

BINDing the lightcone: A suite of astrophysical ray-traced weak lensing and SZ maps

MAX E. LEE ¹, SHY GENEL ², ZOLTÁN HAIMAN ^{1,3,4}, GREG L. BRYAN ^{1,2} 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

Recent multiwavelength observations of galaxy group and cluster gas suggest stronger baryonic feedback than our best-calibrated hydrodynamical simulations produce, while modeling this feedback remains a primary source of uncertainty in Stage-IV weak-lensing (WL) analyses. We present a suite of ray-traced maps generated with BIND (Baryonic INpainting with Deep learning), a conditional flow-matching model that paints baryonic mass and gas thermodynamics onto the halos of dark-matter-only simulations, and which was developed in a companion paper. Applied to IllustrisTNG300-Dark and ray-traced, we generate convergence, optical depth, and Compton- y maps at five source redshifts, each with 1000 pseudo-independent realizations. We build lightcones across a 256-node Sobol sequence spanning the thirty-dimensional IllustrisTNG galaxy formation prior, with individual parameter variations and at the fiducial model. In validation, the maps match those built from the IllustrisTNG300 halos to within LSST-Y10-like precision for a range of WL statistics. Across the prior, the response to feedback exceeds Stage-IV statistical precision by more than an order of magnitude on small scales, and different statistics respond to different model sectors: galactic winds control the WL power spectrum and gas auto- and cross-spectra, while the stellar initial mass function slope and AGN parameters shape the morphological statistics (PDF, peaks, minima, and Minkowski functionals). Finally, we find that the response of statistics can be compressed into seven halo properties which linearly predict a range of WL and SZ statistics. We publicly release the maps, statistics, and model tables.

Keywords: Galaxies (573) — Cosmology (343)

1. INTRODUCTION

Modeling baryonic feedback is a leading source of systematic uncertainty in Stage-IV weak lensing analyses (e.g. A. Amon & G. Efstathiou 2022; C. Preston et al. 2023; A. H. Wright et al. 2025). Unfortunately, the small scales that drive this uncertainty, where galactic winds and active galactic nuclei (AGN) redistribute gas, are furthermore the same scales that carry a wealth of cosmological information. Mismodeling these processes therefore has dire consequences for cosmological analyses. A simulation with AGN feedback tuned to reproduce the X-ray and optical properties of galaxy groups induces a $\sim 40\%$ bias in w_0 if that feedback is left unmodeled (E. Semboloni et al. 2011), and neglect-

ing baryons biases Ω_m and σ_8 by more than 5σ for a Stage-IV survey (A. Schneider et al. 2020). Avoiding small scales by removing them from analyses and relying only on large-scale modes inflates the Ω_m - σ_8 contour area by more than an order of magnitude (A. Schneider et al. 2020). This makes analysis pipeline requirements demanding, requiring the matter power spectrum to be predicted to roughly one percent accuracy out to $k \simeq 10 h \text{ Mpc}^{-1}$ (D. Huterer & M. Takada 2005; A. P. Hearin et al. 2012; N. E. Chisari et al. 2019), the same regime where galaxy formation model decisions dominate.

The past decade of multiwavelength observations has begun to narrow down where the real universe sits in this vast galaxy formation model space, although consensus on feedback strength remains elusive. Kinetic Sunyaev-Zel’dovich (kSZ) profiles, on their own and

jointly with *eROSITA* X-ray gas masses, suggest that group-scale halos eject gas more efficiently and to greater radii than the weakest-feedback state-of-the-art hydrodynamical simulations, such as IllustrisTNG, predict (B. Hadzhiyska et al. 2024, 2025; J. Siegel et al. 2025a). For example, the fiducial FLAMINGO run, calibrated to pre-*eROSITA* X-ray gas fractions, was shown to be $> 8\sigma$ discrepant with stacked kSZ profiles of groups, while its strongest-feedback variant (fgas- 8σ) agrees to within $\sim 2\sigma$; likewise, *eROSITA* X-ray gas fractions in groups are ~ 2 times lower than the fiducial FLAMINGO predicts (J. Siegel et al. 2025a). By stacking optically selected groups, recent literature finds that hot gas accounts for a small fraction (20–40%) of the cosmic baryon budget within R_{200} . Every flagship high-resolution, large-volume hydrodynamical simulation overpredicted this by up to a factor of three, except MAGNETICUM and SIMBA (P. Popesso et al. 2026). For weak lensing surveys, semi-analytic baryon correction models can translate these measurements into a matter power-spectrum suppression, although the translation rests on an assumed gas profile and halo model. Under those assumptions, the measurements imply a suppression of $10 \pm 2\%$ at $k = 1 h \text{ Mpc}^{-1}$, double the $\sim 5\%$ that most current lensing pipelines assume (J. Siegel et al. 2025b).

For thermal measurements of gas through the thermal Sunyaev-Zel’dovich (tSZ) effect, joint analyses of DES Year 3 weak lensing with ACT tSZ detect the shear $\times$ tSZ cross-correlation at 21σ and find $2\text{--}4\sigma$ tension with mild-feedback simulations (S. Pandey et al. 2025). The implication is that group-sized halos ($10^{13} M_{\odot} h^{-1} \leq M < 10^{14} M_{\odot} h^{-1}$) require significantly reduced thermal gas pressure (S. Pandey et al. 2022, 2025). The newest and highest signal-to-noise kSZ measurement, DESI DR2 crossed with ACT DR6, detects the effect at 18σ around luminous red galaxies (F. J. Qu et al. 2026), with companion measurements around bright and emission-line galaxies (B. Hadzhiyska et al. 2026), and favors gas ejection from group-scale halos more efficiently than the standard AGN gas profile (N. Battaglia 2016) predicts. These results are consistent with earlier weak lensing plus kSZ constraints, which preferred gas fractions systematically below simulation predictions (B. Hadzhiyska et al. 2024, 2025; L. Bigwood et al. 2024), and with the kSZ benchmark established against DESI+ACT data, in which the fiducial feedback calibrations of FLAMINGO, ANTILLES, BAHAMAS, and FABLE appear disfavored at $> 3\sigma$, while stronger-AGN variants of FLAMINGO and BAHAMAS, as well as SIMBA, reproduce the measured signal (L. Bigwood et al. 2025a; I. G. McCarthy et al. 2024). These analyses point to stronger baryonic feedback than the weakest-feedback simulations predict, al-

though how strong it is and whether every probe agrees on a single calibration remain open questions.

Beyond the weak lensing angular power spectrum analyses, Stage-IV surveys are designed to exploit non-Gaussian information through statistics such as peak counts, minimum counts, Minkowski functionals, and the one-point probability distribution function, each of which outperforms two-point statistics in forecast precision (V. Euclid Collaboration: Ajani et al. 2023; D. Zürcher et al. 2021; N. Martinet et al. 2021), and has now been demonstrated across a wide range of surveys (J. Liu et al. 2015; T. Kacprzak et al. 2016; N. Martinet et al. 2018; H. Shan et al. 2018; J. Harnois-Déraps et al. 2021; M. Gatti et al. 2022; T. Lu et al. 2022, 2023). Peak counts combined with the angular power spectrum tighten S_8 by 38% over the power spectrum alone in DES Year 3 (D. Zürcher et al. 2022), and peaks and minima combined with the power spectrum yield a 35% improvement in HSC Year 1 (T. Lu et al. 2023; G. A. Marques et al. 2024). Each of these statistics is affected by baryonic physics (K. Osato et al. 2021; W. R. Coulton et al. 2020; T. Lu et al. 2022; J. C. Broxterman et al. 2024; D. Grandón et al. 2024; M. Marinichenko et al. 2025), however, these impacts vary across the statistic space. Ignoring baryons biases peak-count inference by $\approx 4\sigma$ while leaving minimum counts nearly unaffected at $\approx 0.5\sigma$ (W. R. Coulton et al. 2020). M. E. Lee et al. (2026b) showed that the effect of baryons on morphological statistics, such as peak counts, is sourced primarily from the cores of massive halos, while the power spectrum is sensitive to the baryonic imprint from a broad range of halo masses and radii, a separation also visible in M. P. van Daalen et al. (2020); X. Yang et al. (2011); J. Liu & Z. Haiman (2016); J. C. Broxterman et al. (2024); M. Marinichenko et al. (2025). A correction calibrated to reproduce the power spectrum is therefore not guaranteed to be correct for other statistics, and in practice, it is not. M. E. Lee et al. (2023) found that baryon correction models (BCMs) tuned on the power spectrum reproduce the baryonic effects in peak counts at the percent level only for peaks with $S/N < 4$. For larger peak amplitudes, the effects of baryons are improperly modeled by BCMs, which implies they will become insufficient for use in LSST and *Euclid* WL analyses.

Baryonic modeling approaches fall into three broad categories. Full hydrodynamical simulations (e.g., FLAMINGO, BAHAMAS, IllustrisTNG, SIMBA) resolve the complex interplay and physics across scales, but are computationally prohibitive for the large volumes and wide parameter-space coverage that inference pipelines require (J. Schaye et al. 2023; F. Villaescusa-Navarro

et al. 2021). Semi-analytic baryon correction models are fast and flexible, providing $\sim 2\%$ power-spectrum accuracy (A. Schneider & R. Teyssier 2015; A. Schneider et al. 2019; G. Aricò et al. 2021; S. K. Giri & A. Schneider 2021; A. Schneider et al. 2025), and have recently been extended to operate directly on maps, where they have been calibrated against higher-order statistics rather than only against $P(k)$. For example, D. Anbajagane et al. (2024) jointly fit the second, third, and fourth moments of the lensing and tSZ fields to within measurement uncertainties above ~ 1 Mpc, and A. J. Zhou et al. (2025) fit two-point statistics, wavelet phase harmonics, scattering coefficients, and the third and fourth moments to within 2% for $\ell < 2000$ using only three parameters. However, these semi-analytic models impose spherically symmetric parametric profiles and therefore cannot represent aspherical morphology or stochasticity in gas properties for a given halo, to which non-Gaussian statistics are sensitive. Deep-learning approaches occupy a middle ground, painting baryonic fields directly onto N -body simulations (T. Tröster et al. 2019; L. Thiele et al. 2020b; B. Dai & U. Seljak 2021; U. Chadayammuri et al. 2023; D. Sharma et al. 2025; R. H. Liu et al. 2025), but existing models are trained on a single galaxy-formation calibration rather than conditioned on the subgrid parameter space, or are trained to reproduce the full volume, sacrificing resolution on small halo scales where the baryonic effects are sourced. Outside of these three categories, a model-free calibration of the suppression has recently been proposed, which cross-correlates shear with direct tracers of the ionized gas: fast radio burst dispersion measures (C. Leung et al. 2025) or the kSZ effect (B. Hadzhiyska 2026; A. Ganguly et al. 2026). While these do measure the two-point suppression without a galaxy formation model, they say nothing about non-Gaussian statistics and how they may respond. Further, none of the models above allow analysis of the connection between physics model parameters, their induced effects on halo populations, and the propagation of these effects into various weak lensing and CMB statistics.

This is the gap we aim to address in this and our companion work (M. E. Lee et al. 2026a) by introducing BIND (Baryonic INpainting with Deep learning). BIND is a conditional generative model trained via flow matching to paint baryonic fields onto the halos of dark-matter-only N -body simulations, conditioned on the full set of Λ CDM cosmological and IllustrisTNG galaxy formation parameters (M. E. Lee et al. 2026a). BIND is halo-centric, allowing for high-resolution emulations of individual halos, generative so that it can sample from the distribution of effects for a given halo, and it gen-

erates simultaneous channels so the same structures, at the same positions, appear in gas, stellar, and dark matter mass distributions as well as in the resulting gas temperatures, pressures, and entropies. The companion paper (M. E. Lee et al. 2026a) introduces the model and validates it at the halo level, while here, we apply BIND to the lightcone.

In this paper, we generate ray-traced weak lensing convergence (κ), electron scattering (τ), and Compton- y (y) maps across the IllustrisTNG galaxy formation parameter space using BIND, and use these maps to measure how the astrophysical parameters propagate into weak lensing and SZ summary statistics. We then use these measurements to build an analytic model of weak lensing statistics as a function of measurable halo quantities. Applied to lightcones built from TNG300-Dark, BIND produces (κ, y, τ) map triplets at a fraction of the cost of a full hydrodynamical simulation, allowing us to cover the parameter space instead of sampling it at a handful of points. We generate one set of map triplets by varying the thirty astrophysical parameters one at a time while keeping all others fixed to their fiducial values, and a second set by varying all thirty parameters simultaneously across the full thirty-dimensional prior volume using a 256-node Sobol sequence. At each parameter-space location, we generate maps for five source redshifts and produce 1000 pseudo-independent realizations. To our knowledge, this is the first systematic study of how a full galaxy formation parameter space propagates into a suite of Stage-IV weak lensing summary statistics and into the gas auto- and cross-spectra on a cosmological lightcone.

We demonstrate three principal results.

1. First, at the fiducial IllustrisTNG parameters, the generated BIND lightcones reproduce every WL statistic from an analogously constructed lightcone with hydrodynamical halos to within the statistical precision of an LSST-Y10-like survey.
2. Second, feedback responses vary by statistic and feedback type. The projected 2D WL statistics and every gas spectrum respond primarily to the galactic-wind sector feedback parameters, while the WL morphological statistics are more sensitive to the slope of the stellar initial mass function and AGN feedback parameters.
3. Third, the statistical response to the astrophysical parameters is compressible. A linear model can be generated from the shapes of parameter effects on statistics and a handful of measured halo properties. This implies that the imprint of feedback on the maps and statistics is carried by a

few physical, observable halo properties, and provides a bridge between the multiwavelength gas measurements discussed above and Stage-IV cosmological inference (J. Siegel et al. 2025a; L. Bigwood et al. 2025a; M. P. van Daalen et al. 2020; A. Schneider et al. 2022).

