. Moreover, this indicates that BIND can be used to investigate the misalignment angle between dark matter and hydrodynamical components, and potentially whether they vary as functions of astrophysics or cosmology.

We close this subsection with three caveats when interpreting these results. First, we measure shapes from projected mass maps, whereas observational shape measurements are made in flux. A rigorous comparison with observational intrinsic alignment measurements would require modeling the halo SED and computing shapes from mock photometric images, or directly modeling the luminosity of stars in BINDed halos (e.g., as in M. E. Lee et al. 2026c). Second, our pixel scale of $\sim 50 h^{-1}$ kpc is comparable to the effective radius of the brightest central galaxies at these halo masses, so the central stellar component is effectively unresolved and occupies roughly a single pixel. Third, we project along the full $50 h^{-1}$ Mpc line of sight, so foreground and background mass contribute to both the true and generated images, diluting the intrinsic halo ellipticity signal.

Intrinsic orientation alignments of galaxies are a key systematic in upcoming weak lensing analyses, and current models exclude dependencies on baryonic effects. The above caveats notwithstanding, the goal of this section is not to present BIND as a model for intrinsic alignments (yet). Instead, it shows that the generative model can capture the intrinsic shapes of these halos across the parameter space, with generated and true shapes in close agreement. This implies that BIND has

learned the morphological mapping from dark-matter-only halos to their baryonic counterparts, a non-trivial result that potentially opens the door to large-scale intrinsic alignment emulation in the future.

5. FIDELITY ACROSS THE PARAMETER SPACE

The previous section established that BIND reproduces the baryonic fields of individual halos when compared with the same halos in our three test sets. We now turn to the second question posed in the introduction. *Has BIND learned how those fields change across the IllustrisTNG parameter space?* We test this at three levels of resolution (integrated masses, radial profiles, and the fields themselves) and close with the redshift dependence.

5.1. Parameter Response of Integrated Masses

To measure how generated summaries change with parameters, we compute the Spearman rank correlation coefficient between the quantities explored in the previous section and the model parameters. In Fig. 8 we show this correlation coefficient between the integrated mass in each component and each of the galaxy formation elements of θ_{TNG} , for both the true and generated halos at identical parameter-space locations. To truly gauge the effect of feedback on these quantities, we exclude the ΛCDM parameters from the correlation analysis, though they are also captured with high fidelity. We use the parameter names defined in Table 1 of S. Genel et al. 2026, which also provides the exact definitions and citations to the papers in which they are introduced. In Fig. 8, we group the parameters by parameter type, indicated by the color and the names at the top of the plot. In each group, we organize the columns by the maximum correlation observed in the stellar channel, which appears most sensitive to the parameters. The top rows show the true correlation, the central rows show BIND's predicted correlation, and the third row shows the residual for each channel.

Across the 102 SB35 test simulations, the true and generated correlations agree, with a root-mean-squared (RMS) residual of 0.014 across all channels and parameters, dominated by deviations in the stellar channel. The largest deviation is $|\Delta\rho_S| = 0.08$, which appears in a few of the parameters. The strongest true responses are recovered almost perfectly. For example, stellar mass against the energy of stellar winds A_{SN2} gives $\rho_S = -0.50$ (true) versus -0.50 (BINDed). The dark matter correlations are perfect for all parameters, and the gas correlations are within 0.02 for all parameters.

Fig. 8 shows that the stellar mass channel is the most sensitive to the parameters. This makes intuitive sense,

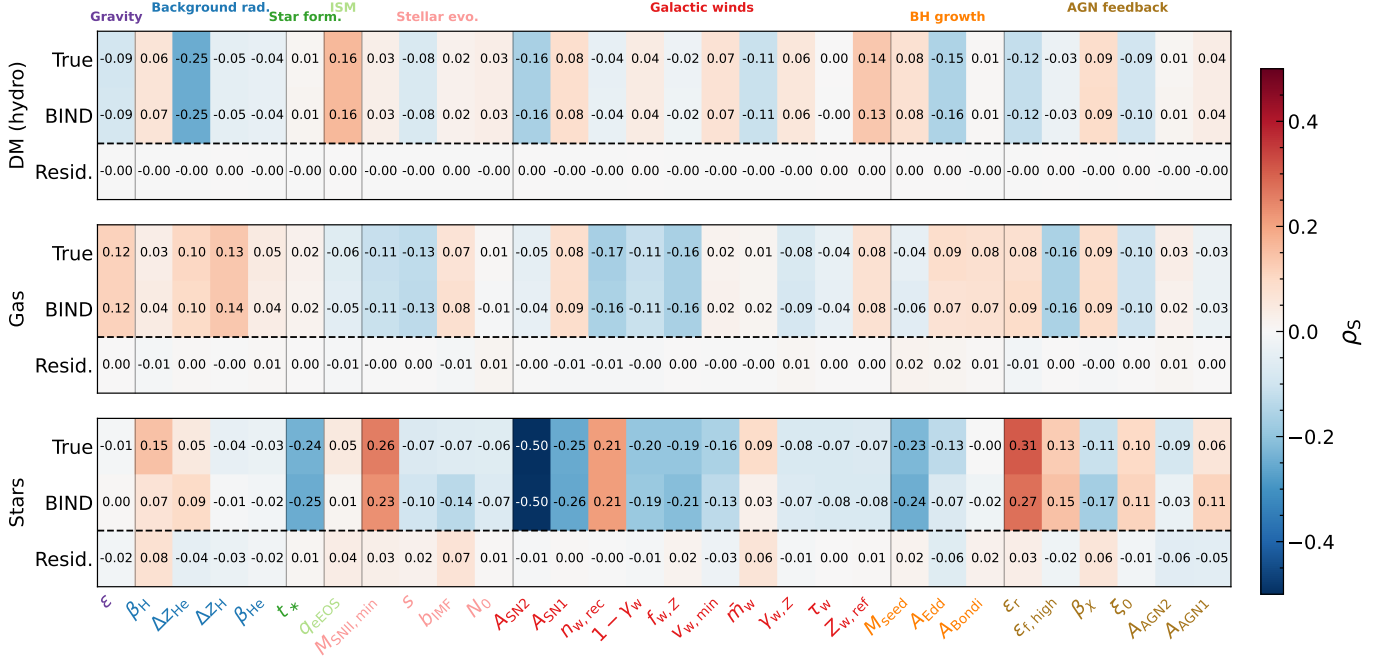


Figure 8. Spearman rank correlation between SB35 test set integrated masses per component and each of the astrophysical parameters. The top rows show the correlations measured in the true halos, while the middle shows those measured from the BIND halos, and the bottom rows show the residuals between the two. We group each parameter by type following the conventions of Table 1 of S. Genel et al. 2026, and within each group, we sort parameters by their largest absolute correlation value. The strong agreement between truth and BIND demonstrates that the model has correctly learned the dependence of baryonic mass on the IllustrisTNG subgrid parameters. The stellar channel (stars) exhibits the strongest parameter sensitivity, followed by gas, while the dark-matter channel shows little response to astrophysical parameter variations.

as changing the parameters directly influences the star formation rate within the halos. For example, increasing $A_{\text{SN}2}$ acts as preventative feedback, starving the central black hole of gas and disrupting star formation. As a result, we observe a negative correlation between stellar mass and the parameter. Stellar mass is then most dependent on the model parameters, whereas the gas channel shows intermediate sensitivity, as it correlates with the dark-matter structure yet remains responsive to variations in stellar and AGN feedback. The dark-matter channel shows little response to the astrophysical parameters. Gravitational collapse determines most of the dark-matter distribution and resulting halo masses, which are ultimately controlled by cosmological parameters such as the total matter density Ω_m and the variance of fluctuations σ_8 . The feedback parameters affect the stellar and gas distributions, which in turn back-react on the dark matter distribution (M. Gebhardt et al. 2026); however, this effect is much smaller than gravity.

5.2. Parameter Response of Radial Profiles

Fig. 9 extends the Spearman correlation analysis of Section 5.1 to the azimuthal radial profiles. For each annular bin and each matter component, we compute the

Spearman correlation between the surface density and each element of θ_{TNG} , for both the true and generated halos. Here, we show the correlation for BIND in the left panels and the residual in the right panels. Over the SB35 test set, the radially resolved grid is reproduced with an RMS residual of 0.029, indicating that BIND has correctly learned the spatially resolved dependence of baryonic structure on the IllustrisTNG subgrid parameters beyond just the integrated masses.

As with the integrated mass comparison of Fig. 8, the stellar channel shows the strongest parameter sensitivity, particularly in the central regions of the halo, where the stellar surface density is highest, and AGN feedback most directly regulates the star formation rate. There is also a clear anticorrelation between the stellar and gas density responses for several feedback parameters (such as $A_{\text{SN}1}$ and $A_{\text{SN}2}$). Parameters that drive higher stellar densities tend to suppress the gas density at similar radii, and vice versa. This reflects the competition between gas cooling and star formation on the one hand and stellar- and AGN-feedback-driven outflows on the other, a balance that BIND appears to have learned from the training data. A more detailed investigation into the parameter dependencies and effects on galaxy group gas and stellar properties is saved for future work,

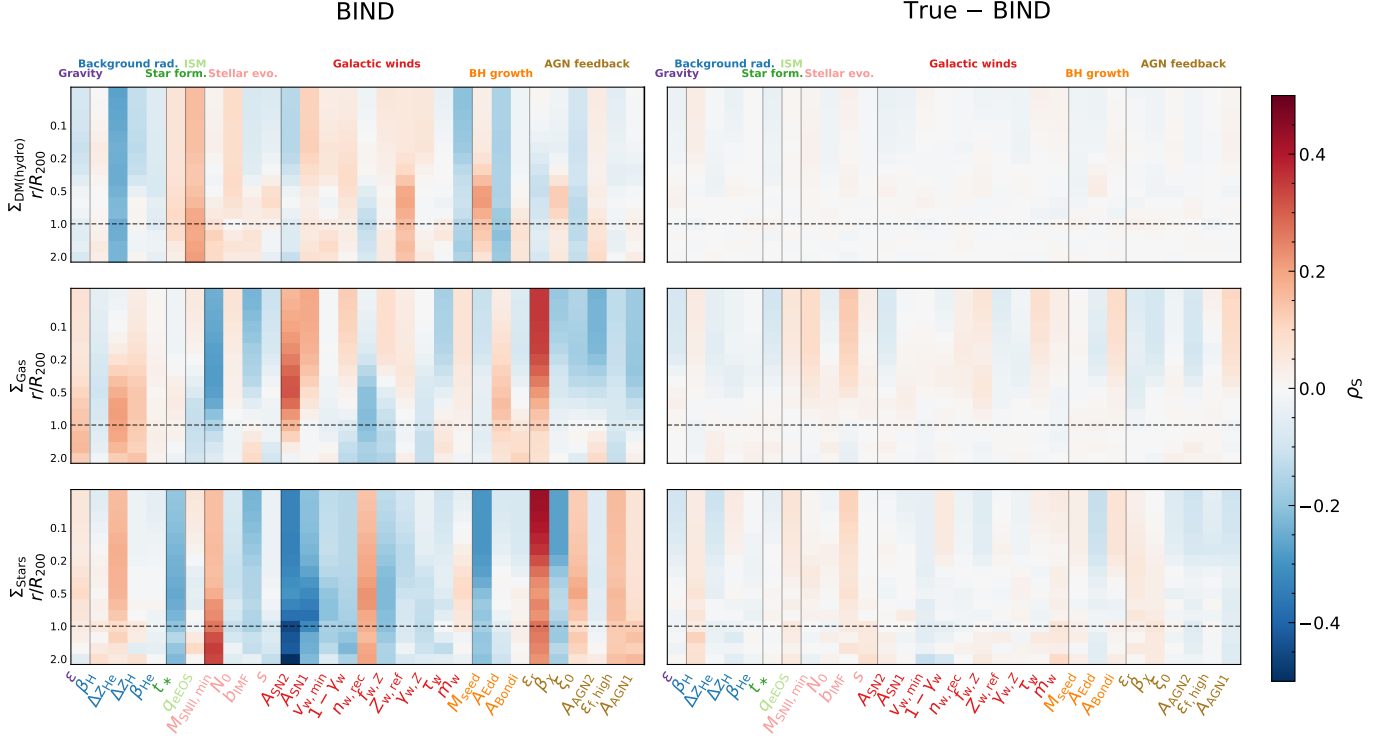


Figure 9. Spearman rank correlation between the azimuthally averaged surface density profile and the 30 astrophysical parameters as a function of radius r/R_{200c} in 19 bins, computed over the 102 SB35 test simulations. For each channel, the left column shows the BIND correlations and the right column the residual (True – BIND). Again, similar to Fig. 8, the stellar channel is most parameter-sensitive; however, we now see that this is localized mostly to the central regions of the halo.

as the goal here is to show that BIND is a tool that can be used to decipher this, given that it accurately captures the parameter dependencies of both the integrated masses and the radial profiles.