This paper is structured as follows. Section 2 reviews BIND, the IllustrisTNG-300 simulation, halo conditioning and pasting procedure, ray tracing, and introduces the various lightcone sets to be generated. Section 3 validates the fiducial generation against the hydro-pasted truth, first at the halo level (scaling relations and radial profiles) and then at the map level (all WL statistics and the gas auto- and cross-spectra). § 4 measures the response of twelve statistics across the 30-dimensional astrophysical parameter space, identifies the dominant parameters (§ 4.1), and categorizes their response into five distinct families (§ 4.2). Section 5 builds an analytic linear model from the results of § 4. We develop a physical interpretation for our results in § 6. Sections 7 and 8 catalog the limits of the maps’ construction and their remedies, and Section 9 summarizes our main conclusions. We publicly release the lightcones, measured statistics, and model tables.

2. METHODS

In this section, we review the BIND methodology, the simulation suites used, and the procedure for generating halos, lightcones, and our ray-traced maps.

2.1. BIND

BIND is a conditional generative model for creating hydrodynamical fields in and around dark-matter-only halos. More specifically, BIND is a trained *conditional flow model* (Y. Lipman et al. 2022, 2024; S. Kannan et al. 2025), which learns to sample from the posterior

$$P(\mathbf{F}_{\text{hydro}}^H | \rho_{\text{DMO}}^H, \boldsymbol{\theta}, z), \quad (1)$$

where $\mathbf{F}_{\text{hydro}}^H$ is a set of hydrodynamical fields such as gas, stellar, and dark matter densities, ρ_{DMO}^H is a dark matter density field from an N -body simulation centered on a halo H , $\boldsymbol{\theta}$ is a set of cosmological and astrophysical parameters, and z is the redshift.

BIND was trained on halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ inside the $50 \text{ Mpc } h^{-1}$ CAMELS suite of 1024 dark-matter-only and hydrodynamical simulation pairs (S. Genel et al. 2026)⁷. Each simulation in the

suite used random initial phase fluctuations, a set of five cosmological parameters, and thirty galaxy formation model subgrid parameters that control astrophysical effects such as star formation, stellar feedback, and AGN feedback. For a detailed list of the parameters and their descriptions, see S. Genel et al. (2026).

In M. E. Lee et al. (2026a), BIND was trained to generate hydrodynamical density fields such as gas, stellar, and dark matter masses on 128×128 pixel grids centered on DMO halos with pixel scales of $48 \text{ kpc } h^{-1}$ such that the full patch is $6.25 \times 6.25 \text{ Mpc } h^{-1}$. In this work, we retrain the same network with the same map resolution and scale, but also include thermodynamic channels for gas temperature, T , pressure, P , and entropy K , such that the total list of fields generated by BIND is $\mathbf{F}_{\text{hydro}}^H = \{\rho_{\text{DM}}, \rho_{\text{gas}}, \rho_{\text{star}}, T_{\text{gas}}, P_{\text{gas}}, K_{\text{gas}}\}$. It is important to note that BIND is a 2D model, and as such, it generates a projected field of $6.25 \times 6.25 \text{ Mpc } h^{-1}$, where each pixel contains the surface density projected over a $50 \text{ Mpc } h^{-1}$ depth. This decision naturally followed from the CAMELS boxes being $(50 \text{ Mpc } h^{-1})^3$ volumes. However, we will see in § 2.2 that the lightcone generation requires simulation projections, which actually turns BIND’s projection mechanisms into an advantage.

For a detailed explanation of the BIND architecture, training, and validation, see M. E. Lee et al. (2026a). We verified that the retraining described above produces analogous results in the mass channels and will validate the thermodynamic channels in § 3 as well.

2.2. TNG simulations and conditioning generation

The goal of this work is to use BIND to generate weak lensing, tSZ, and τ maps with various astrophysical parameter changes, which requires a dark-matter-only simulation with halos for BIND to paste, and which is large enough to produce these maps. IllustrisTNG (R. Weinberger et al. 2017; A. Pillepich et al. 2018; V. Springel et al. 2018) is the galaxy formation model that BIND was trained on, and so we use the largest IllustrisTNG dark-matter-only simulation, IllustrisTNG-300 Dark (henceforth TNG300-Dark) (D. Nelson et al. 2019). We also use the paired hydrodynamical simulation, IllustrisTNG300 (henceforth TNG300), to validate our procedure. Both TNG300 and TNG300-Dark cover a simulated volume $(205 \text{ Mpc } h^{-1})^3$ with dark matter resolutions of $4.0 \times 10^7 h^{-1} M_{\odot}$ and, $4.73 \times 10^7 h^{-1} M_{\odot}$ respectively, and both are generated with the cosmological parameters inferred from Planck Collaboration et al. (2016) ($\Omega_m = 0.3089$, $\sigma_8 = 0.8158$, $\Omega_b = 0.0486$, and $H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$). The TNG suite contains 100 simulation snapshots output between redshifts $z = 0$ and $z = 20$, each containing halo and subhalo catalogs

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

computed on the fly and saved using the FoF/Subfind algorithms.

This work leverages the TNG300-Dark simulation suite for pasting, but in § 3, we test our procedure by comparing maps generated at the fiducial TNG300 parameter space location with the true TNG300 hydrodynamical simulation. Another advantage of using the IllustrisTNG suite is that several previous works have built and validated ray-tracing pipelines (K. Osato et al. 2021; M. E. Lee et al. 2023, 2026c,b). We treat TNG300 as the ground-truth simulation; however, TNG300 is only a single hydrodynamical realization of a field generated from the fiducial set of 30 astrophysical parameters. In this work with BIND, we can arbitrarily modify any of these parameters to generate different hydrodynamical halo realizations, which we then paste back into the dark-matter-only maps and feed through the ray-tracing pipeline.

To BIND⁸ a simulation, one only requires a projection of a dark-matter-only simulation, a catalog of halo positions within the projection, the cosmological parameters of the simulation, the desired astrophysical sub-grid parameters, and a random seed from which the flow matching starts its sampling. As TNG300-Dark contains the snapshots between $z = 0$ –20, and is a volume of $(205 \text{ Mpc } h^{-1})^3$ per snapshot, we can turn each snapshot into four slabs projected over a depth of $\delta\chi = 51.25 \text{ Mpc } h^{-1}$ and, using Cloud-in-cell (CIC; R. W. Hockney & J. W. Eastwood 1988) interpolation, generate dark matter surface density maps for each, which are 4096×4096 pixels. The pixel resolution and projection depth for each slab then closely match (with a mere 2.5% offset) those used to train BIND, allowing it to be used as conditioning in the model. We find that this small discrepancy doesn’t affect our emulated results, as we show in our validations in § 3.1. To identify the halos, once the pixelized slabs are generated, the FoF catalog positions of halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ can be identified in pixel coordinates, and the conditioning field, ρ_{DMO}^H , for each halo extracted as a 128×128 pixelized grid.

In practice, to build a lightcone, we require only 20 snapshots between $z = 0$ and $z = 2.5$. From these 20 snapshots, we typically want to reduce repeated structures along a given line of sight, which can occur when we naively project along the same axis in each snapshot. To avoid this, we follow K. Osato et al. (2021); M. E. Lee et al. (2026c), using the simulation’s periodic boundary conditions and applying random translations

and rotations to each snapshot before projecting. This ensures that when the slabs are stacked to perform the ray tracing described in § 2.4, there are fewer repeated structures along a line of sight. Each of the 20 snapshots used to generate the lightcone receives a unique rotation, translation, and randomly chosen projection direction (along x, y, or z axes), from which the catalog of halo conditioning grids is extracted and used to generate with BIND.

Once the full halo conditioning catalog is created, we can choose a set of astrophysical parameters, the TNG300 set of cosmological parameters, and the given redshift of the snapshot and generate each of the halos. We show this procedure in Fig. 1, where the leftmost field shows a slab of TNG300-Dark rotated, translated, and projected over $51.25 \text{ Mpc } h^{-1}$ to a 4096×4096 grid. Each small square contains a halo identified by the FoF catalog, for which a 128×128 -pixel patch is extracted, as shown in the center panel. With this conditioning, we can use BIND to draw from the posterior and generate density fields such as dark matter, gas, and stellar density, or thermodynamic fields such as the Compton- y field, for any chosen parameter combination. We show each field for low, fiducial, and high values of a single galaxy formation model parameter⁹, while keeping all other parameters fixed to the TNG fiducial. We can see clear differences right away. The stellar density is significantly suppressed, and the gas concentration and Compton- y become more centrally concentrated in the halo when this parameter is turned up. When the parameter is turned low, there are more small-scale features in all channels, including a dramatic change for the stellar field; however, in the halo core, the gas and Compton- y distributions appear to be more diffuse. For a more detailed exploration of the individual effects of the parameters on the halos, we refer the reader to M. E. Lee et al. (2026a).

2.3. Generating halos across the Sobol sequence

After creating the halo catalog for conditioning described in the previous section, we can generate halos for each slab across all 20 snapshots. The generated halos are then pasted back into the TNG300-Dark simulation following the pasting scheme outlined in §5 of M. E. Lee et al. (2026a), where each generated density field blends into the dark-matter-only density field following a cosine tapering beginning at $3R_{200c}$. For the

⁸ We use BIND as a verb frequently throughout the text, as we believe it is useful and clear.

⁹ While irrelevant for this discussion and Fig. 1, the parameter that is varied is A_{SN1} . We simply wanted to present the pipeline and qualitative changes for a parameter; a full analysis of the effects due to parameters can be found in § 4

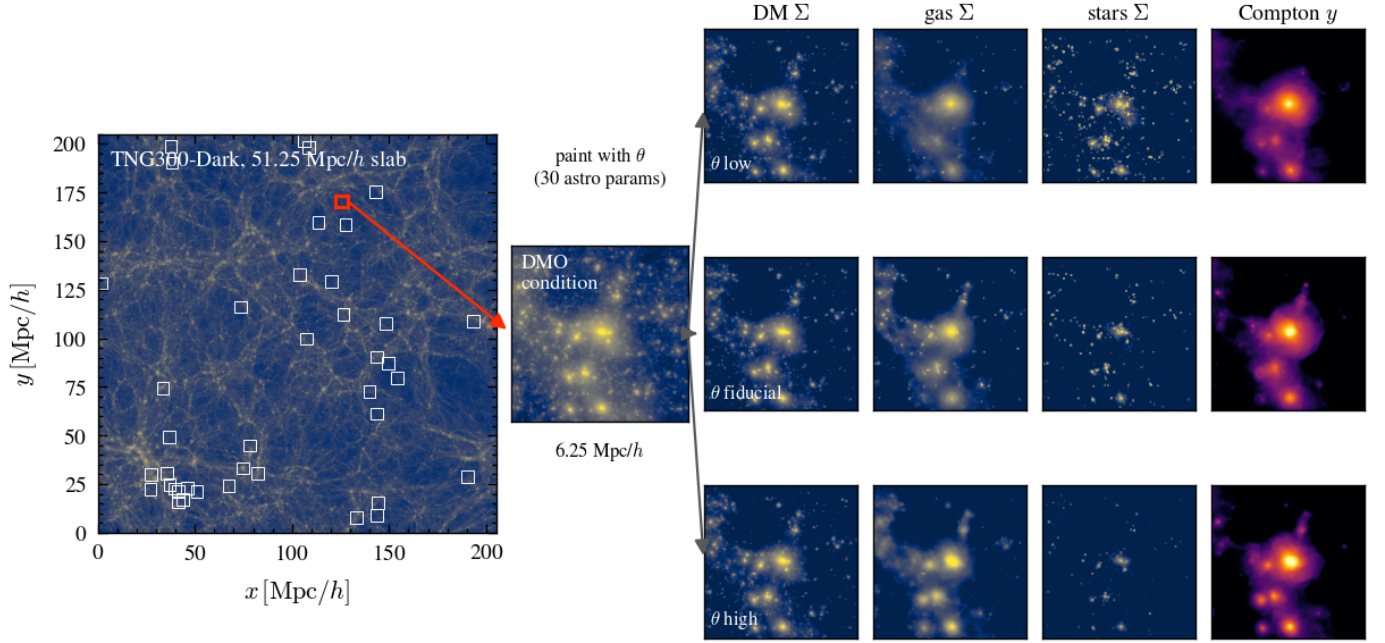


Figure 1. The lightcone BINDing pipeline. *Left:* a $51.25 \text{ Mpc } h^{-1}$ slab of TNG300-Dark at $z = 0.0337$, rotated, translated, and projected onto a 4096×4096 grid. Squares mark halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ identified in the FoF catalog. *Center:* the 128×128 pixel conditioning field ρ_{DMO}^H extracted around one halo. *Right:* fields generated by BIND from this conditioning (dark matter, gas, and stellar surface masses, and the Compton- y field) for low, fiducial, and high values of a chosen parameter while all other parameters are fixed to the TNG fiducial values.

thermodynamic fields, the painted halo gas follows the same tapering. However, because there is no gas outside the painted halo regions in the TNG300-Dark, for the τ and Compton- y fields, we approximate the electron density outside halo regions by multiplying the dark matter field by the baryon fraction. In this way, we paint each slab with baryons centered on the halos and stack the slabs as in K. Osato et al. 2021 to create a lightcone stretching from $z = 0$ to $z = 2.5$. We create five sets of lightcones for use in this work, which we describe below.

- **DMO set:** TNG300-Dark is used with no halo painting to create the lightcone.
- **Hydro-pasted set:** TNG300-Dark has its halos replaced with TNG300 halos to create the lightcone.
- **Fiducial set:** Each halo is generated with the fiducial TNG300 parameters to create the lightcone. This can then be validated against the hydro-pasted set.
- **1P set:** Each halo is generated with all TNG300 parameters at their fiducial values except for one of the 30 parameters, which is set to one of its prior bounds from S. Genel et al. (2026). This yields 57 lightcones (rather than 60, as three parameters have their fiducial values at a prior bound, see S.

Genel et al. (2026) for details) at different locations in parameter space, allowing us to isolate the effects of individual parameters.

- **Sobol set:** Each of the 30 astrophysics parameters is varied simultaneously over a 256-node Sobol sequence to generate 256 lightcones. This is a smaller version of the S. Genel et al. (2026) Sobol sequence, which contains 1024 simulations over a 35-dimensional parameter space.

In all of the sets described above, the cosmological parameters are fixed to the TNG300 fiducial values, and only the 30 astrophysical parameters are varied. This is because generating halos with DMO in a different cosmology would be out-of-distribution from the training set. BIND is not a tool for cosmological rescaling, but could in the future be used on rescaled cosmological simulations (R. E. Angulo & S. D. M. White 2010; M. Zenaro et al. 2019).

We also generate our lightcones for each set in the same way. They contain the same rotations, translations, and projections of the snapshots and the same set of replaced halos (either with the generative model's field or the true hydro field). Further, each halo has a unique random seed for flow matching, used to generate it once per parameter-space location, and that seed is shared across all lightcones, so differences between lightcones are purely parameter-driven. The result is then

314 lightcones at different parameter space locations that use BIND and a single lightcone from the TNG300-Dark simulation, and a single lightcone that has replaced the TNG300-Dark halos of $M_{200c} \geq 10^{13} h^{-1} M_{\odot}$ with their true hydrodynamical field counterparts.¹⁰

2.4. Ray tracing and map generation

We follow the procedure of K. Osato et al. (2021), which uses the multiple lens-plane ray-tracing algorithm (B. Jain et al. 2000; S. Hilbert et al. 2009; A. Petri 2016) to account for lensing deflections and generate the convergence κ , Compton- y , and τ maps. We ray-trace each of the surface density stacks described in the previous section, where each slab in a stack, projected over a depth of $\delta\chi = 51.25 h^{-1} \text{Mpc}$, has surface density Σ after the pasting procedure. From each slab we compute the lensing potential at plane k via

$$\nabla_{\beta^k}^2 \psi^k(\beta^k) = 2\Sigma^k(\beta^k), \quad (2)$$

where β^k is the angular position on the k^{th} plane. At each plane we then compute the deflection angle α ,

$$\nabla_{\beta^k} \psi^k(\beta^k) = \alpha^k(\beta^k), \quad (3)$$

which, summed across all planes, gives the final lensed position of the ray given the original pointing angle on the sky ϕ ,

$$\beta^k(\phi) = \phi - \sum_{i=1}^{k-1} \frac{\chi^k - \chi^i}{\chi^k} \alpha^i(\beta^i), \quad (k = 2, 3, \dots), \quad (4)$$

where χ^i is the comoving distance at the i^{th} lensing plane, and χ^k is the comoving distance of the source. The lensed position at a given pointing angle from the observer, and out to a comoving distance is then related to the convergence and shear of the field via the lensing Jacobian,

$$A_{ij}(\phi, \chi) \equiv \frac{\partial \beta_i(\phi, \chi)}{\partial \phi_j}, \quad (5)$$

from which the convergence follows as $\kappa = 1 - \frac{1}{2}(A_{11} + A_{22})$ (for the full recursion and higher-order terms, we refer the reader to K. Osato et al. 2021 and S. Hilbert et al. 2009). For each surface density stack, we generate 1000 pseudo-independent lightcones of 25 deg^2 by following the same rotation convention as K. Osato et al. (2021), randomly rotating each slab by 0° , 90° , 180° , or 270° about each of the three axes and translating the pixels in each direction. This rotation and translation process allows for a large statistical sample of map

realizations, with up to 10,000 pseudo-independent realizations demonstrated in A. Petri et al. (2016). The resulting κ , y , and τ maps are output on 1024^2 pixel grids, downsampled from the 4096^2 planes following the same scheme as K. Osato et al. (2021), at five source redshifts $z_s = 0.5, 1.0, 1.5, 2.0$, and 2.44 .

In the ray tracing procedure, the convergence is determined by the weighted deflections accumulated by a ray along its line of sight. However, we can follow the same light ray and, instead of computing a lensing term, simply sum the gas density in each pixel the ray crosses. Accounting for the ionization state of the gas, this amounts to an integral of the ionized gas along the (deflected) path of the ray, the optical depth,

$$\tau(< z_s) = \sigma_T \int_0^{z_s} n_e dl_{\text{prop}} = \frac{\sigma_T x_e}{m_p} \sum_k \Sigma_{\text{gas}}^k, \quad (6)$$

where Σ_{gas}^k is the gas surface density of plane k evaluated along the deflected ray, and $x_e = X_{\text{H}} + Y_{\text{He}}/2 = 0.88$ electrons per proton mass, assuming fully ionized gas of primordial composition ($X_{\text{H}} = 0.76$, $Y_{\text{He}} = 0.24$). The end result is a map of the optical depth accumulated to a given source redshift along the same line of sight, deflected by gravitational lensing. Similarly, as the Compton- y parameter is the integrated electron pressure along the line of sight, we follow the same ray and sum the pressure planes,

$$y(< z_s) = \frac{\sigma_T}{m_e c^2} \int_0^{z_s} P_e dl_{\text{prop}} = \frac{\sigma_T}{m_e c^2} \sum_k \Sigma_P^k, \quad (7)$$

where Σ_P^k is the integrated electron pressure of plane k . In practice, the pipeline evaluates the electron pressure directly from the generated gas density and temperature channels, $P_e = n_e k_B T_e$, with the ionization prefactors given above, i.e. $P_e = \frac{2+2X_{\text{H}}}{3+5X_{\text{H}}} P_{\text{th}} \approx 0.52 P_{\text{th}}$ for the total thermal pressure P_{th} of a fully ionized primordial gas.

Because κ , τ , and y are accumulated along the same deflected rays and output at the same source redshifts, every released map triplet of κ , τ , and y is mutually consistent.

Of course, the optical depth τ and Compton- y maps suffer from the fact that the box is not entirely filled with hydrodynamical gas. Only regions centered on the $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ halos contain painted gas tapering to an approximation of the gas density outside by using the dark matter field and cosmological baryon fraction, and a constant temperature approximation of 10^4 K , roughly the temperature to which the photoionized intergalactic medium is held by the ultraviolet background (M. McQuinn 2016), and consistent with the cool diffuse phase that dominates the gas outside massive halos in

¹⁰ For more information on this replacement procedure, see M. E. Lee et al. (2026b).

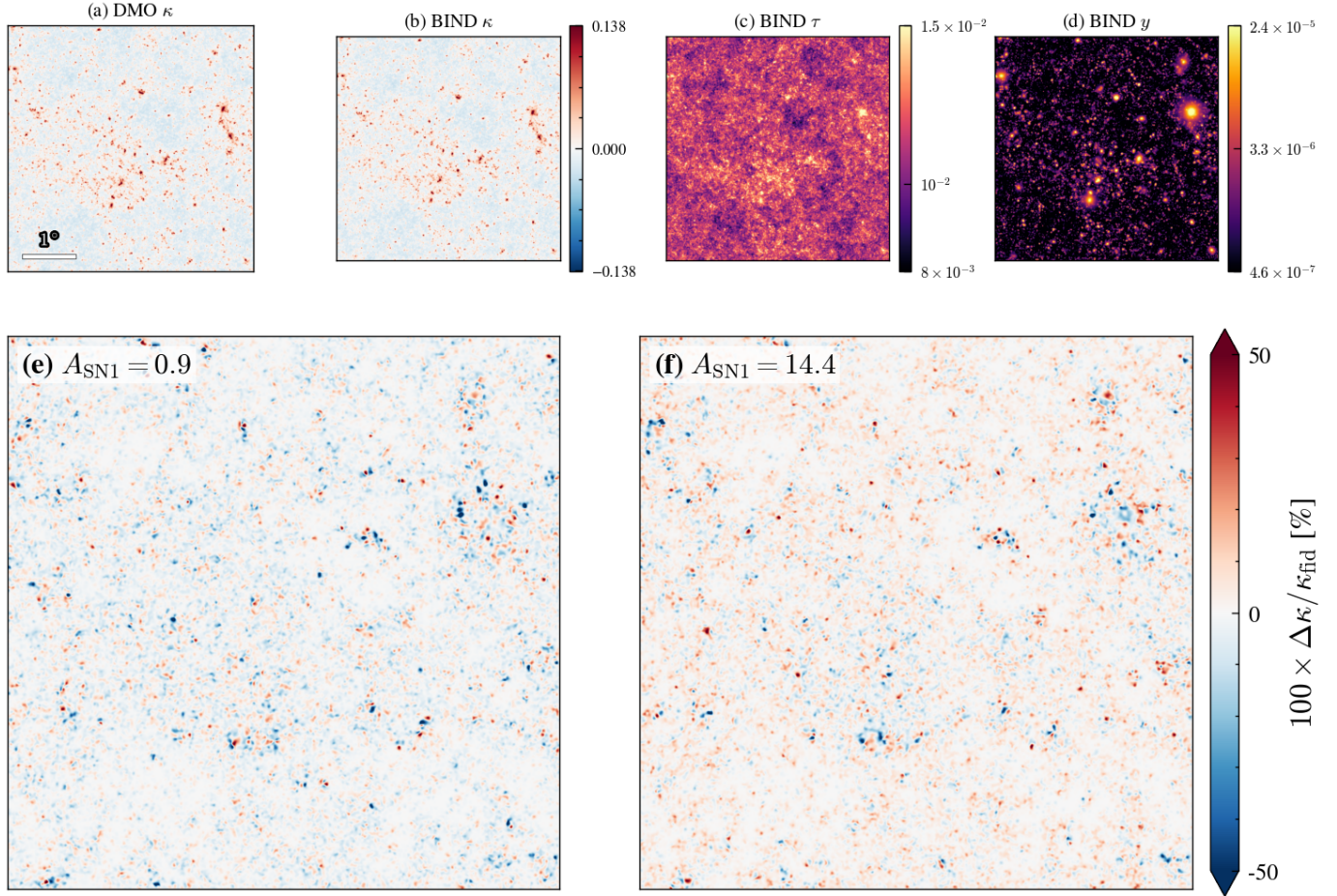


Figure 2. Example maps from the pipeline described in § 2.4. *Top row left to right:* a dark-matter-only convergence map, and a BInded convergence κ , optical depth τ and Compton- y map generated at the fiducial parameter space location for TNG300. The halos are consistently traced, with peaks appearing at the same locations across all three observables. *Bottom row:* differences between convergence maps with a single parameter set to each of its two prior bounds while all others are fixed to the TNG300 fiducial, normalized by the fiducial convergence map, with clear differences appearing around the halos. All maps shown in this plot are at $z = 1$

IllustrisTNG (D. Martizzi et al. 2019). However, each of the halo regions contains the integrated gas and pressure across the full $51.25 \text{ Mpc } h^{-1}$ slab, allowing a compensated aperture photometry (CAP; E. Schaen et al. 2021; S. Amodeo et al. 2021) style evaluation of the Compton- y and τ signals of halos in the final maps. In future work, we plan to use these maps to perform CAP-style measurements of the halos along the lightcone.

We further note that we do not model any velocities. Our maps are of the optical depth itself, and the translation to a kSZ observation, or to a dispersion-measure observation in the case of fast radio bursts, is left to the observational side, where velocity assumptions or measurements come into play. We return to both modeling statements and their consequences for which statistics the maps support in § 7.

In Fig. 2, we show examples of the resulting maps from the procedure described in this section. The left top

panels show a dark-matter-only convergence map compared to the fiducial BInded convergence map. The top right two panels show the BInded optical depth and Compton- y maps. The halos are traced consistently across the maps, with peaks appearing in the convergence, τ , and Compton- y maps at the same locations. In the bottom two panels, we hint at the effects of changing the parameters, which are discussed in § 4, by showing the difference between convergence maps where a single parameter is set to each of its two prior bounds (all others at the fiducial) and the fiducial convergence map. Clear differences appear, particularly around the halos, suggesting that variations in individual parameters propagate to the full weak lensing maps and are likely observable in the resulting summary statistics.

3. VALIDATION

In this section, we perform a series of validations comparing the fiducial set to the hydro-pasted set. We begin with halo-level relations, computed at the snapshot level before ray tracing, then move to map-level validations. Because TNG300 is only run at a single parameter space location, we are limited to comparing our maps at only this fiducial parameter space location, however, in the companion paper, [M. E. Lee et al. \(2026a\)](#), we focused on applying a battery of tests and showed that BIND accurately captured the effects of IllustrisTNG model parameters on the generated fields. Ideally, we would compare against a full suite of parameter-varying, ray-traced maps, but this is currently computationally infeasible.

One subtlety for the following sections is that they are out-of-distribution tests. The TNG300-Dark projections are sourced from the TNG300 box, which is $205 \text{ Mpc } h^{-1}$. Breaking this into four surface density slices following § 2.4 means that each projection is $\sim 51 \text{ Mpc } h^{-1}$ containing an extra $\sim 1 \text{ Mpc } h^{-1}$ of matter in the projection compared to the training set used for BIND in [M. E. Lee et al. 2026a](#). Before performing any ray tracing, we need to be sure that this out-of-distribution conditioning can still be used with BIND.

3.1. Halo-level validation

We first validate the BINDed halos by comparing various relations to the true TNG300 hydro halos. We illustrate results only at redshift $z = 0.0337$ for clarity of presentation, although we have performed the same comparisons at higher redshifts and find similar results (we also explore redshift validation and effects in the companion paper, [M. E. Lee et al. 2026a](#)). At $z = 0.0337$, we identify 2933 halos with $M_{200c} \geq 10^{13} \text{ M}_{\odot} h^{-1}$, which are used in the following analyses.