5.3. Field-Level Parameter Response

One of the hopes for BIND is that, unlike radial baryon correction models, it captures the asymmetric response of fields to changes in the underlying galaxy-formation parameters. To test this, we use the 1P test set, in which we vary a single IllustrisTNG parameter at a time while holding all others at their fiducial values and keeping the initial conditions fixed. For each parameter, we select a halo at the two prior bounds (the minimum and maximum parameter values) and compute the signed field-level residual

$$\delta M(\mathbf{x}) = M_{\text{high}}(\mathbf{x}) - M_{\text{low}}(\mathbf{x}), \quad (18)$$

where M_{high} and M_{low} are the projected mass maps at the upper and lower parameter bounds, respectively. Because the initial conditions are identical across the 1P simulations, any structure in δM is attributable to

baryonic processes alone rather than cosmic variance⁸. We compute this difference for each of the three output channels (dark matter, gas, and stellar mass) for both the true hydrodynamical fields and the BINDed counterparts and show the results in Fig. 10.

The left two columns of Fig. 10 show the dark matter residuals following Eq. 18 for both the true and generated fields. In each panel, we apply 1-pixel Gaussian smoothing to bring out the shapes of the effects. We find that the dark-matter residual maps are largely consistent with small pixel-level shifts between the parameter bounds, likely corresponding to subhalo positional shifts, and show no coherent large-scale structure. This is expected, as dark matter responds primarily to gravity, and, at fixed initial conditions, the dark-matter distribution is nearly insensitive to subgrid parameters over the mass and redshift range probed here. We note that this insensitivity is not exact. Baryonic processes do back-react on dark matter, and *M. Gebhardt et al. (2026)* measure this effect across the CAMELS suites,

⁸ However, there is a noise floor that is imposed by the butterfly effect in *S. Genel et al. (2019)*. Comparing the butterfly-effect floor to these residuals would require understanding how the effect changes with model parameters, which we save for future work.

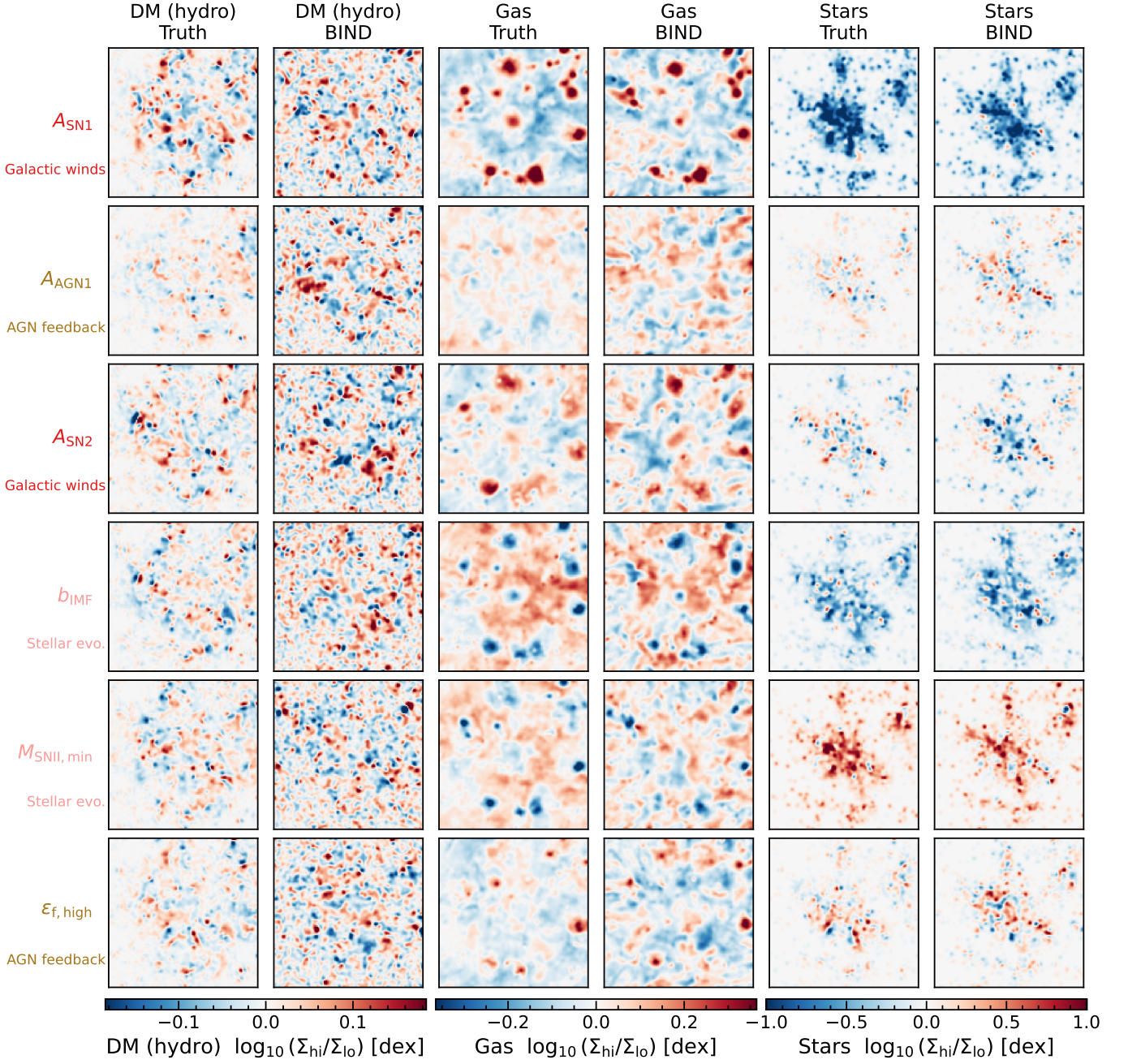


Figure 10. Field-level parameter response for the most massive 1P halo shown as the $\log_{10}$ difference between parameters at their upper and lower prior bounds. Each row is one IllustrisTNG parameter, and the six columns pair truth and BIND for the dark-matter, gas, and stellar channels. Each panel shows the response $\delta\mathbf{M}$ of Eq. 18 between the parameter’s prior maximum and minimum on a symmetric per-channel color scale. The dark-matter responses are consistent with noise in both truth and BIND, as expected. For the wind and IMF parameters (A_{SN1} , A_{SN2} , IMFslope , SNII_MinMass_Msun) the generated gas response reproduces the sign and morphology of the truth (pixel-level $r = 0.3\text{--}0.6$), and for the AGN parameters, the true single-halo response is weaker than the stochastic floor of a single generated sample, and the corresponding BIND panels are noise-dominated.

finding a real but small redistribution compared to the gas response. Our result is consistent with that picture, in which any back-reaction present is at or below the noise level of a single generated sample. The back-reaction is also likely affecting the dark matter at pixel

scales ($\sim 50 \text{ kpc } h^{-1}$), causing noisier patterns than the stronger gas and stellar redistribution driven by the parameters. We defer a detailed study to future work.

The middle two panels of Fig. 10 show the gas residuals, which contain rich, spatially coherent structure

varying significantly across the parameters. For example, in A_{SN1} , the cores of both the central and exterior halos become more condensed as the parameter increases, while the halo exteriors are suppressed. This is in line with expectations from M. E. Lee et al. (2024), who have found that an increase in A_{SN1} in massive halos starves the black hole, preventing it from ejecting matter far outside the central galaxy.

In contrast, a shallower IMF slope allows more massive stars to form, leading to higher metallicity, cooling, and black hole growth. The core density of the halos is depleted, but feedback pushes gas further out of the halos, as shown by the red rings. Similarly, the stellar fields in the rightmost panels show that changes to the feedback parameters have net positive or negative impacts on total star formation across the fields, and that this response is consistent with changes in gas. Taken together, this figure demonstrates that BIND has encoded the directional dependence of baryonic processes at the field level.

BIND recovers the major trends across these parameters with good fidelity. The generated residual maps share the sign, spatial morphology, and approximate amplitude of the true residuals in both the gas and stellar channels, capturing features that span a dynamic range of order 10^1 – 10^2 in surface density. The main discrepancy is that the BIND residuals are somewhat noisier than the ground truth. Small-scale features in the true maps that arise from the deterministic response of the same halo to a parameter change are partially washed out in the generated fields by the stochastic nature of the sampling. This is not unexpected, as the model generates independent samples from the learned posterior for each parameter value rather than computing a deterministic perturbation. The residual is therefore noisy at the single-sample level even if the mean response is correct. Still, the bulk spatial structure is preserved, and the qualitative parameter responses are clearly recognizable. This provides strong evidence that BIND has learned the mapping between the IllustrisTNG subgrid parameters and their field-level baryonic signatures, beyond what could be inferred from integrated-mass or azimuthally averaged statistics alone.

5.4. Redshift Dependence

Up until this point, we have presented results only at a single redshift, $z = 0$. However, throughout this work and in training, we treat the redshift as a model parameter that BIND is trained on. Especially for our companion paper, M. E. Lee et al. (2026a), where an entire lightcone is BINDed out to redshift $z = 2.44$, we must validate that the redshift conditioning works

as intended. We focus on whether BIND retains model fidelity at each redshift in isolation and whether it has learned the evolution of the baryonic fields with redshift.

Fig. 11 repeats the integrated-mass test of Fig. 4 at each of the eight redshifts, reported per set. The dark matter and total mass are accurate and flat at every epoch (medians within +0.3–1.1% and +0.2–2.6% over $0 \leq z \leq 2$), and the held-out SB35 suite shows the least evolution with redshift in every channel. One interesting trend is that the 1P set appears to degrade for the stellar channel as redshift increases. Given that the 1P set has all parameters fixed to the fiducial values, it makes sense that the CV and 1P lines track each other, but at higher redshifts, the effect of individual parameters on the stellar distribution is clearly degrading. The 1P set inherently probes the edges of the parameter prior where BIND has fewer training samples, which is the likely cause of this trend.

We also verified directly that the conditioning is used and that it interpolates. Holding the dark-matter conditioning, the parameter vector, and the initial ODE noise fixed while varying only the scale-factor label produces a monotonic response, and evaluations at five redshifts that correspond to no training snapshot ($z = 0.10$ – 1.75) fall on the same smooth monotonic curve. This implies that the model has learned a continuous function of the scale factor rather than memorizing a discrete set of snapshots, which allows it to be used on lightcones and across redshift. Taken together, these tests validate the redshift conditioning over $0 \leq z \leq 2.5$ spanned by the lightcones of the companion paper (M. E. Lee et al. 2026a).

6. EMERGENT JOINT STRUCTURE: COMPONENT CORRELATIONS

We now turn to the third question posed in the introduction. *To what extent does BIND capture correlations between physically distinct baryonic components?* We begin by examining the halo baryon fraction across all three test sets to probe whether the model correctly partitions mass among gas, stars, and dark matter as a function of halo mass and cosmic baryon fraction. We then focus on the 1,111 CV halos generated at the fiducial IllustrisTNG parameters to investigate a set of inter-component scaling relations: the stellar-to-halo mass relation (SHMR), the baryonic-to-halo mass relation, and the gas-to-stellar mass relation. Finally, we decompose the scatter around these relations and show that BIND correctly reproduces the joint distribution of halos that deviate coherently from multiple relations simultaneously, including the most pathological cases.