We first show the $Y - M$, $\tau - M$, and $M_* - M$ scaling relations for both the BINDed and TNG300 halos in Fig. 3. Each quantity is computed as a sum over the projected R_{200c} of each halo. Of course, these relations include foreground and background material generated in the full projection, making them “contaminated” halo scaling relations. We paste these patches into the multi-lensplane algorithm of § 2.4, which is itself a projection, so we want to validate the full projection; these contaminated relations are a useful metric for that validation.

Each panel shows halos in ten bins evenly log-spaced between the mass threshold of $M_{200c} = 10^{13} \text{ M}_{\odot} h^{-1}$ and the most massive halo in the sample at $M_{200c} = 10^{14.98} \text{ M}_{\odot} h^{-1}$. The solid line represents the median in a given halo mass bin, and the shaded regions, representing the scatter, are computed as the 16th–84th per-

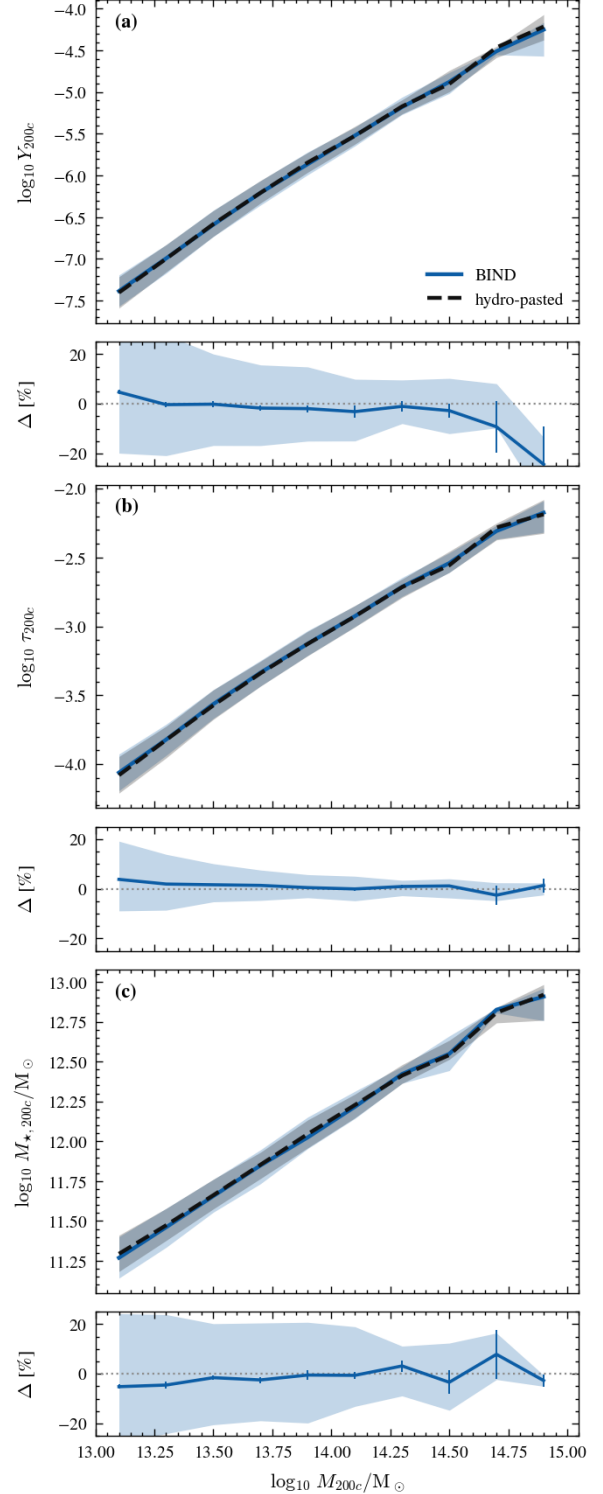


Figure 3. the $Y_{200c} - M$, $\tau_{200c} - M$, and $M_{*,200c} - M$ relations for the BINDed halos (blue) and the TNG300 halos (black dashed) computed within R_{200c} at redshift $z = 0.0337$. The lines are computed as the median across all halos in ten evenly log-spaced halo mass bins; the shaded bands show the 16th–84th percentile halo-to-halo scatter. Below each panel, we show the relative difference, $\Delta = (\text{BIND} - \text{hydro-pasted})/\text{hydro-pasted}$, with error bars computed as bootstrap uncertainty on the median. Masses on the horizontal axis are in $\text{M}_{\odot} h^{-1}$, as in the text.

centiles. Below each scaling relation panel, we show the relative difference between the BINDED halos and TNG300 halos $\Delta = (\text{BIND} - \text{TNG300})/\text{TNG300}$. Error bars in the lower panels are computed with bootstrap resampling to estimate median uncertainty in each bin.

In each panel, we see that the bootstrap uncertainty increases at the highest masses; however, these bins also contain few halos: 149 above $10^{14} M_{\odot} h^{-1}$, 14 above $3 \times 10^{14} M_{\odot} h^{-1}$, and only 6 above $4 \times 10^{14} M_{\odot} h^{-1}$. BIND is a generative model that draws from a conditional distribution, and the resulting draw may differ by up to the scatter associated with a given halo mass. Most bins show that BIND captures this halo-to-halo scatter well, with small biases; however, it clearly underpredicts the Compton- y parameter in the highest-mass bin. We suspect this is a training-set effect rather than a ray-tracing one. The CAMELS boxes are $(50 \text{ Mpc } h^{-1})^3$, so halos above a few $\times 10^{14} M_{\odot} h^{-1}$ are rare in training, and the Y - M relation is the steepest of the three ($Y \propto M^{5/3}$ for self-similar halos, against roughly linear for τ and $M_{\star}$), so it is the relation where a draw that leans toward the better-sampled, lower-mass part of the conditional distribution costs the most. Fig. 4 shows where in the halo this deficit sits.

In Fig. 4, we turn to halo profiles in four mass bins and compare them between the BINDED and TNG300 halos. We compute radial profiles for each halo by integrating the masses or pressures within fifteen radial bins from $0.1R_{200c}$ to $3R_{200c}$. Each profile begins at the pixel scale, so the curves for the different mass bins start at different fractional R_{200c} radii. Each line again is the median in the mass bin, and the scatter is the 16th–84th percentiles for a given quantity.

Similar to Fig. 3, we find a close agreement across the profile quantities and mass bins. The gas-related profiles (the top two quantities) agree to within a few percent inside R_{200c} in every mass bin. The one exception is the y profile of the highest-mass bin, which falls 5–10% low between R_{200c} and $3R_{200c}$. This is the Y - M deficit discussed in Fig. 3, where BIND reproduces the pressure in the cores of the most massive halos and underproduces it in their outskirts, where the projected pressure is carried by extended, low-density gas that the model sees least often in training. The stellar channel shows a clear bias, particularly at intermediate radii and out beyond R_{200c} . Although its integrated mass is only $\sim 6\%$ low for the lowest mass bins, the $\Sigma_{\star}$ profiles sit 25–40% low in individual annuli, meaning BIND misplaces stellar mass radially more than it underproduces it. This is consistent with the stellar channel being the hardest to reconstruct because of its sparsity (M. E. Lee et al. 2026a). As stars contribute only a small fraction

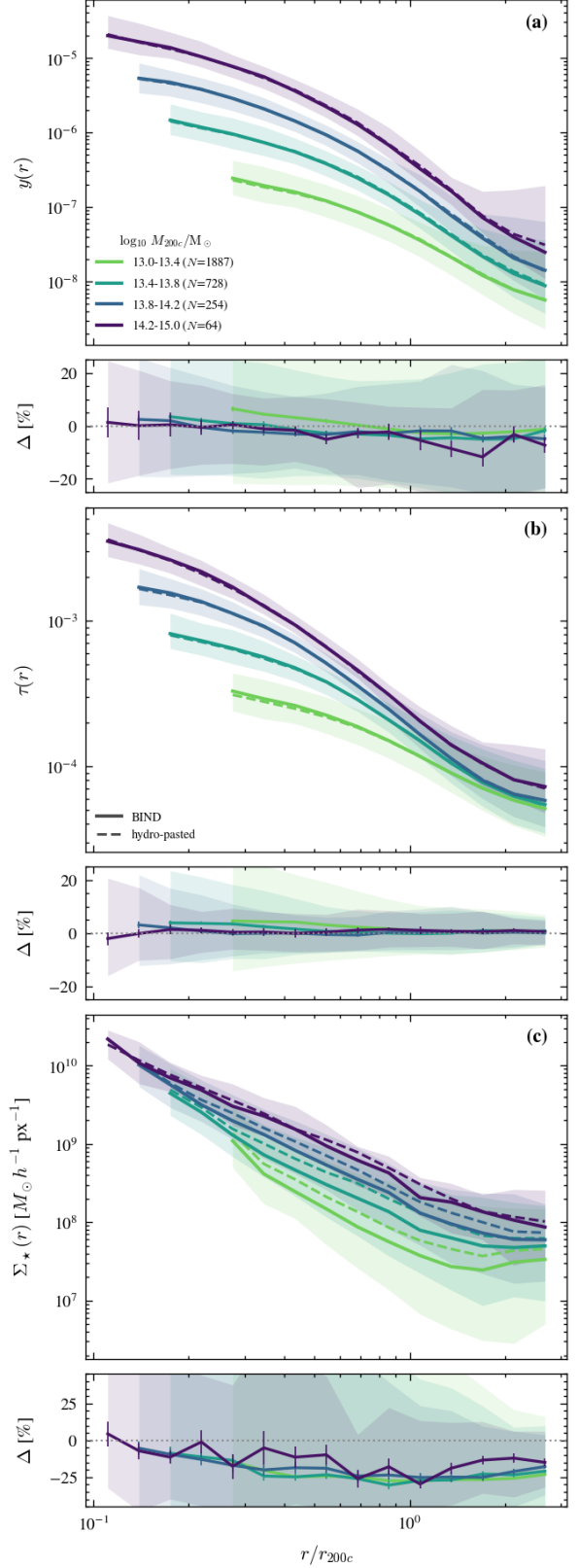


Figure 4. Radial profiles for y , τ , and $\Sigma_{\star}$ for the BINDED (solid) and TNG300 (dashed) halos at $z = 0.0337$, in four mass bins shown as different colors. Similar to Fig. 3, the bands show the 16th–84th percentile halo-to-halo scatter, and the error bars are computed with bootstrap resampling. Each curve begins at the pixel scale of its mass bin and extends to $3R_{200c}$, where the taper for pasting into the dark-matter-only simulation begins (see § 2.3 for more details). Masses in the legend are in $M_{\odot} h^{-1}$, as in the text.

of the total projected mass, this does not appreciably affect the κ maps.

3.2. Map-level validation

We now validate the BINDed fiducial maps against the hydro-pasted maps. Recall that, as described in § 2.4, the fiducial and hydro-pasted sets contain 1000 pseudo-independent maps at five redshifts. For each map and redshift, we compute several weak-lensing statistics, including both Gaussian and non-Gaussian statistics. No smoothing is applied to any map, and all statistics are computed on the native 1024^2 pixel grid (0.293 arcmin pixels).

For each convergence map, we calculate the angular power spectrum,

$$C^{\kappa\kappa}(\ell_i) = \langle |\tilde{\kappa}(\ell)|^2 \rangle_{\ell \in [\ell_i^{\min}, \ell_i^{\max}]}, \quad (8)$$

where $\tilde{\kappa}(\ell)$ is the two-dimensional Fourier transform of the convergence field $\kappa(\phi)$ in the flat-sky approximation, so that ℓ is the two-dimensional wavevector conjugate to ϕ rather than a spherical-harmonic index (ℓ, m) , and the angle brackets denote averaging over all Fourier modes ℓ within the multipole bin $[\ell_i^{\min}, \ell_i^{\max}]$. From the TNG300-Dark set of maps, we define the power spectrum suppression,

$$S(\ell) = \frac{C_{\text{BIND}}^{\kappa\kappa}(\ell)}{C_{\text{DMO}}^{\kappa\kappa}(\ell)}, \quad (9)$$

computed realization by realization. Because each map between simulation sets contains matched rotations, translations, and generative seeds, the power-spectrum suppression division cancels realization noise in the ratio.

We use the *LensTools* package (A. Petri 2016)¹¹ to compute non-Gaussian statistics from the sets of convergence maps. All are computed as functions of the signal-to-noise ratio $\nu = \kappa/\sigma_0$, where σ_0 is the root-mean-square convergence of the *fiducial* map set, held fixed across all map sets so that bins remain directly comparable between runs. We use the same set of statistics as in M. E. Lee et al. (2026b),

1. **Peak counts** $N_{\text{pk}}(\nu)$: The number of local maxima in the convergence field, defined as pixels with greater values than their 8 nearest neighbors (J. M. Kratochvil et al. 2010; M. Maturi et al. 2010).
2. **Minimum counts** $N_{\text{min}}(\nu)$: The number of local minima in the convergence field, defined as pixels with smaller values than their 8 nearest neighbors (W. R. Coulton et al. 2020).

3. **One-point PDF** $p(\nu)$: The probability distribution of convergence values, representing the full histogram of ν values (L. Thiele et al. 2020a).
4. **Minkowski functionals** (MFs): Three descriptors of convergence excursion sets (J. M. Kratochvil et al. 2012; A. Petri et al. 2013) (area $V_0(\nu)$, boundary length $V_1(\nu)$, and genus $V_2(\nu)$) that encode shape information.

For the Compton- y and τ maps, we compute the two-point auto- and cross power spectra with the same estimator,

$$C^{AB}(\ell_i) = \langle \text{Re}[\tilde{A}(\ell)\tilde{B}^*(\ell)] \rangle_{\ell \in [\ell_i^{\min}, \ell_i^{\max}]}, \quad (10)$$

$A, B \in \{\kappa, y, \tau\},$

giving C^{yy} , $C^{\tau\tau}$, $C^{\kappa y}$, $C^{\kappa\tau}$, and $C^{y\tau}$ in addition to $C^{\kappa\kappa}$.

Each ℓ -statistic is computed in log-spaced bins between $\ell = 10^2$ and 4×10^4 . The Nyquist mode of the 1024^2 maps is $\ell_{\text{Ny}} = \pi/\theta_{\text{pix}} \approx 3.7 \times 10^4$, and we only present results up to this Nyquist mode. For the non-Gaussian statistics, we compute all of them in 22 bins between $-2 \leq \nu \leq 8$.

Note that the hydro-pasted maps are **not** the full TNG300 hydro ray-traced maps. Instead, they are TNG300-Dark maps where each halo with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ has been replaced exactly with the dark matter, stellar, and gas particles of the TNG300 simulation. This is the same procedure that was followed in M. E. Lee et al. (2026b), where the authors found that replacing halos of $10^{13} M_{\odot} h^{-1}$ and above accounts for ~ 60 – 65% of the response between dark-matter-only and hydrodynamical simulations at the level of the power spectrum, but virtually all of the response for peak counts and Minkowski functionals. We return to the consequences of this construction in § 7. The comparison here is also noiseless, meaning no shape noise, beam, or survey systematics enter at any stage. By comparing BINDed with the hydro-pasted maps, we are able to directly observe this halo-centric and field-level generative model approach in an apples-to-apples comparison.

In Fig. 5, we compare the weak lensing statistics between the hydro-pasted and BIND maps at source redshift $z_s = 1.0$. We generate maps at all five source redshifts but limit the comparison to $z_s = 1$ for clarity. However, the results are comparable at the other redshifts. Each figure shows the given statistic in the top panels, while the bottom panels show the residual at $z_s = 1$.

In the bottom panels, we assess the statistical distinguishability of the statistics given an LSST-like covariance matrix. To compute this, we follow a similar procedure to M. E. Lee et al. (2023), and use the hydro-pasted

¹¹ lenstools.readthedocs.io

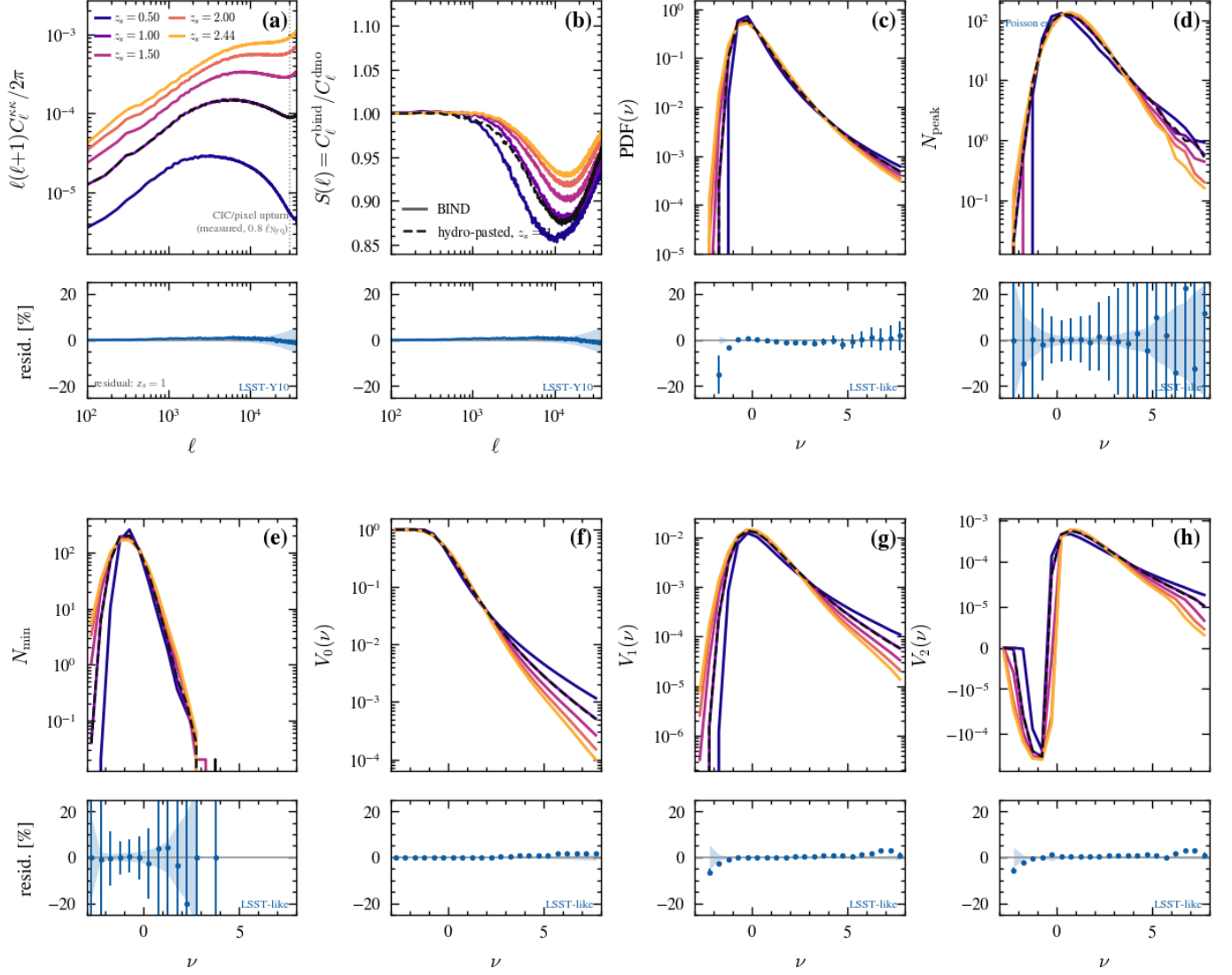


Figure 5. Weak lensing statistics for the BINDed (solid) and hydro-pasted (dashed) maps at source redshift $z_s = 1$. The weak lensing angular power spectrum (a), the power spectrum suppression $S(\ell)$ (b), the one-point convergence PDF (c), weak lensing peak counts (d), minimum counts (e), and the three Minkowski functionals (f-h). Panels (a, b) are functions of multipole ℓ ; panels (c-h) are functions of the signal-to-noise $\nu = \kappa/\sigma_0$. Bottom panels show the mean residual between BIND and hydro-pasted at the same ℓ or ν , with error bars scaled to an LSST-Y10-like survey area of $\sim 18,000 \text{ deg}^2$ following Eq. 13.

convergence maps to estimate, for a given statistic,

$$C_{ij} = \frac{1}{N-1} \sum_{k=1}^N \left(n_i^{(k)} - \langle n_i \rangle \right) \left(n_j^{(k)} - \langle n_j \rangle \right), \quad (11)$$

where $n_i^{(k)}$ is the statistic measured in the k^{th} hydro-pasted realization, the sum runs over the $N = 1000$ pseudo-independent realizations, and the subscripts refer to the ℓ or ν bins of the data vector. When a precision matrix is required, we debias it as

$$\hat{C}^{-1} = \frac{N-d-2}{N-1} C^{-1}, \quad (12)$$

where d is the length of the data vector (J. Hartlap et al. 2007). The lightcones are limited to 25 deg^2 , but we can

test statistical distinguishability for LSST-like areas A by scaling the covariance matrix as

$$C(A) = \left(\frac{A}{25 \text{ deg}^2} \right)^{-1} C(25 \text{ deg}^2), \quad (13)$$

with $A = 18,000 \text{ deg}^2$ for LSST-Y10. This should not be read as a realistic prediction for LSST, which would require super-sample covariance and the inclusion of many other systematics; however, it is a good test of statistical distinguishability against the hydro-pasted model with an LSST-like covariance.

The bottom panels of each plot show the mean residual with error bars associated with an LSST-Y10-like survey, with Poisson error bars for the number counts

in a given bin. We can read off by eye the statistical distinguishability relative to an LSST-like survey by comparing the residual lines to the error contours in the bottom panels, and we clearly see that the BINDed statistics match the hydro-pasted maps to within LSST-like precision for all weak lensing statistics.

Beyond weak lensing statistics, we are interested in the trend in baryon correction models used to generate thermodynamic predictions on the lightcone, as in D. Anbajagane et al. (2024) and A. Schneider et al. (2025). We compare our generated τ and Compton- y maps with respect to the hydro-pasted set at $z_s = 1$ in Fig. 6. We follow the same format as above and show BINDed maps at each source redshift, with comparisons at $z_s = 1$ to the hydro-pasted maps. The top panels show the three auto power spectra, while the bottom three panels show the cross spectra. In the absence of an LSST-like analog for distinguishability, we show the percent error of the prediction relative to the truth in the bottom panels, with the standard error on the median curve. Unlike the weak lensing statistics, the thermodynamic spectra, both auto and cross, are more biased. The yy spectrum is the worst case: it sits $\sim 20\%$ low at $\ell \simeq 10^2$, crosses zero near $\ell \simeq 5 \times 10^3$, and rises to $\sim 20\%$ high at the Nyquist scale. The tSZ power at low ℓ is dominated by the most massive halos, and C_ℓ^{yy} scales as Y^2 , so the 10–15% underprediction of Y in the highest-mass bin, concentrated in the halo outskirts, is enough to account for the $\sim 20\%$ deficit. The τ spectra trace gas mass rather than pressure and have no such deficit at the halo level, and are correspondingly less biased.

Given these biases, it is fair to ask what the τ and Compton- y maps are useful for. We are primarily interested in how the statistics respond to parameter variations, and the results in §§ 4–5 are largely insensitive to these offsets. The rank correlations of § 4.1 are invariant to any bin-wise bias constant across the parameters. Moreover, the biases are an order of magnitude below the actual feedback response found in § 4, so the maps resolve feedback responses at high contrast. However, predictions of the absolute C_ℓ^{yy} or the mean Compton- y inherit these offsets, as well as the approximated diffuse gas discussed in § 7.

Finally, we emphasize again that the validation above is performed only at the fiducial parameters because TNG300 provides the only available hydrodynamical truth and is available only at this parameter-space location. The parameter response of BIND itself, that the generated halos move correctly as the parameters vary, was validated against the CAMELS simulations in M. E. Lee et al. (2026a), where held-out parameter locations were recovered across the prior volume, and correlations

between generated quantities and the parameters were matched almost perfectly. Validation in both this section and the companion paper supports the validity of the parameter variations we now turn to. In the following section, we use the Sobol and 1P sets to measure how weak lensing and gas statistics respond across the full 30-dimensional astrophysical parameter space.

4. ASTROPHYSICAL EFFECTS

In the previous sections, we showed that BIND generates lightcones comparable to TNG300 at the fiducial parameter-space location; we now turn to the effects on weak lensing and CMB statistics as functions of IllustrisTNG galaxy formation model parameters.

In this work, we do not explore varying cosmologies or correlations between astrophysics and cosmological parameters. TNG300-Dark was run with a single set of cosmological parameters, so we only explore astrophysical parameter variations independently of cosmology here. In future work, we will consider the effect of cosmology by applying cosmological rescaling methods or by running additional dark matter simulations with varying cosmologies to use as conditioning.

Our first question is whether, at a fixed cosmology, the IllustrisTNG galaxy formation model parameters induce statistically distinguishable effects for an LSST-like survey. Current gas and lensing data already prefer stronger feedback than the TNG300 fiducial calibration (L. Bigwood et al. 2024; B. Hadzhiyska et al. 2024; J. Siegel et al. 2025a; L. Bigwood et al. 2025b), but with the generated Sobol set of lensing maps we can directly test whether parameter changes induce a detectable shift in weak lensing statistics given LSST-like uncertainty. As a reminder, the amplitude of the power spectrum suppression is limited by the incomplete pasting, which only includes more massive halos of $M \geq 10^{13} M_\odot h^{-1}$, but the relative variation of statistics with respect to the parameters is accurate.

For this measurement, we assume that the TNG300 fiducial parameter-space location is the “true” observed data, and we build the survey precision exactly as in § 3.2 with the covariance of $C_\ell^{\kappa\kappa}$ measured across the 1,000 hydro-pasted realizations and scaled to the area of an LSST-like survey (exactly as in Eq. 13).