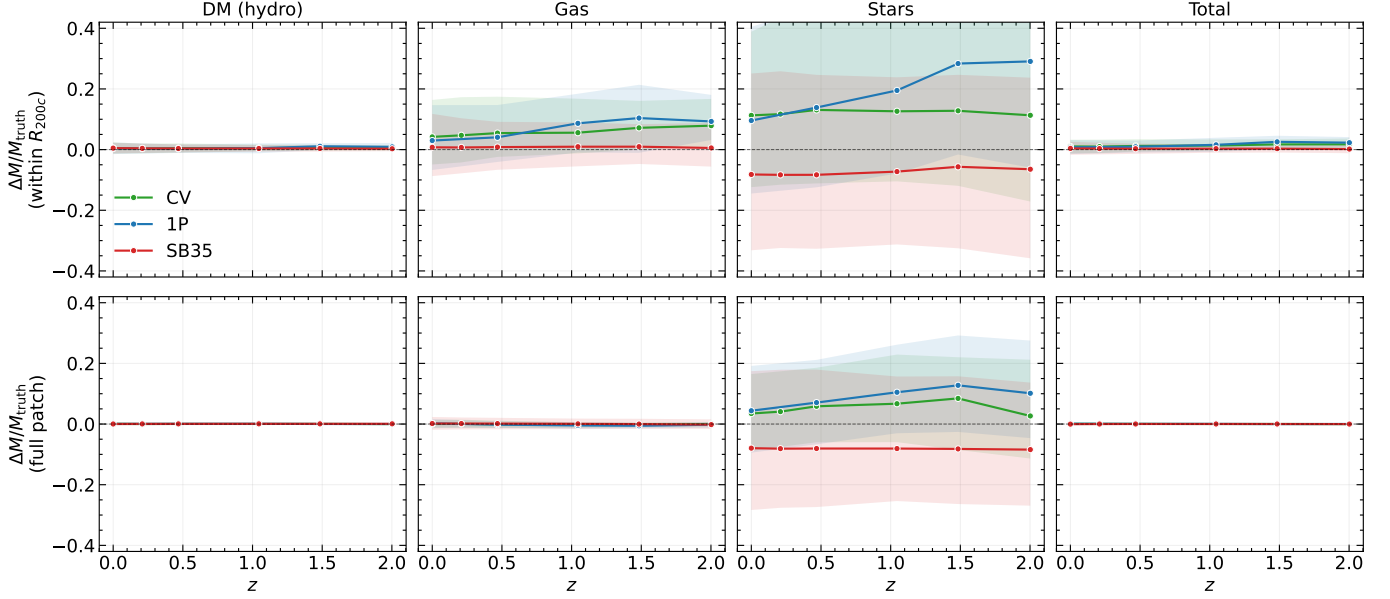


Figure 11. Relative error in integrated mass, $M_{\text{BIND}}/M_{\text{hydro}} - 1$, as a function of redshift, for each matter component and for the total (columns), measured within R_{200c} (top row) and over the full 128×128 patch (bottom row). Lines and points show the median over the halos of each suite at each snapshot, and shaded bands show the halo-to-halo 16th–84th percentile scatter. The suites are shown separately. Dark matter and total mass are flat and accurate everywhere, while the held-out SB35 suite is flat in every channel.

6.1. Baryon Fractions

For each halo in the three test sets, we compute the projected baryon fraction within R_{200c} as

$$f_b(< r) = \frac{M_{\text{gas}}(< r) + M_{\text{star}}(< r)}{M_{\text{total}}(< r)}, \quad (19)$$

and normalize by the cosmological baryon fraction $f_{b,0} \equiv \Omega_b/\Omega_m$ computed from the parameter tables of each simulation. We focus on $f_b/f_{b,0}$ as a function of halo mass because we expect groups to retain a smaller fraction of their baryons than clusters (e.g., [L. Lovisari et al. 2021](#); [B. Hadzhiyska et al. 2025](#); [J. Siegel et al. 2025](#)), and because the mass dependence of f_b is one of the quantities that varies strongly across subgrid models ([I. Medlock et al. 2024b](#); [E. T. Lau et al. 2024](#)).

We bin the halos in logarithmic M_{200c} intervals, computing the median normalized baryon fraction and the halo-to-halo scatter in each bin for both the true hydrodynamical halos and the BINDed counterparts. Fig. 12 shows the resulting $f_b/f_{b,0}$ vs. M_{200c} relation for each test set. The BINDed halos reproduce the median and scatter of the true relation with excellent fidelity across more than a decade in halo mass and across the diverse parameter coverage of the SB35 suite.

The baryon fraction involves the ratio of two separately generated fields (gas and stellar mass) to the total halo mass. We did not supply halo mass information as explicit conditioning to the network. Still, the model has learned the halo-mass dependence of the baryon frac-

tion, the sensitivity of f_b to the cosmic baryon fraction Ω_b/Ω_m , and the correct relative normalization between the gas, stellar, and dark-matter channels simultaneously. The close agreement between the true and generated relations across all three suites therefore reflects a non-trivial internal consistency in the generative model, similar to that shown in § 4.2.

6.2. Inter-component Scaling Relations

We next investigate whether BIND reproduces the joint structure of the baryonic mass components. For this analysis, we focus on the 1,111 CV halos generated from the 27 CV simulations. We measure three scaling relations within R_{200c} of each halo: the gas-to-stellar mass relation, $M_{\text{star}}-M_{\text{gas}}$, the SHMR ($M_{\text{star}} - M_{200c}$), and the baryonic mass $M_{\text{bar}} - M_{200c}$ relation (unnormalized quantity shown in Fig. 12). These relations, and the SHMR in particular, are among the most heavily studied summaries of the galaxy-halo connection (for a review, see [R. H. Wechsler & J. L. Tinker 2018](#)). Each relation is roughly described by a power law over the mass range, so we fit a simple linear model in log-log space for each relation and overlay the best-fit regression on both the true and BINDed halo populations.

Fig. 13 shows these three panels for the true halos (top) and BIND halos (bottom). The generated population reproduces not only the slope and normalization of each scaling relation but also the scatter amplitude and asymmetry about the best-fit line. Halos that

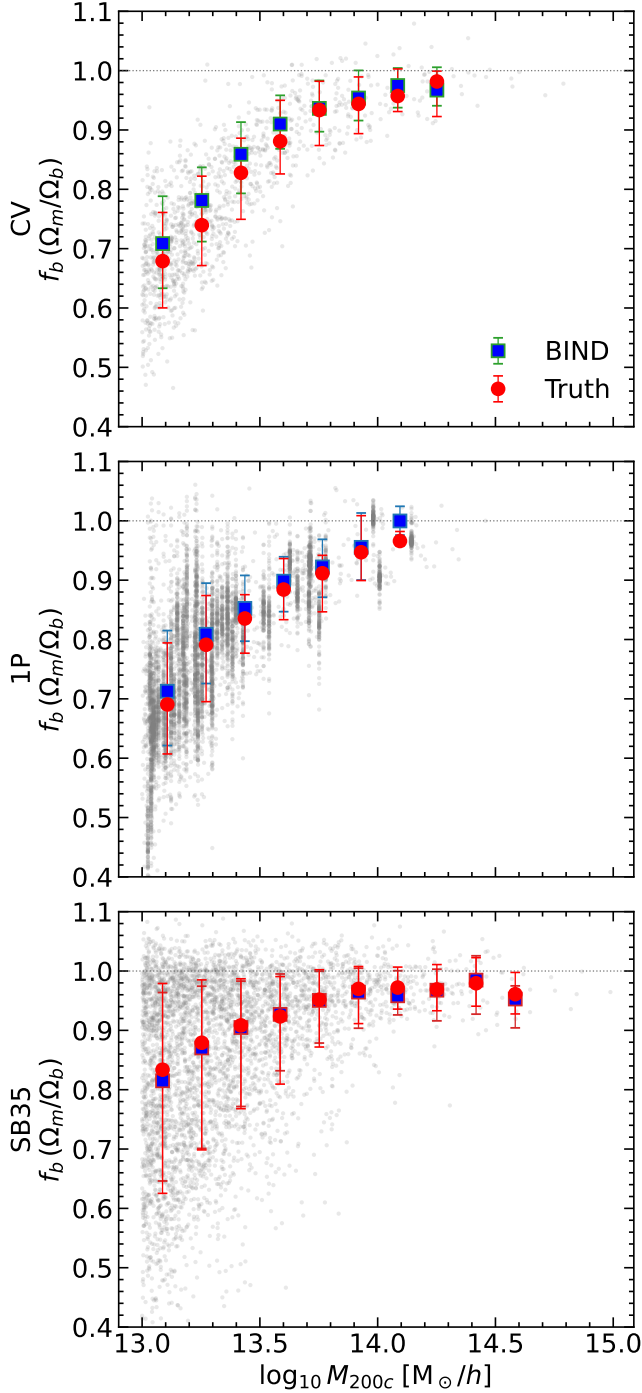


Figure 12. Normalized baryon fraction $f_b(< R_{200c})/f_{b,0}$ as a function of halo mass M_{200c} for the three test sets. Points show individual halos, while lines and error bars show the median and 16th–84th percentile scatter computed in logarithmic mass bins. *True* hydrodynamical halos are shown with black scatter points to help guide the eye. BIND reproduces both the median trend and scatter of the baryon fraction across the full range of halo masses and parameter values, despite halo mass never being used as a conditioning variable.

scatter significantly above or below the mean relation in the hydrodynamical simulations have corresponding analogs in the BIND output, indicating that the model has learned the physical correlations that drive halo-to-halo scatter in baryonic content at fixed mass. Scatter in these relations is driven by the assembly history of each halo, gas cooling, star formation, and feedback, none of which are directly encoded in the N -body input beyond the projected dark-matter distribution. That we can recover this almost perfectly suggests the model has, in some way, learned these non-linear effects.

6.3. Residual Correlations

To examine the degree to which scatter in different scaling relations is correlated, we identify halos that lie more than 1σ above or below the best-fit linear relation in each of the three panels of Fig. 13, classifying them as the “high” and “low” populations for each relation. For each pair of relations, we then visualize the joint distribution of the residuals $\Delta \log M_{\text{star}}$, $\Delta \log M_{\text{gas}}$, and $\Delta \log M_{\text{bar}}$ for both the high and low subsamples as a corner plot. We smooth the distributions with a Gaussian KDE and overplot the 1σ , 2σ , and 3σ contours. We apply the same procedure to the BINDed population and overlay it for direct comparison.

Fig. 14 presents this for the true and BINDed CV halos. The generated contours reproduce the ground-truth distributions across all residual pairs and for both the high and low populations. The halos that scatter high in the SHMR also tend to scatter high in the baryonic mass relation, reflecting a correlated excess of both stellar and total baryonic content. The halos that scatter low in the gas-to-stellar mass relation tend to scatter high in the SHMR, consistent with AGN-driven gas expulsion redistributing baryons out of the halo while leaving the stellar component largely intact. These correlations span multiple physically distinct processes, making them non-trivial for the network to recover from projected dark-matter and IllustrisTNG parameter-set inputs alone.

The close agreement between the truth and BIND across all of these joint distributions is perhaps the most demanding consistency test. It demonstrates that the model has not merely learned marginal distributions in each channel independently but has captured the full multivariate structure of the baryonic fields as a function of halo properties. This follows directly from BIND’s joint, multi-channel generative architecture, which produces all three mass fields simultaneously from a single sample of the learned probability path and naturally preserves inter-channel correlations.