Fig. 7 panel (a) shows the extent of the power spectrum suppressions from the entire Sobol set in gray, the fiducial suppression in black, and the two survey precisions as envelopes drawn on the fiducial. Panel (b) shows the same information but rewritten as a detection significance per band-power for each of the Sobol nodes, $|S_{\text{node}} - S_{\text{fid}}|/\sigma$. We take the uncertainty, σ , to be

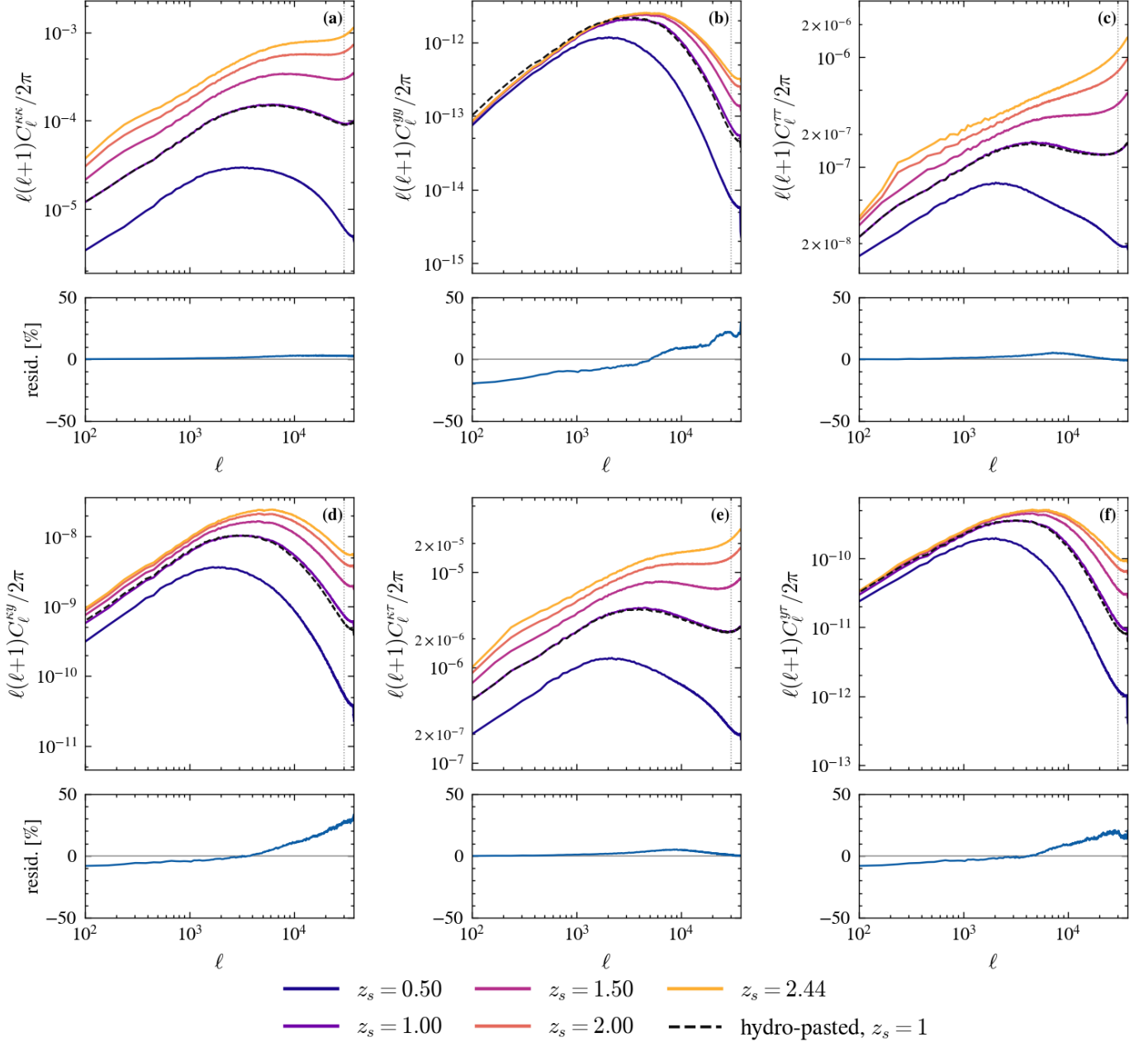


Figure 6. Auto and cross power spectra of the BINDED maps at all five source redshifts (colors), compared with the hydro-pasted maps at $z_s = 1$ (black dashed), though each panel shows all five source redshifts. The $\kappa\kappa$ (a), $\gamma\gamma$ (b), and $\tau\tau$ (c) auto spectra, and the $\kappa\gamma$ (d), $\kappa\tau$ (e), and $\gamma\tau$ (f) cross spectra. Residual panels show the $z_s = 1$ comparison. Unlike Fig. 5, no survey scaling is applied, as no LSST-like covariance analog exists for the gas observables.

the square root of the diagonal of the covariance matrix (Eq. 11).

Over $\ell \simeq 2.2 \times 10^3$ – 1.8×10^4 , the Sobol set prior spread exceeds the LSST-Y10 precision by more than 10σ , peaking at 22σ near $\ell \simeq 5 \times 10^3$. For *Euclid* the window is narrower owing to the smaller area of the survey, with $\ell \simeq 3 \times 10^3$ – 1.3×10^4 and a peak of 17σ . All 256 parameter space locations of the Sobol set vary from the fiducial by more than 1σ somewhere in

the range of $300 < \ell < 0.8\ell_{\text{Ny}}$, and 84% (73%) of the prior is statistically distinguishable by more than 5σ for LSST-Y10 (*Euclid*). Ultimately, this plot shows that the galaxy formation model parameter space yields two-point statistics that are distinguishable from the ground truth for an LSST- or *Euclid*-like survey. This holds beyond power-spectrum suppression; in fact, we find the same for all statistics we study in this work. We show the power-spectrum suppression to connect with previ-

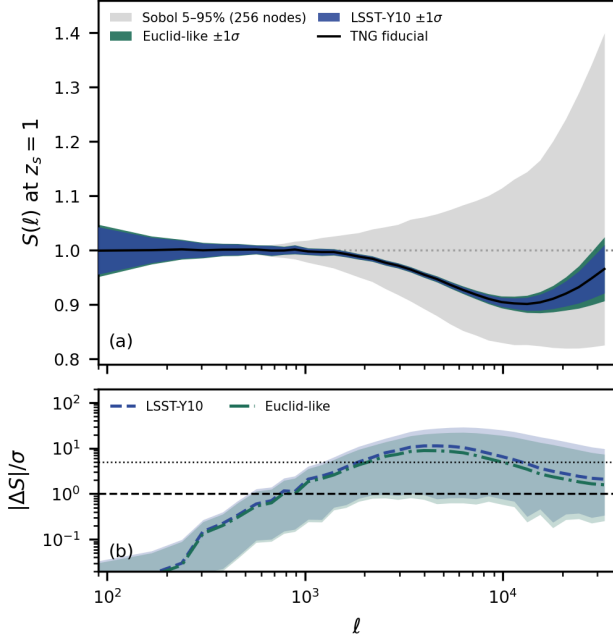


Figure 7. The power spectrum suppression, $S(\ell)$, across 256 IllustrisTNG model parameter space locations at redshift $z_s = 1$ with respect to LSST-Y10-like uncertainty. **(a)** The 5–95% spread of $S(\ell)$ for the Sobol sequence (gray), the fiducial (black), and the $\pm 1\sigma$ LSST-Y10 and Euclid-like precisions drawn on the fiducial. **(b)** The same Sobol sequence spread, but shown as a measure of significance of statistical distinguishability $|S_{\text{node}} - S_{\text{fid}}|/\sigma$. The bands are the 5–95% range across nodes, while the lines represent the medians.

ous literature (M. P. van Daalen et al. 2011; M. L. van Loon & M. P. van Daalen 2024; M. P. van Daalen et al. 2025).

In the following sections, we examine which TNG300 parameters drive the effects we see in the various weak lensing statistics. We start by looking for correlations between the statistics and parameter changes, as in M. E. Lee et al. (2026a). We then consider if individual parameter variations fall into similar families of effects and hypothesize about the physical reasoning behind this.

4.1. Correlations across the Sobol set

Our goal is to address a few concrete questions: which galaxy model parameters impact which statistics? Is this consistent across all statistics, or does each have unique sensitivities? For hypotheses on the physical origins of these effects, see § 6.

Following a similar routine to M. E. Lee et al. (2026a), we compute the signed Spearman rank correlation $\rho[\theta_j, n_i]$ between each of the 30 parameters and each bin of each measured statistic across the 256 Sobol nodes, at $z_s = 1$. For each (statistic, parameter) pair,

we then extract the peak $|\rho|$ across that statistic’s bins, restricted to the trusted scale range. Fig. 8 shows this construction for the power spectrum suppression alone. Each column is one of the thirty parameters, each row an ℓ bin, and the circles mark the bin with the strongest correlation for each parameter. Some parameters clearly induce effects more prominently on smaller scales, while others on larger angular scales. Some parameters also appear to have little to no effect across the full spectrum.

However, we now want to consider the 12 commonly explored statistics¹² introduced in § 3.2, each with its own bins in ℓ or ν , its own noisiness, and its own correlations across bins. One could imagine eleven more versions of Fig. 6, each for a different statistic.

A maximum as represented by the open circles in Fig. 8 taken over many bins is biased high. Even a parameter that affects nothing still draws one chance correlation per bin, and by taking the peak, we report its luckiest. A search of this size can therefore be fooled by peaks produced by random fluctuations or a shared systematic effect. What we are really after is a map of parameters against statistics in which a peak correlation appears only if chance alone would not have produced it. To this end, we subject the peak correlations to three levels of statistical rigor.

1. *Tile level, from the scatter across Sobol nodes.* With $N = 256$ nodes, a single *pre-chosen* bin of a statistic that responds to nothing still shows chance correlations of order $1/\sqrt{N}$. We take the 2σ level, $|\rho| = 2/\sqrt{256} = 0.125$. This level applies before any bin search.
2. *Row level, from the look-elsewhere across a statistic’s own bins.* Neighboring bins in either ℓ or ν for a statistic are correlated, so the number of effectively independent bins matters when considering the bin with peak correlation. We measure this empirically by randomly assigning the Sobol parameter vectors to the measured statistics, breaking the connection between parameters and statistics while preserving correlations among bins. Any peak correlation in these shuffled sets arises solely from bin covariance and noise. We set each statistic’s threshold at the 95th percentile of its shuffled peaks, with $|\rho|$ ranging from 0.14 to 0.182 across the twelve statistics.

¹² Note that we show $S(\ell)$, which has an identical response to $C_{\ell}^{\kappa\kappa}$. Hence, the 12 statistics are the auto-correlations in Fig. 6, and the weak lensing statistics in Fig. 5

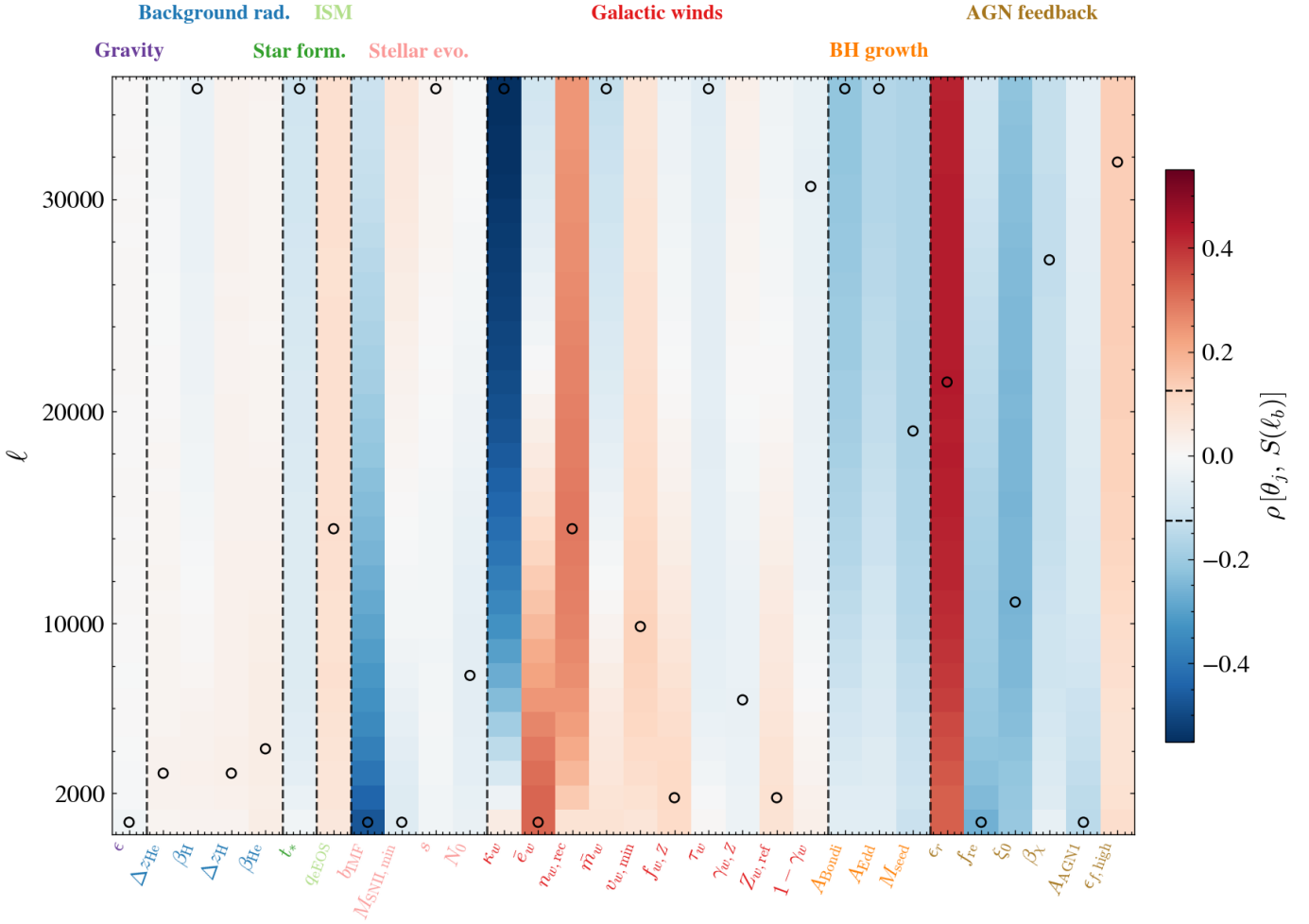


Figure 8. The peak-correlation construction of § 4.1 for the power spectrum suppression alone. Each column is one of the thirty galaxy formation parameters, each row an ℓ bin, and the color shows the signed Spearman rank correlation $\rho[\theta_j, S(\ell_b)]$ measured across the 256 Sobol nodes at $z_s = 1$. Circles mark the bin with the strongest correlation for each parameter, whose peak- $|\rho|$ values are the entries that feed the statistic-parameter grid of Fig. 9. Dashed lines on the colorbar mark the chance-correlation level for a single pre-chosen bin, $|\rho| = 2/\sqrt{256} = 0.125$ (§ 4.1). The wind sector (VarWindVelFactor, WindFreeTravelDens) drives the strongest and most scale-coherent responses.

3. *Column level, from the look-elsewhere across all statistics.* The full grid of twelve statistics and thirty parameters shares the same random assignments as in the row level, so correlations are shared not only across a statistic’s bins but between the statistics themselves. From this, we obtain the per-parameter threshold across statistics at the 95th percentile, $|\rho| = 0.200$, and the whole-grid null, $|\rho| = 0.253$, which is the level the single largest cell of a pure-noise grid would reach only 5% of the time.

A larger search always yields a luckier peak, so a parameter matters for a statistic only when it clears the level appropriate to the claim being made.

In Fig. 9 we show the peak correlations for each of the 12 statistics across all 30 parameters, with parameters

sorted by their strongest effect on any statistic. Each cell shows the signed correlation at the peak- $|\rho|$ bin for that pair. A parameter matters for a statistic only when it crosses the appropriate threshold. We find 11 parameters to the left of the dashed line clear the per-parameter look-elsewhere null (0.200) for at least one statistic. Of these, a few contain at least one cell beyond the whole-grid null (0.253), an effect stronger than 5% of the entire null space. Cells marked $\times$ fall below their statistic’s individual null and are consistent with bin covariance rather than parameter influence, and blank cells without numerals lie between their own-statistic null and the whole-grid level.

Fig. 9 compresses all statistics and parameter effects into a concise image. A handful of parameters dominate the correlations across every statistic, while the major-

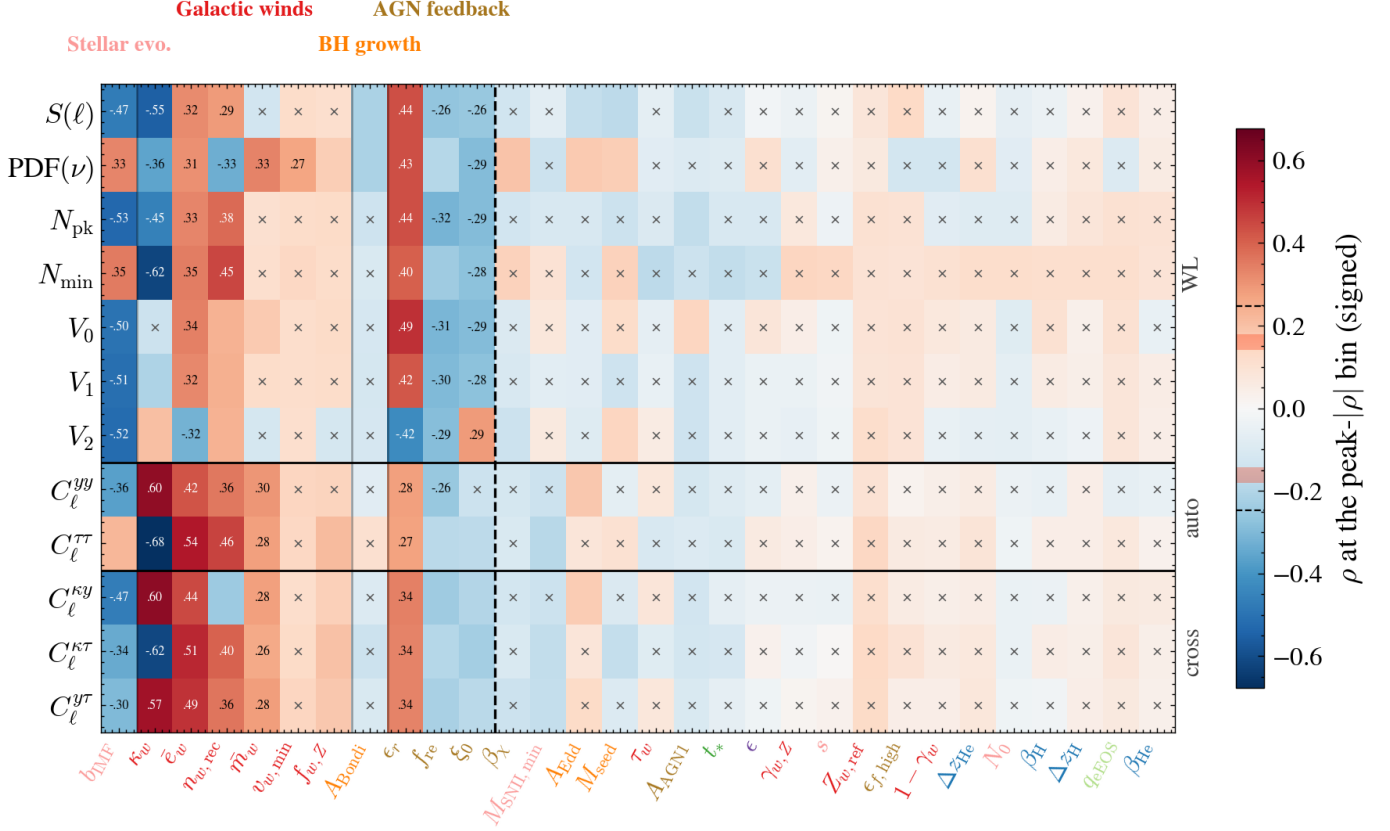


Figure 9. The signed Spearman correlation ρ evaluated at each pair’s peak- $|\rho|$ bin, for the 12 statistics (rows) against the 30 parameters (columns, sorted by peak $|\rho|$). The null levels described in the text of § 4.1 are marked on the colorbar. Printed values mark cells clearing the whole-grid null ($|\rho| \geq 0.253$), and cells marked $\times$ fall below that statistic’s null. Blank, uncrossed cells lie between the two. The vertical dashed line marks the per-parameter look-elsewhere null (0.200) such that parameters to its right never clear it for any statistic.

ity never clear their look-elsewhere null (19/30 deemed insignificant with $|\rho| \leq 0.2$).

The row structure provides more insight into the connection between the galaxy formation model and the statistics. The ℓ -domain statistics and every gas spectrum appear to be dominated by the wind parameters, while the ν -domain morphological statistics (peaks, minima, and Minkowski functionals) are sourced primarily by the IMF slope b_{IMF} and the black hole radiative efficiency ϵ_r . Different statistic families listen to different feedback channels, and in § 6, we theorize more about the physical causes.

4.2. Response families

In the previous section, we saw that 11 parameters (those to the left of the dashed line in Fig. 9) control the effects on the statistics, based on peak correlations across bins, and that the wind parameters dominate the ℓ -domain statistics while the ν -domain statistics are shaped primarily by the AGN parameters and the IMF slope. Here, instead of the amplitude of the parameter effects, we explore their scale dependence. Our goal

is to determine whether the parameters act in distinct ways or whether separate families of parameters drive the statistics to respond similarly.

We start with the ℓ -domain lensing power spectrum $C_\ell^{\kappa\kappa}$, computed across the 57 1P lightcones at $z_s = 1$. For each parameter, we compute the fractional response of the statistic,

$$R_i(\ell) = \frac{C_\ell^{\kappa\kappa}|_{\theta_{i,\max}} - C_\ell^{\kappa\kappa}|_{\theta_{i,\min}}}{C_\ell^{\kappa\kappa}|_{\theta_{\text{fid}}}}, \quad (14)$$

where, for the three parameters whose fiducial value sits at a prior bound, the response is one-sided, $(C_\ell|_{\theta_{\text{bound}}} - C_\ell|_{\theta_{\text{fid}}})/C_\ell|_{\theta_{\text{fid}}}$. We compute this for each of the 50 realizations of every 1P lightcone and take the median. Since we are interested here in the shape of each parameter’s effect rather than its amplitude, we normalize every curve by its extremal value,

$$\hat{R}_i(\ell) = \frac{R_i(\ell)}{\max_\ell [R_i(\ell)]}. \quad (15)$$

Finally, to identify connections between the responses, we compute the correlation between each normalized

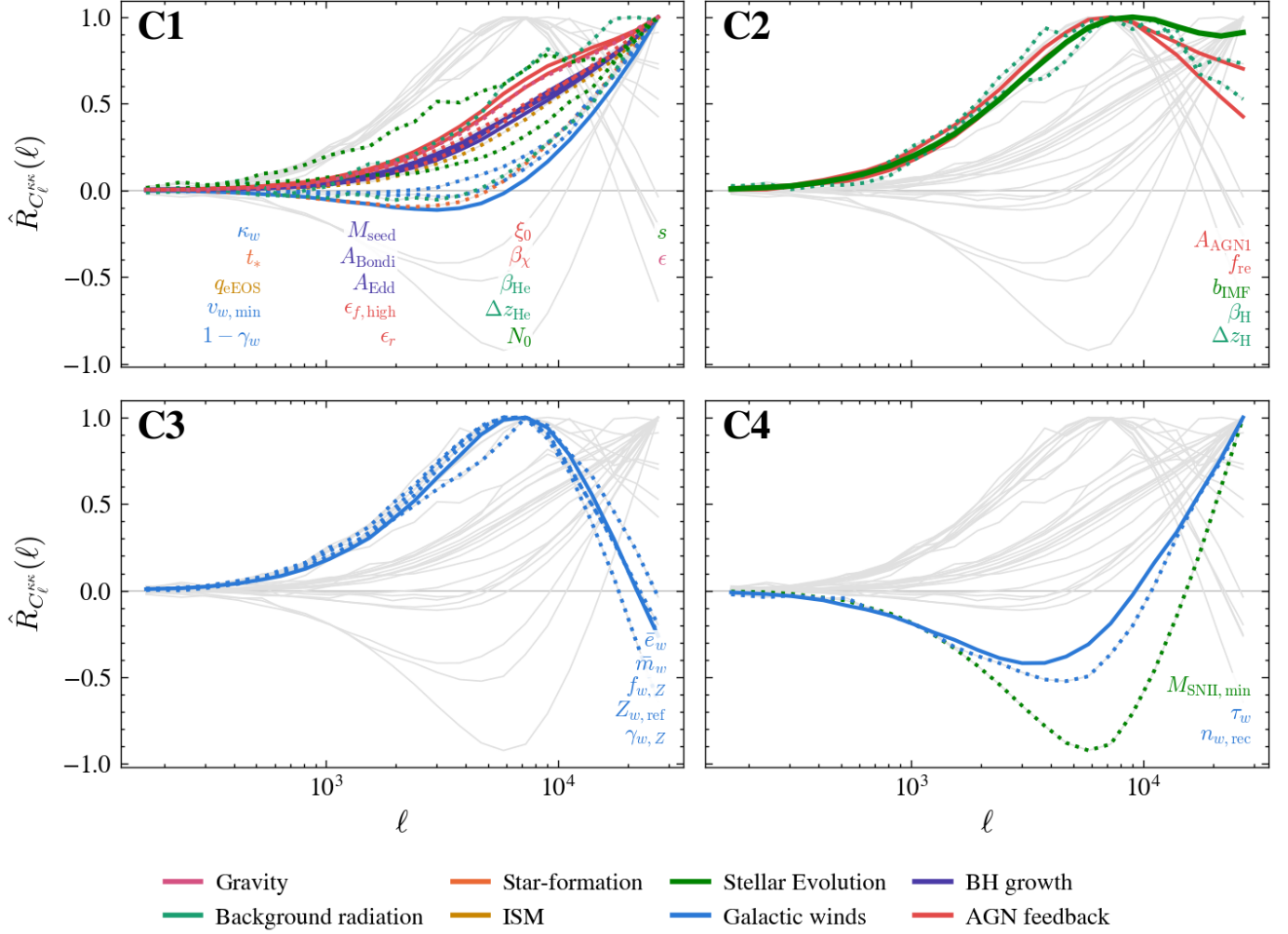


Figure 10. Response families of the convergence power spectrum. Each panel shows all thirty normalized 1P responses $\hat{R}_{C_\ell^{\kappa\kappa}}(\ell)$ (gray), with the panel’s family members colored by their astrophysical sector from Table 1 of S. Genel et al. (2026) (legend). Solid lines mark parameters passing the statistic-level noise floor of Fig. 9, while dotted lines do not.

curve and every other curve, and consider curves with $r \geq 0.85$ to belong to the same family.

In Fig. 10, we show the resulting four distinct families. Each panel shows all thirty normalized response curves, with the panel’s family members colored by their astrophysical sector, taken exactly from Table 1 of S. Genel et al. (2026): gravity, background radiation, star formation, ISM, stellar evolution, galactic winds, black hole growth, and AGN feedback. Solid lines mark parameters that pass the statistical noise floor of Fig. 9, and dotted lines do not. Because Eq. 15 normalizes each curve by its value at the scale of its largest difference, every curve reaches +1 at that scale regardless of the sign of the underlying response, so the families group parameters by the *shape* of their scale dependence and not by its sign. The signs and amplitudes are those of the $S(\ell)$ row of Fig. 9.

The four families sort the galaxy formation parameters by the radius at which they act, which we discuss in more detail in § 6. C1 controls clustering on the core scales, C2 on the halo volume, C3 on the halo radius scale, and C4 on both the core and halo radius scales. The unique responses of the families are the structure that the analytic model will exploit in S 5, and it is also why a single suppression amplitude cannot describe the parameter space. Two parameters can produce the same suppression at $\ell \simeq 6 \times 10^3$ while moving the core in opposite directions, and the peaks and Minkowski functionals, which M. E. Lee et al. (2026b) showed baryonic effects are dominated by the cores of massive halos, will tell them apart where the power spectrum cannot.