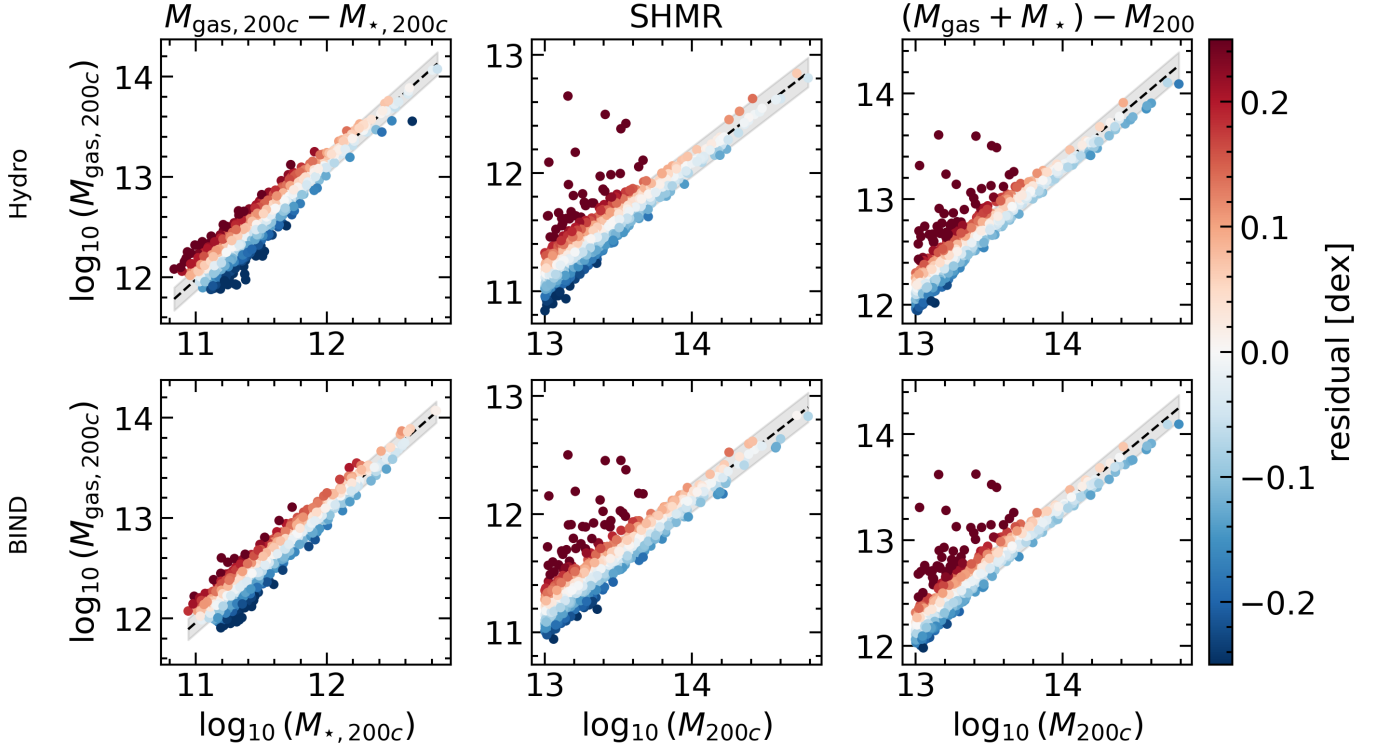


Figure 13. Inter-component mass scaling relations for the 1,111 CV halos at the fiducial IllustrisTNG parameters. Each column shows a different relation in log-log space: gas-to-stellar mass (*left*), stellar-to-halo mass (SHMR; *center*), and baryonic mass- M_{200c} (*right*). The *top* panels show the relations measured from the hydro simulations in the CV set, while the bottom panels show the generated BINDed realizations. Dashed lines show the best-fit linear regressions to the true and generated populations, with $\pm 1\sigma$ scatter indicated by shaded bands, and the colors represent the residuals of a given halo from the best-fit line. BIND accurately reproduces both the slopes and normalizations of all three relations as well as the amplitude and morphology of the scatter, indicating that the model has learned the joint covariance structure of the multi-component baryonic fields.

7. USING BIND IN THE WILD

One of the primary motivations for BIND, and the fourth question posed in the introduction, is to serve as a practical baryonification engine. *Given an arbitrary N -body simulation and a choice of IllustrisTNG galaxy-formation parameters, can BIND produce a baryonified matter field whose statistical properties match those of a true hydrodynamical run at that parameter point?* The key downstream observable we target is the projected matter power spectrum suppression,

$$S(k) = \frac{P_{\text{hydro}}(k)}{P_{\text{DMO}}(k)}, \quad (20)$$

which encodes the net impact of baryonic feedback on the matter distribution as a function of scale. In the companion paper, we extend this to non-Gaussian statistics (M. E. Lee et al. 2026a). Accurately reproducing $S(k)$ across the full IllustrisTNG parameter space is a prerequisite for using BIND as a baryonic correction tool in cosmological inference pipelines.

7.1. BINDing an N -body Simulation

To construct a full baryonified field from an N -body simulation, we apply BIND halo by halo and paste the generated patches back into a global projected mass map. The procedure proceeds in four steps.

1. **Halo extraction.** We read in the DMO particle catalog for each CV or SB35 test simulation and identify all halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ using the same halo finder and centering procedure described in Section 3.1. For each halo, we extract the 128^2 -pixel central patch.
2. **BIND generation.** We pass each extracted patch through BIND, conditioned on the cosmological and IllustrisTNG parameter vector θ_{TNG} of the parent simulation, and generate a single sample of the three-channel baryonic output (dark matter, gas, stellar mass).
3. **Replacement and pasting.** The N -body DMO projection of the full $50 h^{-1}$ Mpc box is first projected onto a 1024^2 grid. We then paste each

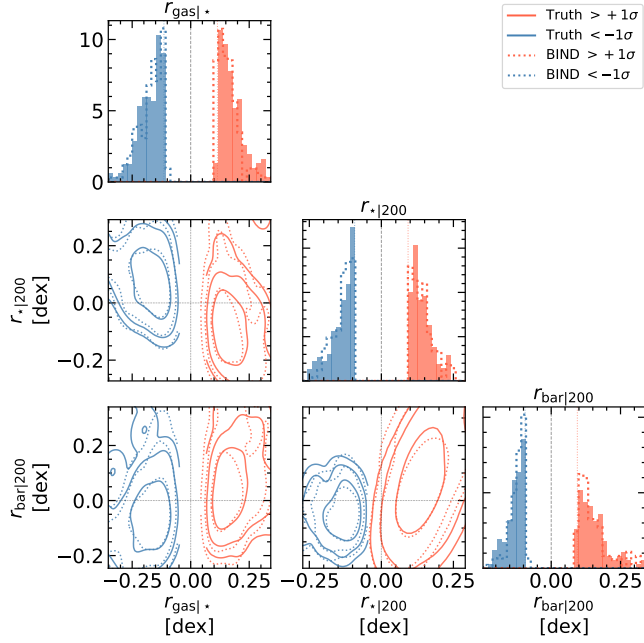


Figure 14. Corner plot of the residuals $\Delta \log M$ about the best-fit scaling relations of Fig. 13 for the CV halo population. Each off-diagonal panel shows the joint distribution of residuals for a pair of relations. The diagonal panels show the marginal distributions. Orange (blue) contours correspond to halos that scatter more than 1σ above (below) the mean scaling relation in any given panel. Filled contours show the true hydrodynamical halos, and open contours show the BIND-generated halos, smoothed with a Gaussian KDE. Contour levels correspond to the 1σ , 2σ , and 3σ enclosed probability. The agreement between the true and generated contours across all panels demonstrates that BIND has learned the full multivariate structure of the baryonic mass components.

generated halo patch back into the global map at the halo center, replacing the corresponding DMO pixels. Each generated field blends into the surrounding DMO field through a raised-cosine taper that begins at $3R_{200c}$ and fades the replacement to zero at the patch edge. In regions where multiple halo patches overlap, we apply a cosine taper function to blend the edges of adjacent patches and assign priority to the more massive halo so that the largest halos take precedence in contested regions. This strategy follows M. E. Lee et al. (2026b), adapted here for BINDed fields rather than the true hydrodynamical patches used in that work. The total baryonic map is obtained by summing the gas, stellar, and modified dark-matter channels of the pasted field.

4. **Power spectrum computation.** We compute the 2D projected matter power spectrum of the full

baryonified map using PYLIANS (F. Villaescusa-Navarro 2018) and compare it to the power spectra of (i) the DMO projection, (ii) the true hydrodynamical map, and (iii) a “hydro-replace” control map constructed by pasting the actual hydrodynamical halo patches in place of the BIND outputs using the same cosine-taper procedure. The hydro-replace map isolates the pasting algorithm from any inaccuracy in the BIND generations.

Fig. 15 shows the resulting maps from this procedure for each of the fields. Each bubble is centered on a halo extracted from the DMO simulation and generated with BIND. We then apply the pasting mechanism cumulatively, but here we show each channel individually to make clear what they look like after being placed back onto the map.

Fig. 16 shows the resulting power spectrum suppression $S(k)$ for each of these maps, along with the ratio of each to the true hydrodynamical power spectrum, for the CV, SB35, and 1P test sets (left, middle, and right columns of Fig. 16). We stress up front that, because only halos above $10^{13} M_{\odot} h^{-1}$ are replaced, neither BIND nor the hydro-replace control is expected to recover the full suppression of the hydrodynamical box; the hydro-replace control sets the ceiling attainable with this halo sample and aperture, and it, rather than the true hydrodynamical curve, is the benchmark against which BIND should be judged. Several conclusions are immediately apparent.

First, the pasted maps recover only part of the suppression measured in the full hydrodynamical simulations, a consequence of the halo sample and paste aperture rather than of the BIND emulator. The hydro-replace control, which is derived from *true* hydrodynamical patches pasted through the identical machinery, itself recovers only $\sim 65\%$ (CV) of the true suppression $1 - S(k)$ at $k = 10\text{--}20 h \text{ Mpc}^{-1}$ at the $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ threshold, consistent with previous findings that the full suppression requires halos down to $M_{200c} \sim 10^{12} M_{\odot}$ and the baryonic field out to several R_{200c} (M. P. van Daalen et al. 2020; K. Miller et al. 2026; M. E. Lee et al. 2026b). Extending BIND to lower halo masses is a straightforward avenue for future work.

Second, BIND approaches but does not fully saturate the ceiling set by halo replacement. Plotting the ratio $P_{\text{BIND}}/P_{\text{hydro-replace}}$ directly (bottom row of Fig. 16) shows BIND carrying 2.5% more small-scale power than the control at $k = 10\text{--}20 h \text{ Mpc}^{-1}$ in the CV suite (1.025, with a 16–84 simulation range of [0.992, 1.067]) and 6.0% more in 1P (1.060 [1.030, 1.084]), while the SB35 test set is consistent with the control there (0.986 [0.927, 1.034]) and falls $\sim 5\%$ below it at $k = 40\text{--}64 h \text{ Mpc}^{-1}$. Because

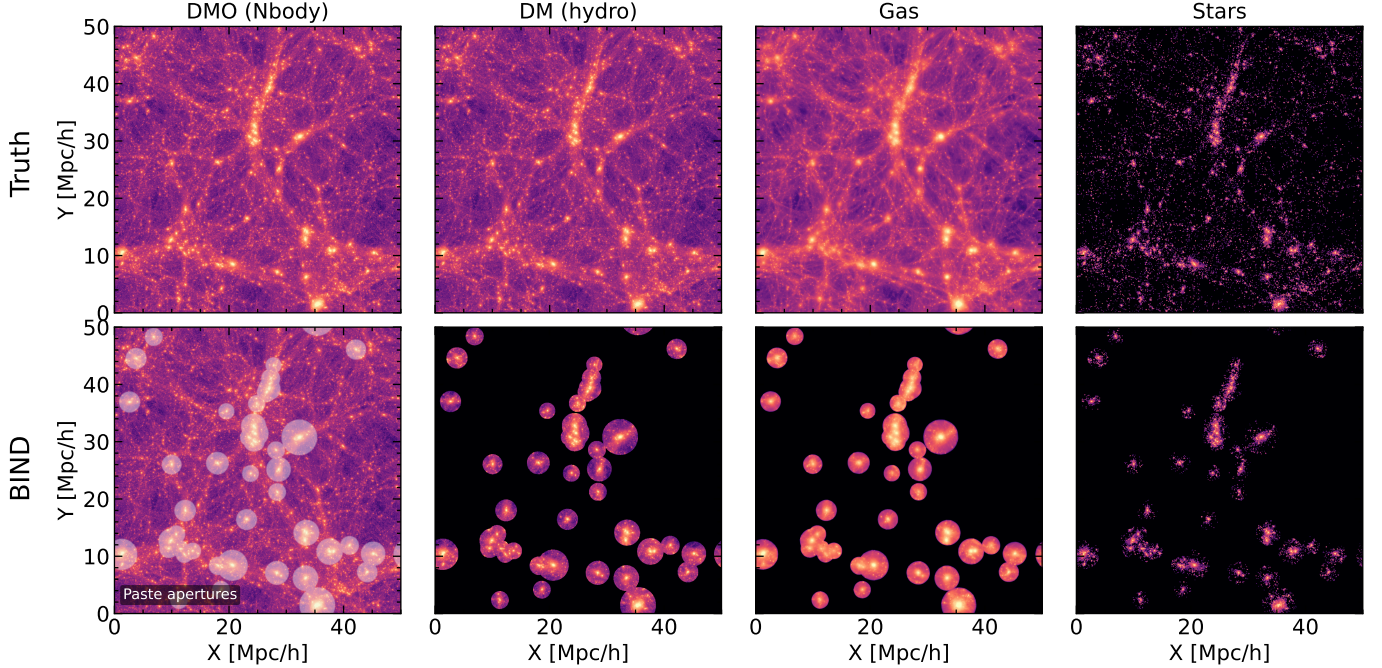


Figure 15. BIND applied to a full N -body volume. *Top row:* The truth, showing the DMO projection of the N -body run followed by the dark matter, gas, and stellar surface densities of its paired hydrodynamical counterpart, each projected over the full $50 h^{-1}$ Mpc box onto a 1024^2 grid. *Bottom row:* the same DMO field, and the three channels generated by BIND from it and pasted back in, halo by halo, for every halo with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$. We show the channels individually rather than cumulatively to make the procedure’s scope explicit. BIND generates the baryonic content of the resolved halos and their immediate surroundings, and the field between them is inherited from the N -body run. Each bubble reproduces the morphology of its hydrodynamical counterpart, with the stellar component far more compact than the gas, and the largest halos tracing the same knots of the cosmic web that the truth does.

$P_{\text{BIND}}/P_{\text{truth}} > 1$ at these scales, the residual error is an *under-suppression*, meaning that the generated halos remove slightly too little power relative to the truth, with the median $P_{\text{BIND}}/P_{\text{truth}}$ peaking at +12% (CV) and +15% (1P) near $k \approx 16\text{--}18 h \text{ Mpc}^{-1}$, and its size tracks the set-dependent stellar and gas mass residuals of Section 4.2. Across simulations, the excess over the control correlates with the per-simulation stellar and gas mass residuals.

Finally, for the SB35 test set (which spans the full 35-dimensional parameter space with parameter combinations and initial conditions entirely withheld from training), the median suppression is reproduced to better than $\sim 5\%$ at all $k < 40 h \text{ Mpc}^{-1}$ (the median $P_{\text{BIND}}/P_{\text{truth}}$ peaks at 1.05 near $k \simeq 13 h \text{ Mpc}^{-1}$), consistent with the parameter-sensitivity results in Section 5.

These results are encouraging. For a model to saturate the information ceiling exactly, it would have to reproduce the fields in and around halos perfectly. This is, of course, a tall order for any model, especially one expected to maintain fidelity across the full high-dimensional parameter space. The advantage of this approach is that, beyond achieving near-percent-level

accuracy across scales relative to this ceiling, it simultaneously captures the profiles, integrated masses, and aspherical redistribution of mass. Future work will first improve the mass threshold of BIND to allow lower masses; this will allow the ceiling to approach that of the true hydrodynamical simulation. Future work will also improve the fidelity of the flow model.

7.2. Implications for the Future

The results in Section 7.1 establish that BIND can serve as a practical, fast baryonification engine for large-scale structure analyses. Generating baryonic fields for all halos in a single $50 h^{-1}$ Mpc simulation takes $\mathcal{O}(\text{minutes})$ on a single GPU (see Table 2), orders of magnitude faster than re-running the hydrodynamical simulation, which makes it feasible to baryonify thousands of N -body realizations in the time required to complete a single hydrodynamical run.

The most immediate application is learning the relationship between baryonic effects in massive halos and weak lensing statistics. This application is already explored in the companion paper (M. E. Lee et al. 2026a). However, that work uses a base DMO simulation from

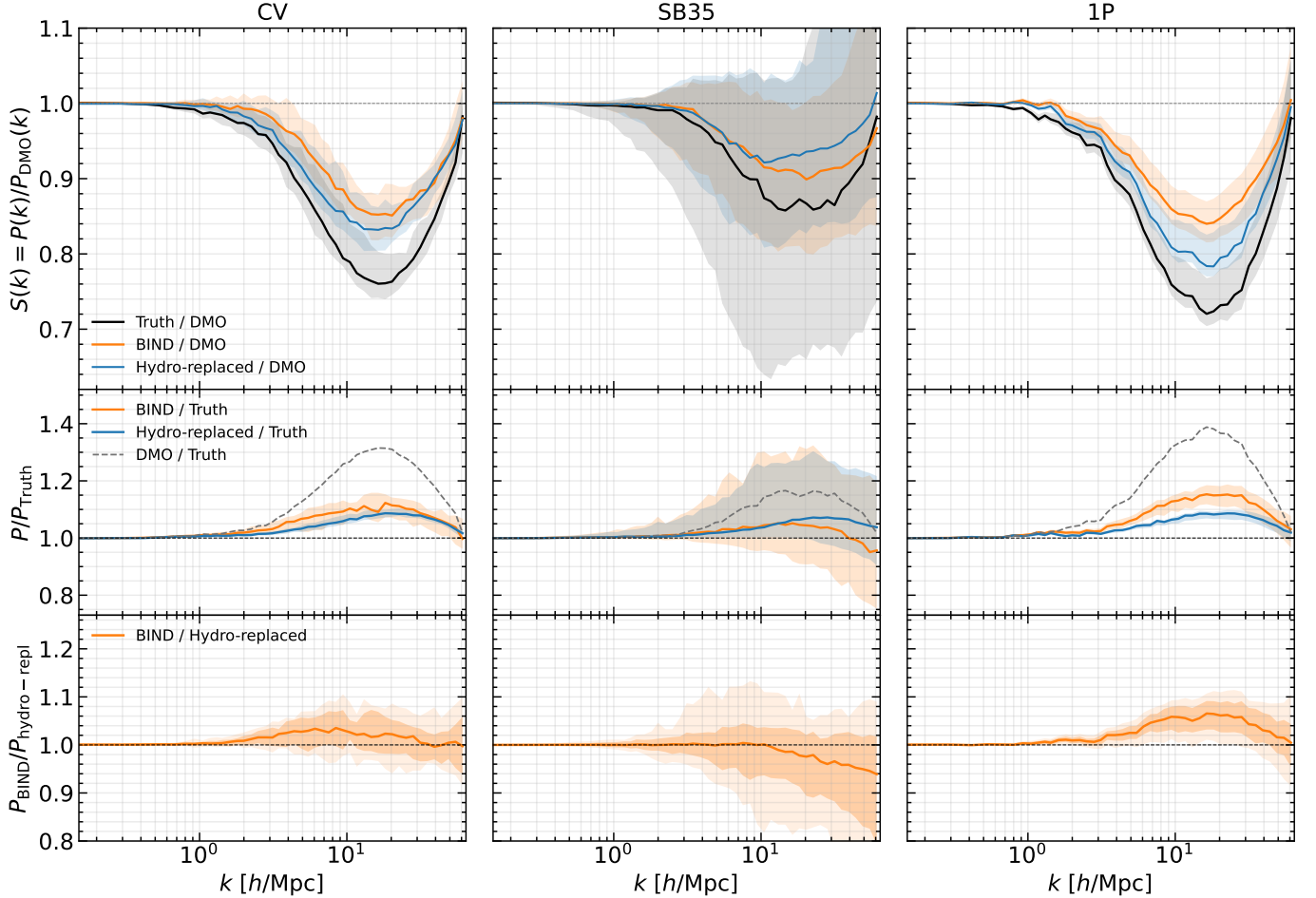


Figure 16. Projected matter power spectrum suppression for the CV, SB35 test, and 1P suites. All curves are medians over simulations of per-simulation ratios, truncated at the Nyquist scale $k = 64 h \text{ Mpc}^{-1}$. *Top row:* $S(k) = P(k)/P_{\text{DMO}}(k)$ for the true hydrodynamical maps (black), the BIND-baryonified maps (orange), and the hydro-replace control where true hydrodynamical patches are pasted through the identical aperture and blending as BIND (blue), each with its 16–84 percentile band over simulations. *Middle row:* the ratio of each to the true hydrodynamical spectrum, with 16–84 percentile bands over simulations and the DMO-to-truth ratio (gray dashed) for reference. *Bottom row:* the ratio $P_{\text{BIND}}/P_{\text{hydro-repl}}$ with 16–84 (dark) and 5–95 (light) bands. Neither BIND nor the hydro-replace control reaches the full hydrodynamical suppression (black), because only halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ are replaced, and only within a finite aperture; the hydro-replace control therefore sets the ceiling attainable with this halo sample, and the relevant comparison for BIND is orange versus blue rather than orange versus black.

IllustrisTNG300-Dark, run at a single point in the cosmological parameter space.

By running a suite of N -body simulations at different cosmological parameters and baryonifying each with BIND at a range of θ_{TNG} values, we can construct a joint likelihood over cosmological and galaxy-formation parameters directly from field-level observables such as weak-lensing convergence maps, without relying on analytic profile prescriptions.

Alternatively, BIND can be used in a forward-modeling framework to constrain the IllustrisTNG subgrid parameters from observations. Multi-wavelength data such as X-ray gas fractions, the kinematic and thermal Sunyaev-Zel’dovich effect, and optical weak lens-

ing jointly constrain different combinations of the feedback parameters (M. Kovač et al. 2025), and fast radio burst dispersion measures are emerging as a further, largely independent handle on how far feedback pushes gas (I. Medlock et al. 2024a, 2025). Emulators built on CAMELS have already been used to turn such observables into constraints on the subgrid parameters directly (E. T. Lau et al. 2024). Because BIND is conditioned on the full 35-dimensional parameter vector, it can, in principle, propagate observational constraints on these parameters into predictions for any field-level observable, which would provide a unified framework for multi-probe baryon modeling. This would require up-

dates to the BIND methodology for the output fields and, ideally, the incorporation of 3D field generation.

A third direction is generating large training sets for summary-statistic emulators or simulation-based inference networks. Field-level neural network summaries trained on cosmological maps require thousands of simulation realizations across diverse parameter combinations, a task that BIND can perform at negligible cost once the N -body simulations are available. Ideally, we can use BINDed simulations to generate accurate beyond-Gaussian statistics.

Finally, a joint training over multiple subgrid prescriptions is an enticing next step. When the CAMELS suite expands to include $50 h^{-1} \text{ Mpc}$ boxes across a range of subgrid prescriptions, BIND can inherit all the different feedback models and, hopefully, become a unified halo-based generative model.

8. CAVEATS WITH THE BIND APPROACH

Despite its encouraging performance, the BIND framework presented here has several limitations to keep in mind when applying it in practice.

The current model is trained and validated on halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$, which contribute the largest individual patches to the baryonified field but represent only a fraction of the halos responsible for the full power spectrum suppression (M. P. van Daalen et al. 2011; A. Schneider & R. Teyssier 2015; A. Schneider et al. 2019; G. Aricò et al. 2020; M. E. Lee et al. 2026b). Extending the training to group- and galaxy-scale halos (10^{12} – $10^{13} M_{\odot}$) would require handling a larger number of objects per simulation volume and a wider dynamic range of halo morphologies, but is a natural and necessary next step for achieving percent-level accuracy in $S(k)$ at all relevant scales.

BIND operates on projected 2D mass maps rather than 3D particle distributions, which introduces line-of-sight confusion from foreground and background matter unassociated with the target halo. While baryonification carried out directly on projected maps has been shown to reproduce two-point and several higher-order weak lensing statistics at the few-percent level (D. Anbajagane et al. 2024; A. J. Zhou et al. 2025), the approximation may be less accurate for higher-order statistics that are sensitive to the three-dimensional geometry of halos and filaments or quantities that care about the full three-dimensional density instead of projected surface density. Extending BIND to 3D inputs and outputs is conceptually straightforward but requires substantially more GPU memory and training data.

BIND is trained exclusively on the IllustrisTNG galaxy-formation model as implemented in the

CAMELS SB35 suite. While this covers a broad, 35-dimensional subgrid parameter space, it is anchored to a single simulation code and a single set of physical prescriptions. Applying BIND to parameter combinations outside the SB35 training range, or to simulation codes with qualitatively different feedback implementations (e.g., SIMBA or EAGLE), would require either retraining on the appropriate simulation suite or demonstrating that the learned mapping generalizes across codes.

Similarly, BIND is trained only on a single map resolution. This limits its applicability to dark matter maps generated at the same resolution as those BIND was trained on. Increasing the training set to include multi-fidelity simulations will be an important next step to generalize BIND. It would also be interesting to see whether lower-resolution (and cheaper) simulations, gridded to the same map resolution as now, would work well for BIND.

9. SUMMARY AND CONCLUSIONS

We have introduced BIND (Baryonic INpainting with Deep learning), a conditional flow matching model that learns the field-level mapping from dark-matter-only halos to their hydrodynamical counterparts, conditioned on the full 35-dimensional cosmological and IllustrisTNG subgrid parameter space of the CAMELS SB35 suite. BIND takes as input a projected N -body patch centered on a halo, together with a parameter vector θ_{TNG} , and outputs simultaneous, physically consistent realizations of the dark-matter, gas, and stellar mass fields without imposing any analytic functional form on any component. Applied patch-by-patch to the halos in an N -body simulation, BIND produces baryonified cosmological fields at $\mathcal{O}(\text{minute})$ cost per simulation volume on a single GPU, orders of magnitude faster than a hydrodynamical re-simulation.

We validated BIND through a comprehensive suite of tests on three independent evaluation sets (the SB35 test set, the 1P single-parameter-variation set, and the CV cosmic-variance set), spanning a wide range of halo masses, cosmological parameters, and galaxy-formation physics. Our main findings are the following.

1. **Accurate integrated masses and parameter sensitivity.** BIND recovers nearly unbiased total masses for all three baryonic components in hydrodynamical simulations across the full evaluation parameter space, with median relative errors below 1% for dark matter and total mass, $\sim 1\%$ for gas, and $\sim 10\%$ for stars. Spearman rank correlations between the generated integrated masses and the IllustrisTNG subgrid parameters closely match the ground-truth correlations across

all channels, showing that the model has learned the dependence of baryonic content on galaxy-formation physics.

2. **Radially resolved profile accuracy.** Azimuthally averaged surface density profiles are reproduced to better than 10% accuracy at all projected radii out to $3.125 h^{-1} \text{Mpc}$ across all test sets. The radially resolved Spearman correlations with the IllustrisTNG parameters match the ground truth at virtually all radii, capturing the physically motivated competition between gas cooling and feedback-driven expulsion as a function of projected radius.
3. **Projected halo morphology.** Distributions of the projected axis ratio q for all three mass channels are reproduced faithfully across all test sets, correctly recovering the ordering of shapes among components. Dark matter peaks near $q \approx 0.87$, the gas is more circular, and stars exhibit the broadest distribution and largest ellipticities. This demonstrates that BIND has learned the anisotropic morphological mapping from DMO to baryonic fields, a non-trivial capability with direct implications for weak-lensing intrinsic alignment modeling, including the direction of the ellipse for stellar fields.
4. **Field-level parameter response.** Using the 1P suite, we showed that BIND correctly reproduces the signed, spatially resolved response of the gas and stellar fields to variations in individual subgrid parameters, including the central gas deficit and peripheral excess driven by AGN feedback (AAGN1, AAGN2) and the enhanced infall gas density and stellar suppression driven by stronger stellar winds (ASN1). The dark-matter channel correctly shows little coherent parameter response, consistent with its gravitational origin.
5. **Inter-component correlations and scaling relations.** Across the 1,111 CV halos at the fiducial parameters, BIND reproduces the slope, normalization, and scatter of all three inter-component scaling relations ($M_{\text{gas}}-M_{\text{star}}$, SHMR, $M_{\text{bar}}-M_{200c}$). Corner plots of the residuals confirm that the joint multivariate distribution of halos that deviate coherently from multiple relations simultaneously is also recovered, including physically motivated correlations such as the coherent SHMR-baryon excess and the gas-stellar anticorrelation driven by AGN feedback. Halo mass

was never used as an explicit conditioning variable, yet the model emergently learned the correct halo-mass dependence of the baryon fraction across more than a decade in M_{200c} .

6. **Power spectrum and baryonification.** Baryonifying full $50 h^{-1} \text{Mpc}$ N -body projections with BIND reproduces the projected matter power spectrum suppression $S(k) = P_{\text{hydro}}(k)/P_{\text{DMO}}(k)$ to within a few percent across the full range of k probed, for both the CV and SB35 test sets. The BIND suppression curve closely tracks the hydro-replace control, indicating that the model approaches the halo-pasting procedure's performance ceiling.

Together, these results answer the four questions posed in the introduction. BIND can learn the field-level DMO-to-hydro mapping at the halo level, correctly encode the dependence of baryonic structure on the IllustrisTNG galaxy-formation parameter space, capture the inter-component correlations and their parameter dependence, and can be deployed as a practical baryonification engine with percent-level accuracy for two-point statistics.

The present work has several limitations that define a natural roadmap for future development. The most important is the halo mass threshold of $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$: extending BIND to group- and galaxy-scale halos ($10^{12}-10^{13} M_{\odot}$) would recover the full power spectrum suppression and open the model to a wider range of observational comparisons. BIND currently operates in 2D projection, a choice that reduced computational costs and simplified the halo-pasting step, but extending to 3D fields is a natural next iteration. The model is also tied to the IllustrisTNG subgrid framework, and extending it to other galaxy-formation models such as SIMBA or EAGLE would require retraining on the appropriate CAMELS suite, though the architecture and training procedure are directly transferable.

Looking forward, BIND is well positioned to address some of the most pressing systematic challenges for upcoming weak-lensing surveys. The model's halo-centric, field-level, parameter-conditioned, and generative design allows it to serve as a forward-modeling engine for joint constraints on cosmological and galaxy-formation parameters using non-Gaussian statistics such as peak counts, Minkowski functionals, and field-level neural summaries, for which analytic baryonic correction tools have been shown to be insufficient (M. E. Lee et al. 2023, 2026b). When combined with a fast N -body simulator and a flexible posterior sampler, BIND could enable full field-level inference pipelines that marginalize over the

IllustrisTNG parameter space at speeds that were previously infeasible, unlocking the full statistical power of surveys such as *Euclid* (R. Laureijs et al. 2011), the Vera C. Rubin Observatory LSST (Ž. Ivezić et al. 2019), and the Nancy Grace Roman Space Telescope (D. Spergel et al. 2015). The companion paper (M. E. Lee et al. 2026a) takes the first step.

ACKNOWLEDGMENTS

We thank Colin Hill, Carolina Cuesta-Lazaro, Daisuke Nagai, Erwin Lau, and Amanda Lue for useful discussions during this project. MEL is supported by NSF grant DGE-2036197. ZH acknowledges financial support from NASA ATP grant 80NSSC24K1093. The Flatiron Institute is supported by the Simons Foundation. GLB acknowledges support from the NSF (AST-2307419) and NASA (80NSSC21K1053), as well as support from the

Simons Foundation through the Learning the Universe Collaboration. C. Lovell was supported by the research environment and infrastructure of the Handley Lab at the University of Cambridge. MEL also thanks Erin Walter for helpful comments and editing. The authors used Claude Opus and sonnet to refine sections of code and text, and Grammarly was used to refine portions of the draft with grammatical edits. The authors take full responsibility for the final content.

AUTHOR CONTRIBUTIONS

MEL performed the analysis, code generation, testing, and text writing under the mentorship and support of SG, ZH, GLB, CL and BH.

Software: astropy (Astropy Collaboration et al. 2013, 2018, 2022), Pylians (F. Villaescusa-Navarro 2018), PyTorch (A. Paszke et al. 2019)

APPENDIX

A. ARCHITECTURE

A.1. Conditioning Embeddings

The flow time t is encoded with sinusoidal positional embeddings (P. Dhariwal & A. Nichol 2021) of dimension 128, then projected to a 512-dimensional embedding $\mathbf{e}_t$ via a two-layer MLP with SiLU activations. The scale factor a is encoded in the same way (128-dimensional sinusoidal features followed by a two-layer MLP with SiLU activations) to yield $\mathbf{e}_a$. The parameter vector $\boldsymbol{\theta}_{\text{TNG}}$ is independently processed by a third two-layer MLP ($35 \rightarrow 512 \rightarrow 512$, SiLU activations) to yield $\mathbf{e}_\theta$. The three embeddings are summed to form a single conditioning vector

$$\mathbf{e} = \mathbf{e}_t + \mathbf{e}_a + \mathbf{e}_\theta, \quad (\text{A1})$$

which is broadcast into every residual block of the U-Net.

A.2. Encoder–Decoder with Adaptive Group Normalization

The encoder consists of four resolution stages with channel multipliers (1, 2, 4, 8) relative to a base width of 128, giving feature-map widths of 128, 256, 512, and 1024. Each stage contains two residual blocks, and multi-head self-attention layers (4 heads; A. Vaswani et al. 2017) are inserted at resolutions 32^2 and 16^2 . Spatial downsampling uses strided convolutions; upsampling uses nearest-neighbor interpolation followed by convolution. Encoder activations are passed to the decoder via skip connections by channel-wise concatenation. Each residual block conditions on $\mathbf{e}$ through Adaptive Group Normalization (P. Dhariwal & A. Nichol 2021),

$$\text{AdaGN}(\mathbf{h}, \mathbf{e}) = (1 + \boldsymbol{\gamma}(\mathbf{e})) \odot \text{GN}(\mathbf{h}) + \boldsymbol{\beta}(\mathbf{e}), \quad (\text{A2})$$

where GN denotes Group Normalization with 32 groups (Y. Wu & K. He 2018), and $\boldsymbol{\gamma}, \boldsymbol{\beta}$ are learned linear projections of $\mathbf{e}$. This allows the network to modulate its internal representations as a joint function of flow time, scale factor, and model parameters. The output projection layer is zero-initialized to stabilize early training.

The full architecture is summarized in Table 5.

Table 5. U-Net architecture hyperparameters.

Parameter	Value
Base channels	128
Channel multipliers	(1, 2, 4, 8)
Residual blocks per stage	2
Attention heads	4
Attention resolutions	32^2 , 16^2
Embedding dimension	512
Dropout	0.1
Output initialization	Zero

Table 6. Training hyperparameters.

Hyperparameter	Value
Learning rate η_0	10^{-4}
Weight decay	10^{-4}
Gradient clip norm	1.0
Warm-up steps	1,000
Max epochs	200
Batch size (per GPU)	64
EMA decay ρ	0.9999
CFG dropout probability	0.1
Euler steps at inference	20

B. OPTIMIZATION DETAILS

The learning rate follows a two-phase schedule: a linear warm-up from zero to $\eta_0 = 10^{-4}$ over the first $N_{\text{warm}} = 1,000$ gradient steps, followed by cosine annealing to zero over the remainder of training,

$$\eta(k) = \begin{cases} \eta_0 \frac{k}{N_{\text{warm}}}, & k < N_{\text{warm}}, \\ \frac{\eta_0}{2} \left[1 + \cos \left(\pi \frac{k - N_{\text{warm}}}{K_{\text{tot}} - N_{\text{warm}}} \right) \right], & k \geq N_{\text{warm}}, \end{cases} \quad (\text{B3})$$

where k is the current gradient step and K_{tot} is the total number of steps (I. Loshchilov & F. Hutter 2017). Weight decay is set to $\lambda = 10^{-4}$, and gradients are clipped to a maximum ℓ_2 norm of 1.0.

An exponential moving average (EMA) of the network weights is maintained throughout training with decay $\rho = 0.9999$,

$$\bar{\phi}_k = \rho \bar{\phi}_{k-1} + (1 - \rho) \phi_k, \quad (\text{B4})$$

and the EMA weights $\bar{\phi}$ are used exclusively at inference (T. Karras et al. 2022). During training, the conditioning vector is dropped with probability 0.1, following the classifier-free guidance construction of J. Ho & T. Salimans (2022), so that the network also learns an unconditional velocity field. We do not apply guidance at inference in this work, and all results shown use the conditional model. Table 6 collects the full set of training hyperparameters.

C. CHARACTERIZING THE LINE-OF-SIGHT PROJECTION

The fiducial training and evaluation maps integrate the density fields over the full $50 h^{-1} \text{Mpc}$ line of sight (Section 3.1), so every halo aperture contains some foreground and background material unassociated with the halo. Here we quantify the projection contribution directly. For every evaluation halo we construct a halo-local reference by re-projecting only the $(6.25 h^{-1} \text{Mpc})^3$ cube centered on the halo, and match halos one-to-one between the two projections (identical friends-of-friends masses and centers), giving 8,273 matched halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ (1,154 CV, 6,823 1P, 296 SB35-test). The line-of-sight excess within R_{200c} , $X = M^{50}/M^{6.25} - 1$, has a pooled median (16th–84th percentile range) of +6.1% [+3.1, +18.5] for dark matter, +12.0% [+4.6, +30.8] for gas, +0.4% [−0.0, +8.1] for stars,

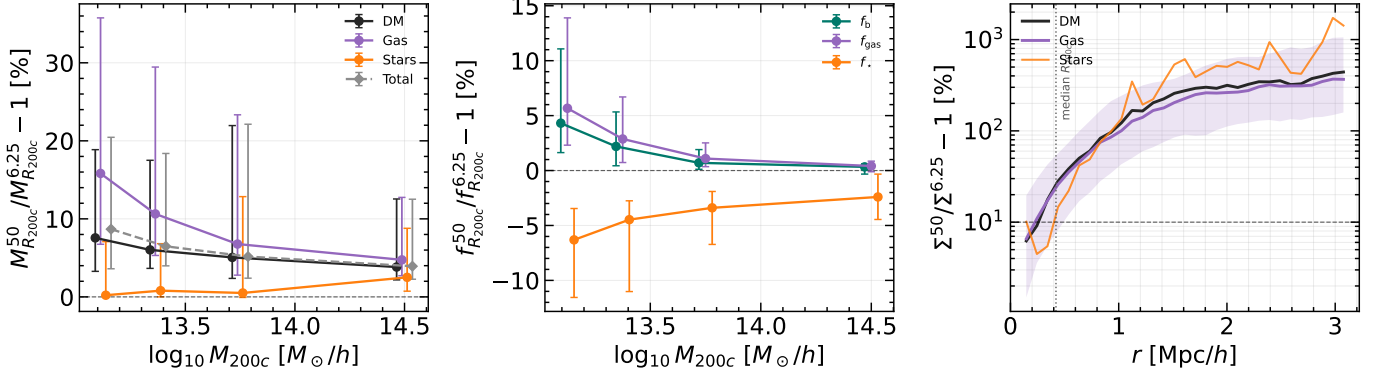


Figure 17. Projection-depth characterization on 8,273 matched halos: the full $50 h^{-1}$ Mpc-depth truth measured against the halo-local $6.25 h^{-1}$ Mpc cube truth of the same halos. *Left:* median R_{200c} -aperture line-of-sight excess $M_{R_{200c}}^{50} / M_{R_{200c}}^{6.25} - 1$ per component as a function of halo mass; bars span the 16th–84th percentiles. *Center:* the induced bias of the projected aperture fractions, $f_{R_{200c}}^{50} / f_{R_{200c}}^{6.25} - 1$. *Right:* median surface-density excess $\Sigma^{50} / \Sigma^{6.25} - 1$ versus radius (32 linear annuli; shaded: gas 16th–84th range; the horizontal line marks 10%); the innermost annulus is omitted because it is dominated by sub-pixel registration differences between the two projections.

and $+6.6\%$ [$+3.3, +20.1$] for the total mass, essentially identical across the three suites (Fig. 17). The excess falls with halo mass as the halo increasingly dominates its own line of sight: the median gas excess drops from $+15.8\%$ at $M_{200c} = 10^{13} - 10^{13.25} M_{\odot} h^{-1}$ to $+4.7\%$ above $10^{14} M_{\odot} h^{-1}$ (total: $+8.7\%$ to $+3.9\%$). A small tail is heavily contaminated — $X_{\text{tot}} > 0.5$ for 0.8% of halos and $X_{\text{tot}} > 1$ for 0.25% — and it is these systems that appear as outliers above the stellar-to-halo and baryonic scaling relations of Section 6.2. The stellar apertures are essentially projection-free, so projection biases the *composition* of the aperture: relative to the halo-local values, the projected baryon fraction is high by a median $+2.4\%$ [$+0.4, +7.9$] (rising to $+4.3\%$ in the lowest mass bin), the gas fraction by $+3.2\%$, and the stellar fraction is diluted by -5.2% [$-10.3, -2.7$]. Radially, the contamination lives in the aperture outskirts: the median surface-density excess $\Sigma^{50} / \Sigma^{6.25} - 1$ is below 10% inside $0.25 h^{-1}$ Mpc, crosses $+50\%$ near $0.7 h^{-1}$ Mpc, and reaches 300–400% at the patch edge, while the aperture-integrated excess stays modest because the aperture mass is core-dominated.

Because the true and generated maps share the identical projection (the same rotation matrices and line of sight, Section 3.1), every truth-versus-BIND comparison in this paper is internally consistent. The line-of-sight term appears on both sides and cancels out in the relative statements. The absolute aperture quantities, however, include it. Aperture masses overestimate the halo-associated mass by the X values above, the baryon fractions of Section 6.1 carry the $+2$ – 4% relative offset, and the projected shapes of Section 4.4 are diluted, as already noted there. Comparisons with three-dimensional, halo-intrinsic quantities from other simulations or from observations should therefore either apply these corrections or forward-model the same projection depth. Finally, the projection depth is not forced on the model by the learning problem. An otherwise identical flow-matching model trained on the $6.25 h^{-1}$ Mpc cube projections (without the large-scale context channels) performs nearly as well against its own halo-local truth, with median total-mass residuals within R_{200c} of $+1.9\%$ versus $+0.5\%$ for the fiducial model (CV suite; gas $+8.8\%$ versus $+3.4\%$), so the full-depth choice is driven by the conditioning and pasting workflow of Section 7 rather than by model accuracy alone.

REFERENCES

- Adam, A., Coogan, A., Malkin, N., et al. 2022, arXiv e-prints, arXiv:2211.03812, doi: [10.48550/arXiv.2211.03812](https://doi.org/10.48550/arXiv.2211.03812)
- Albergo, M. S., & Vanden-Eijnden, E. 2023, in The Eleventh International Conference on Learning Representations (ICLR), doi: [10.48550/arXiv.2209.15571](https://doi.org/10.48550/arXiv.2209.15571)
- Anbajagane, D., Pandey, S., & Chang, C. 2024, The Open Journal of Astrophysics, 7, 108, doi: [10.33232/001c.126788](https://doi.org/10.33232/001c.126788)
- Aricò, G., Angulo, R. E., Contreras, S., et al. 2021a, MNRAS, 506, 4070, doi: [10.1093/mnras/stab1911](https://doi.org/10.1093/mnras/stab1911)
- Aricò, G., Angulo, R. E., Contreras, S., et al. 2021b, MNRAS, 503, 3596, doi: [10.1093/mnras/stab699](https://doi.org/10.1093/mnras/stab699)

- Aricò, G., Angulo, R. E., Hernández-Monteagudo, C., et al. 2020, *MNRAS*, 495, 4800, doi: [10.1093/mnras/staa1478](https://doi.org/10.1093/mnras/staa1478)
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, 935, 167, doi: [10.3847/1538-4357/ac7c74](https://doi.org/10.3847/1538-4357/ac7c74)
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- Bigwood, L., Amon, A., Schneider, A., et al. 2024, *MNRAS*, 534, 655, doi: [10.1093/mnras/stae2100](https://doi.org/10.1093/mnras/stae2100)
- Chadayammuri, U., Ntampaka, M., ZuHone, J., Bogdán, Á., & Kraft, R. P. 2023, *MNRAS*, 526, 2812, doi: [10.1093/mnras/stad2596](https://doi.org/10.1093/mnras/stad2596)
- Chisari, N. E., Mead, A. J., Joudaki, S., et al. 2019, *The Open Journal of Astrophysics*, 2, 4, doi: [10.21105/astro.1905.06082](https://doi.org/10.21105/astro.1905.06082)
- Cuesta-Lazaro, C., & Mishra-Sharma, S. 2024, *PhRvD*, 109, 123531, doi: [10.1103/PhysRevD.109.123531](https://doi.org/10.1103/PhysRevD.109.123531)
- Dalal, R., Li, X., Nicola, A., et al. 2023, *PhRvD*, 108, 123519, doi: [10.1103/PhysRevD.108.123519](https://doi.org/10.1103/PhysRevD.108.123519)
- Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, *MNRAS*, 486, 2827, doi: [10.1093/mnras/stz937](https://doi.org/10.1093/mnras/stz937)
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371, doi: [10.1086/163168](https://doi.org/10.1086/163168)
- Dhariwal, P., & Nichol, A. 2021, in *Advances in Neural Information Processing Systems 34 (NeurIPS)*, doi: [10.48550/arXiv.2105.05233](https://doi.org/10.48550/arXiv.2105.05233)
- Fang, W., & Haiman, Z. 2007, *PhRvD*, 75, 043010, doi: [10.1103/PhysRevD.75.043010](https://doi.org/10.1103/PhysRevD.75.043010)
- Gebhardt, M., Anglés-Alcázar, D., Genel, S., et al. 2026, arXiv e-prints, arXiv:2601.06258, doi: [10.48550/arXiv.2601.06258](https://doi.org/10.48550/arXiv.2601.06258)
- Genel, S., Bryan, G. L., Springel, V., et al. 2019, *ApJ*, 871, 21, doi: [10.3847/1538-4357/aaf4bb](https://doi.org/10.3847/1538-4357/aaf4bb)
- Genel, S., Jo, Y., Oh, B. K., et al. 2026, arXiv e-prints, arXiv:2606.10038, doi: [10.48550/arXiv.2606.10038](https://doi.org/10.48550/arXiv.2606.10038)
- Hadzhiyska, B., Ferraro, S., Farren, G. S., Sailer, N., & Zhou, R. 2025, arXiv e-prints, arXiv:2507.14136, doi: [10.48550/arXiv.2507.14136](https://doi.org/10.48550/arXiv.2507.14136)
- Hadzhiyska, B., Ferraro, S., Ried Guachalla, B., et al. 2024, arXiv e-prints, arXiv:2407.07152, doi: [10.48550/arXiv.2407.07152](https://doi.org/10.48550/arXiv.2407.07152)
- Ho, J., Jain, A., & Abbeel, P. 2020, in *Advances in Neural Information Processing Systems*, Vol. 33, 6840–6851, doi: [10.48550/arXiv.2006.11239](https://doi.org/10.48550/arXiv.2006.11239)
- Ho, J., & Salimans, T. 2022, arXiv e-prints, arXiv:2207.12598, doi: [10.48550/arXiv.2207.12598](https://doi.org/10.48550/arXiv.2207.12598)
- Hockney, R. W., & Eastwood, J. W. 1988, *Computer Simulation Using Particles* (Bristol: Hilger)
- Hsu, A., Ho, M., Lin, J., et al. 2024, arXiv e-prints, arXiv:2410.02857, doi: [10.48550/arXiv.2410.02857](https://doi.org/10.48550/arXiv.2410.02857)
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111, doi: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c)
- Jing, Y. P., Zhang, P., Lin, W. P., Gao, L., & Springel, V. 2006, *ApJL*, 640, L119, doi: [10.1086/503547](https://doi.org/10.1086/503547)
- Joachimi, B., Cacciato, M., Kitching, T. D., et al. 2015, *SSRv*, 193, 1, doi: [10.1007/s11214-015-0177-4](https://doi.org/10.1007/s11214-015-0177-4)
- Karras, T., Aittala, M., Aila, T., & Laine, S. 2022, in *Advances in Neural Information Processing Systems 35 (NeurIPS)*, doi: [10.48550/arXiv.2206.00364](https://doi.org/10.48550/arXiv.2206.00364)
- Kovač, M., Schneider, A., Giri, S. K., et al. 2025, *JCAP*, 2025, 046, doi: [10.1088/1475-7516/2025/11/046](https://doi.org/10.1088/1475-7516/2025/11/046)
- Lau, E. T., Nagai, D., Bogdán, Á., et al. 2024, arXiv e-prints, arXiv:2412.04559, doi: [10.48550/arXiv.2412.04559](https://doi.org/10.48550/arXiv.2412.04559)
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193, doi: [10.48550/arXiv.1110.3193](https://doi.org/10.48550/arXiv.1110.3193)
- Lee, M. E., Genel, S., Haiman, Z., & Bryan, G. L. 2026a, arXiv e-prints
- Lee, M. E., Haiman, Z., & Genel, S. 2026b, arXiv e-prints, arXiv:2603.11815, doi: [10.48550/arXiv.2603.11815](https://doi.org/10.48550/arXiv.2603.11815)
- Lee, M. E., Haiman, Z., Pandey, S., & Genel, S. 2026c, *ApJ*, 996, 36, doi: [10.3847/1538-4357/ae1ca7](https://doi.org/10.3847/1538-4357/ae1ca7)
- Lee, M. E., Lu, T., Haiman, Z., Liu, J., & Osato, K. 2023, *MNRAS*, 519, 573, doi: [10.1093/mnras/stac3592](https://doi.org/10.1093/mnras/stac3592)
- Lee, M. E., Genel, S., Wandelt, B. D., et al. 2024, *ApJ*, 968, 11, doi: [10.3847/1538-4357/ad3d4a](https://doi.org/10.3847/1538-4357/ad3d4a)
- Legin, R., Ho, M., Lemos, P., et al. 2023, *MNRAS*, 527, L173, doi: [10.1093/mnras/lsad152](https://doi.org/10.1093/mnras/lsad152)
- Leung, C., Borrow, J., Masui, K. W., et al. 2025, arXiv e-prints, arXiv:2509.19514, doi: [10.48550/arXiv.2509.19514](https://doi.org/10.48550/arXiv.2509.19514)
- Li, X., Zhang, T., Sugiyama, S., et al. 2023, *PhRvD*, 108, 123518, doi: [10.1103/PhysRevD.108.123518](https://doi.org/10.1103/PhysRevD.108.123518)
- Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. 2023, in *The Eleventh International Conference on Learning Representations (ICLR)*, doi: [10.48550/arXiv.2210.02747](https://doi.org/10.48550/arXiv.2210.02747)
- Liu, X., Gong, C., & Liu, Q. 2023, in *The Eleventh International Conference on Learning Representations (ICLR)*, doi: [10.48550/arXiv.2209.03003](https://doi.org/10.48550/arXiv.2209.03003)
- Loshchilov, I., & Hutter, F. 2017, in *The Fifth International Conference on Learning Representations (ICLR)*, doi: [10.48550/arXiv.1608.03983](https://doi.org/10.48550/arXiv.1608.03983)
- Loshchilov, I., & Hutter, F. 2019, in *The Seventh International Conference on Learning Representations (ICLR)*, doi: [10.48550/arXiv.1711.05101](https://doi.org/10.48550/arXiv.1711.05101)
- Lovell, C. C., Wilkins, S. M., Thomas, P. A., et al. 2022, *MNRAS*, 509, 5046, doi: [10.1093/mnras/stab3221](https://doi.org/10.1093/mnras/stab3221)

- Lovisari, L., Ettori, S., Gaspari, M., & Giles, P. A. 2021, *Universe*, 7, 139, doi: [10.3390/universe7050139](https://doi.org/10.3390/universe7050139)
- Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, *MNRAS*, 502, 1401, doi: [10.1093/mnras/stab082](https://doi.org/10.1093/mnras/stab082)
- Medlock, I., Nagai, D., Anglés-Alcázar, D., & Gebhardt, M. 2025, arXiv e-prints, arXiv:2501.17922, doi: [10.48550/arXiv.2501.17922](https://doi.org/10.48550/arXiv.2501.17922)
- Medlock, I., Nagai, D., Singh, P., et al. 2024a, arXiv e-prints, arXiv:2403.02313, doi: [10.48550/arXiv.2403.02313](https://doi.org/10.48550/arXiv.2403.02313)
- Medlock, I., Neufeld, C., Nagai, D., et al. 2024b, arXiv e-prints, arXiv:2410.16361, doi: [10.48550/arXiv.2410.16361](https://doi.org/10.48550/arXiv.2410.16361)
- Miller, K., More, S., & Jain, B. 2026, *JCAP*, 2026, 092, doi: [10.1088/1475-7516/2026/05/092](https://doi.org/10.1088/1475-7516/2026/05/092)
- Mishra, A., et al. 2026, arXiv e-prints, arXiv:2601.14377, doi: [10.48550/arXiv.2601.14377](https://doi.org/10.48550/arXiv.2601.14377)
- Moews, B., Davé, R., Mitchell, S., et al. 2021, *MNRAS*, 504, 4024, doi: [10.1093/mnras/stab1120](https://doi.org/10.1093/mnras/stab1120)
- Mudur, N., Cuesta-Lazaro, C., & Finkbeiner, D. P. 2024, arXiv e-prints, arXiv:2405.05255, doi: [10.48550/arXiv.2405.05255](https://doi.org/10.48550/arXiv.2405.05255)
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, *Computational Astrophysics and Cosmology*, 6, 2, doi: [10.1186/s40668-019-0028-x](https://doi.org/10.1186/s40668-019-0028-x)
- Ni, Y., Genel, S., Anglés-Alcázar, D., et al. 2023, *ApJ*, 959, 136, doi: [10.3847/1538-4357/ad022a](https://doi.org/10.3847/1538-4357/ad022a)
- Ono, V., Park, C. F., Mudur, N., et al. 2024, arXiv e-prints, arXiv:2403.10648, doi: [10.48550/arXiv.2403.10648](https://doi.org/10.48550/arXiv.2403.10648)
- Pandey, S., Lovell, C. C., Modi, C., & Wandelt, B. D. 2025, arXiv e-prints, arXiv:2511.08438, doi: [10.48550/arXiv.2511.08438](https://doi.org/10.48550/arXiv.2511.08438)
- Paszke, A., Gross, S., Massa, F., et al. 2019, in *Advances in Neural Information Processing Systems 32 (NeurIPS)*, 8024–8035, doi: [10.48550/arXiv.1912.01703](https://doi.org/10.48550/arXiv.1912.01703)
- Perez, E., Strub, F., de Vries, H., Dumoulin, V., & Courville, A. C. 2018, in *Proceedings of the AAAI Conference on Artificial Intelligence*, doi: [10.48550/arXiv.1709.07871](https://doi.org/10.48550/arXiv.1709.07871)
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018a, *MNRAS*, 475, 648, doi: [10.1093/mnras/stx3112](https://doi.org/10.1093/mnras/stx3112)
- Pillepich, A., Springel, V., Nelson, D., et al. 2018b, *MNRAS*, 473, 4077, doi: [10.1093/mnras/stx2656](https://doi.org/10.1093/mnras/stx2656)
- Remy, B., Lanusse, F., Jeffrey, N., et al. 2023, *A&A*, 672, A51, doi: [10.1051/0004-6361/202243054](https://doi.org/10.1051/0004-6361/202243054)
- Ronneberger, O., Fischer, P., & Brox, T. 2015, in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241, doi: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521, doi: [10.1093/mnras/stu2058](https://doi.org/10.1093/mnras/stu2058)
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, *MNRAS*, 526, 4978, doi: [10.1093/mnras/stad2419](https://doi.org/10.1093/mnras/stad2419)
- Schaye, J., Chaikin, E., Schaller, M., et al. 2026, *MNRAS*, 548, stag375, doi: [10.1093/mnras/stag375](https://doi.org/10.1093/mnras/stag375)
- Schneider, A., & Teyssier, R. 2015, *JCAP*, 2015, 049, doi: [10.1088/1475-7516/2015/12/049](https://doi.org/10.1088/1475-7516/2015/12/049)
- Schneider, A., Teyssier, R., Stadel, J., et al. 2019, *JCAP*, 2019, 020, doi: [10.1088/1475-7516/2019/03/020](https://doi.org/10.1088/1475-7516/2019/03/020)
- Schneider, A., Kovač, M., Bucko, J., et al. 2025, *JCAP*, 2025, 043, doi: [10.1088/1475-7516/2025/12/043](https://doi.org/10.1088/1475-7516/2025/12/043)
- Secco, L. F., Samuroff, S., Krause, E., et al. 2022, *PhRvD*, 105, 023515, doi: [10.1103/PhysRevD.105.023515](https://doi.org/10.1103/PhysRevD.105.023515)
- Sharma, D., Dai, B., Villaescusa-Navarro, F., & Seljak, U. 2025, *MNRAS*, 538, 1415, doi: [10.1093/mnras/staf294](https://doi.org/10.1093/mnras/staf294)
- Siegel, J., Amon, A., McCarthy, I. G., et al. 2025, arXiv e-prints, arXiv:2509.10455, doi: [10.48550/arXiv.2509.10455](https://doi.org/10.48550/arXiv.2509.10455)
- Sobol', I. M. 1967, *USSR Computational Mathematics and Mathematical Physics*, 7, 86, doi: [10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9)
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., et al. 2021, in *International Conference on Learning Representations (ICLR)*, doi: [10.48550/arXiv.2011.13456](https://doi.org/10.48550/arXiv.2011.13456)
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757, doi: [10.48550/arXiv.1503.03757](https://doi.org/10.48550/arXiv.1503.03757)
- Springel, V. 2010, *MNRAS*, 401, 791, doi: [10.1111/j.1365-2966.2009.15715.x](https://doi.org/10.1111/j.1365-2966.2009.15715.x)
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, 328, 726, doi: [10.1046/j.1365-8711.2001.04912.x](https://doi.org/10.1046/j.1365-8711.2001.04912.x)
- Troxel, M. A., & Ishak, M. 2015, *PhR*, 558, 1, doi: [10.1016/j.physrep.2014.11.001](https://doi.org/10.1016/j.physrep.2014.11.001)
- van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2020, *MNRAS*, 491, 2424, doi: [10.1093/mnras/stz3199](https://doi.org/10.1093/mnras/stz3199)
- van Daalen, M. P., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, *MNRAS*, 415, 3649, doi: [10.1111/j.1365-2966.2011.18981.x](https://doi.org/10.1111/j.1365-2966.2011.18981.x)
- Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in *Advances in Neural Information Processing Systems 30 (NeurIPS)*, 5998–6008, doi: [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762)
- Velliscig, M., Cacciato, M., Schaye, J., et al. 2015, *MNRAS*, 453, 721, doi: [10.1093/mnras/stv1690](https://doi.org/10.1093/mnras/stv1690)
- Villaescusa-Navarro, F. 2018, *Pylans: Python libraries for the analysis of numerical simulations*, Astrophysics Source Code Library, record ascl:1811.008
- Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al. 2021, *ApJ*, 915, 71, doi: [10.3847/1538-4357/abf7ba](https://doi.org/10.3847/1538-4357/abf7ba)
- Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., et al. 2023, *ApJS*, 265, 54, doi: [10.3847/1538-4365/acbf47](https://doi.org/10.3847/1538-4365/acbf47)

- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518, doi: [10.1093/mnras/stu1536](https://doi.org/10.1093/mnras/stu1536)
- Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, Nature Reviews Physics, 2, 42, doi: [10.1038/s42254-019-0127-2](https://doi.org/10.1038/s42254-019-0127-2)
- Wayland, A., Alonso, D., & Reischke, R. 2026, MNRAS, 547, stag557, doi: [10.1093/mnras/stag557](https://doi.org/10.1093/mnras/stag557)
- Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435, doi: [10.1146/annurev-astro-081817-051756](https://doi.org/10.1146/annurev-astro-081817-051756)
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291, doi: [10.1093/mnras/stw2944](https://doi.org/10.1093/mnras/stw2944)
- Wu, Y., & He, K. 2018, in Proceedings of the European Conference on Computer Vision (ECCV), doi: [10.48550/arXiv.1803.08494](https://doi.org/10.48550/arXiv.1803.08494)
- Zhou, A. J., Gatti, M., Anbajagane, D., et al. 2025, JCAP, 2025, 073, doi: [10.1088/1475-7516/2025/09/073](https://doi.org/10.1088/1475-7516/2025/09/073)