5. AN ANALYTIC MODEL OF THE STATISTICS

Fig. 10 showed that the full IllustrisTNG parameter space affects the power spectrum with roughly four

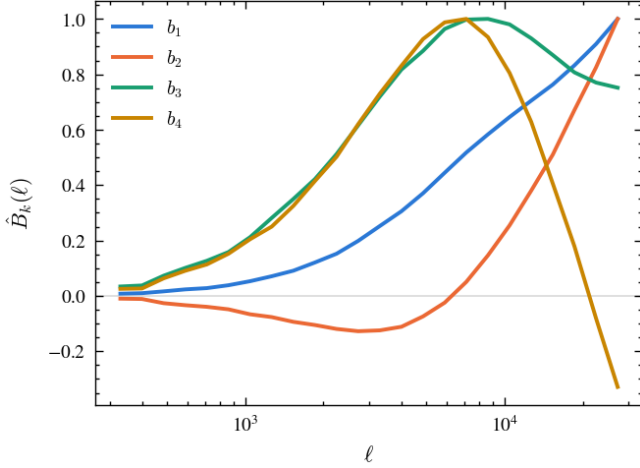


Figure 11. The basis functions $\hat{B}_k(\ell)$ for $S(\ell)$, built as the correlation-weighted mean of each family’s 1P responses shown in Fig. 10 with weights computed from Fig. 9.

unique shapes as a function of ℓ . Low-dimensional descriptions of baryonic effects have precedent, such as in previous works that investigated principal-component-based compressions of feedback’s imprint on shear data vectors (T. Eifler et al. 2015; H.-J. Huang et al. 2019), and single-amplitude models that capture much of the two-point response (A. J. Mead et al. 2021; A. Amon & G. Efstathiou 2022; C. Preston et al. 2023). Here, the basis instead emerges from the measured parameter-response families and extends to every statistic considered. With that knowledge, we can build a simple linear model by expanding these families as basis functions,

$$\tilde{C}_\ell^{\kappa\kappa} = \bar{C}_\ell^{\kappa\kappa} + \sum_k a_k \hat{B}_k(\ell), \quad (16)$$

with the $k = 1-4$ shape families as basis functions $\hat{B}_k(\ell)$, an amplitude a_k for each, and a reference power spectrum $\bar{C}_\ell^{\kappa\kappa}$ which we describe below.

We are not limited to the power spectrum here. In fact, in Appendix A we show that the parameter effects on the convergence PDF also decompose into correlated families from the 1P set, in the same way as in § 4.2. In general, then, for each statistic we can define

$$\tilde{S}(x) = \bar{S}(x) + \sum_k a_k \hat{B}_k(x), \quad (17)$$

where x denotes the statistic’s bin coordinate (ℓ or ν). Moving forward requires three choices to define in Eq. 17 – the reference statistic $\bar{S}(x)$, the combination of a family’s individual parameter responses into its basis $\hat{B}_k(x)$, and the amplitudes a_k .

5.1. Computing the basis functions

The basis functions are computed from the individual parameter responses of the 1P set as the weighted mean

over a family’s members, normalized as in Eq. 15:

$$B_k(x) = \frac{\sum_{j \in F_k} w_j R_j(x)}{\sum_{j \in F_k} w_j}, \quad \hat{B}_k(x) = \frac{B_k(x)}{\max_x [B_k(x)]}, \quad (18)$$

where F_k is the member set of the family k and the weights are the peak correlation magnitudes of § 4.1, $w_j = |\rho_j|$, evaluated for the statistic being modeled. This lets the family’s basis be shaped most by its statistically strongest members, while low-amplitude members are naturally down-weighted. One can simply compute a weighted mean for each panel of Fig. 10.

We show in Fig. 11 the resulting basis functions for the power spectrum obtained from the same lines in Fig. 10. From the curves in Fig. 11, we argue that we can generate any weak lensing power spectrum (or power spectrum suppression, or any other statistic for that matter, as long as we have computed the basis functions from the 1P set) in the entire Sobol set given the correct set of amplitude factors and reference suppression.

The reference statistic we take to be the mean over the full Sobol set of $N = 256$ runs,

$$\bar{S}(x) = \frac{1}{N} \sum_{r=1}^N S_r(x), \quad (19)$$

though the choice is somewhat arbitrary, as any reference within the suite, such as the fiducial or an arbitrary prior point, would only shift the amplitudes and be absorbed.

For the amplitudes, we consider the deviation of each Sobol run’s statistic from the reference, $\mathbf{d}_r = \mathbf{S}_r - \bar{\mathbf{S}}$, written as a vector over the statistic’s bins, and solve the ordinary least squares problem per run,

$$\mathbf{a}_r = \arg \min_{\mathbf{a}} \left\| \mathbf{d}_r - \sum_k a_k \mathbf{b}_k \right\|^2 = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{d}_r, \quad (20)$$

with $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_4]$ the basis matrix.

Solving for the amplitudes at all Sobol points gives a matrix of 256 fitted vectors $\mathbf{a}_r$, from which any Sobol-set statistic can be regenerated. In Fig. 12 we show the power spectrum suppression for a set of Sobol lightcones along with the fit model of Eq. 17. With the above linear prescription, we achieve sub-percent-level accuracy at the tested locations, a trend that holds across the entire Sobol set. This is in and of itself a non-trivial result. We computed the basis vectors from parameter variations in the 1P set, while the amplitudes came from the Sobol set. Recovering the statistics here means the statistic can be written as a linear decomposition of individual parameter variations, with an amplitude factor. We caution, though, that this does not imply a linear

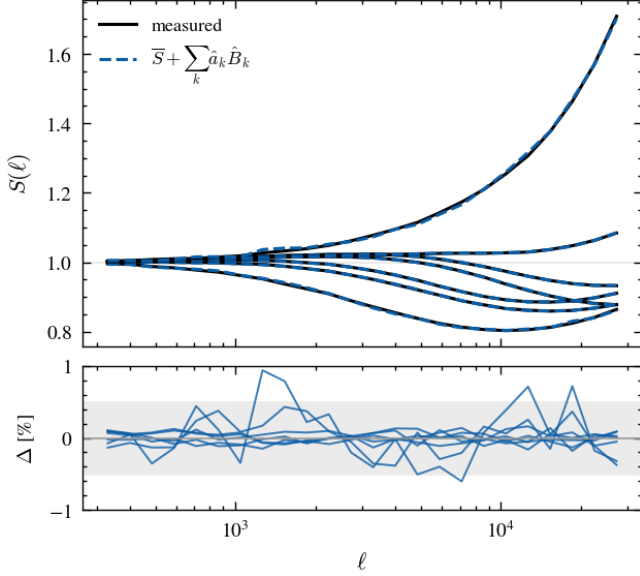


Figure 12. *Top:* the measured suppression $S(\ell)$ (solid) for a set of Sobol nodes spanning the parameter priors, against the linear model $\bar{S} + \sum_k \hat{a}_k \hat{B}_k$ (dashed), with the amplitudes fit to each run directly via Eq. 20. *Bottom:* the fractional residual, with the shaded band marking $\pm 0.5\%$.

mapping between the model parameters and the statistics themselves. If this were the case, then there would be a perfect mapping between the thirty model parameters and the amplitudes which were fit in the above ordinary least squares. We do not find this to be the case. Instead, *the resulting field statistics*, which arise from nonlinear effects induced by the parameters, can be written as a linear model.

To generalize the model across the entire prior, however, we want the amplitudes as functions of measurable quantities or simulation parameters rather than a lookup table. To do this, we need a transformation that converts simulation information into amplitude factors. A neural network is one option; however, here we opt for the simpler and more interpretable approach of a linear model

$$\alpha_k(\boldsymbol{\lambda}) = \mathbf{M}_k \cdot (\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}}), \quad (21)$$

where $\boldsymbol{\lambda}$ is a latent vector of simulation-level properties extracted from each simulation in the Sobol set (discussed in depth in § 5.2), $\bar{\boldsymbol{\lambda}}$ is the Sobol-mean latent vector of $\boldsymbol{\lambda}$, and $\mathbf{M}_k$ is a transformation matrix. where k indexes the components of the latent vector to be discussed in the following section. For $\mathbf{M}$, we stack all latent vectors $\boldsymbol{\lambda}$ and fitted amplitudes a_r from each simulation to generate $\boldsymbol{\Lambda}$ and $\mathbf{A}$, respectively, and again use least squares to solve for the linear transformation,

$$\mathbf{M} = (\boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda}^\top \mathbf{A}. \quad (22)$$

The full model is then a three-stage chain. We first extract latent quantities from each simulation (see the following sections for details). We then use these latents to find amplitude values, which we multiply into our basis vectors to obtain our statistics.

5.2. Searching for the latents

Now the question becomes, *what properties should we extract from each simulation to form the $\boldsymbol{\lambda}$ vector?* The argument above rests on the idea that there exists some linear mapping between the simulations and the resulting statistics we care about. In M. P. van Daalen et al. (2020), the authors find strong correlations between power-spectrum suppression and the baryon fraction of group-scale halos, and the trend in weak-lensing two-point statistics also suggests a potential linear mapping between halo quantities and weak-lensing statistics. Subsequent work has made this mapping quantitative, either by predicting the suppression directly from measured baryon fractions and cluster gas properties (J. Salcido et al. 2023; M. L. van Loon & M. P. van Daalen 2024; S. Grandis et al. 2024; M. P. van Daalen et al. 2025) or by learning it from halo gas observables in the CAMELS suites (A. M. Delgado et al. 2023; S. Pandey et al. 2023).

Because of this, we explore whether a linear mapping exists between some set of halo-level quantities and our weak lensing statistics. However, instead of looking only for the set of quantities that work for the power spectrum, we aim to find the full set that works for all statistics. In our suite, we have access to various halo-level observable proxies. For example, X-ray and tSZ/kSZ campaigns can provide access to halo group and cluster temperatures, pressures, Compton Y - M relations, gas masses, and baryon fractions, and photometric surveys can probe stellar masses for massive halos. We then seek the best combination of observable halo-level properties to define each simulation's $\boldsymbol{\lambda}$.

Rather than cherry-picking halo properties and potentially introducing selection bias, we let the data from the generated halos choose. From each simulation's halo catalog, we build a library of candidate summaries, listed in Table 1. The library covers the gas budget (the baryon, stellar, and gas fractions), its arrangement (the gas concentration), its thermal state (temperature, entropy, pressure, and integrated Compton parameter, each also in self-similar-scaled form), and its halo-to-halo scatter, every one measured in seven evenly log-spaced mass bins from groups to clusters, together with the cross-bin mass-trend slopes of the fractions and of T , P_e , and Y , for ninety candidates in total.

Table 1. The candidate library used for connecting amplitudes in Eq. 17 to halo quantities measured from generated halos. Twelve quantities are measured within R_{500c} for every halo in seven bins of $\log_{10} M_{500c}/M_{\odot}$ from 13.0 to 14.9, and six cross-bin slopes. The first four groups are computed as the median over halos in a mass bin, the scatter group as the 16th–84th percentile half-width across halos in the mass bin, and the slopes as linear fits of the bin medians against $\log_{10} M_{500c}$ across the seven bins. Subscripts mw denote mass weighting within the aperture, and Y is integrated within it. The last column gives the bin at which the search of § 5.2 selected the quantity (Fig. 13).

Symbol	Quantity	Definition	Picked
<i>Mass fractions</i>			
f_{bar}	baryon fraction	$(M_{\text{bar}}/M_{\text{tot}})/(\Omega_b/\Omega_m)$	[13.0]
$f_{\star}$	stellar fraction	$(M_{\star}/M_{\text{tot}})/(\Omega_b/\Omega_m)$	–
f_g	gas fraction	$(M_{\text{gas}}/M_{\text{tot}})/(\Omega_b/\Omega_m)$	–
<i>Structure</i>			
c_{gas}	gas concentration	$M_{\text{gas},500c}/M_{\text{gas},200c}$	–
<i>Thermodynamics</i>			
T	temperature	$\log_{10} T_{\text{mw}}$	[13.2]
K	entropy	$\log_{10} K_{\text{mw}}$	–
P_e	electron pressure	$\log_{10} P_{e,\text{mw}}$	[14.0]
Y	Compton parameter	$\log_{10} Y$	–
<i>Self-similar scaled</i>			
T_{ss}	scaled temperature	$\log_{10} (T_{\text{mw}}/M_{500c}^{2/3})$	–
Y_{ss}	scaled Compton parameter	$\log_{10} (Y/M_{500c}^{5/3})$	[13.4]
<i>Scatter</i>			
$\sigma(f_{\text{bar}})$	scatter of baryon fraction	half-width of f_{bar}	[13.2], [14.0]
$\sigma(Y)$	scatter of Compton parameter	half-width of $\log_{10} Y$	–
<i>Cross-bin slopes</i>			
$df_{\text{bar}}/d \log M$	slope of baryon fraction	$df_{\text{bar}}/d \log_{10} M_{500c}$	–
$df_{\star}/d \log M$	slope of stellar fraction	$df_{\star}/d \log_{10} M_{500c}$	yes
$df_g/d \log M$	slope of gas fraction	$df_g/d \log_{10} M_{500c}$	–
$d \log T/d \log M$	slope of temperature	$d \log_{10} T_{\text{mw}}/d \log_{10} M_{500c}$	–
$d \log P_e/d \log M$	slope of electron pressure	$d \log_{10} P_{e,\text{mw}}/d \log_{10} M_{500c}$	–
$d \log Y/d \log M$	slope of Compton parameter	$d \log_{10} Y/d \log_{10} M_{500c}$	–

We then try to build the candidate set of latents from this catalog. To determine whether a set of latents is actually good, we need a test set the model has not seen during training. We randomly assign the 256 Sobol simulations to five groups (five-fold cross-validation). For each group in turn, the latent map $\mathbf{M}$ of Eq. 22 is fit on the other four groups’ latent vectors, and the held-out group’s statistics are predicted from its measured latents alone. (The basis and the per-run amplitude fits of Eq. 20 involve no latents, so only $\mathbf{M}$ needs to be held out.) The score of a candidate set is then built from these held-out predictions via an R^2 per statistic. The set’s total score is the average R^2 over all statistics.

This procedure produces a noisy score because it depends on how the simulations were assigned to the five groups. We measure that noise by re-drawing the random fold partition twenty times and recomputing the score each time. We find that the resulting scatter in the score is 6×10^{-4} , which we treat as one standard deviation (σ). We use this level to set the acceptance criterion for our search, admitting a candidate latent into the group only if it raises the pooled score by more than roughly 3σ , or 2×10^{-3} . An accepted candidate then confidently increases the group score beyond what chance would expect at the 3σ level.

We start the latent search with an empty set, which grows one candidate at a time. At each step, we try

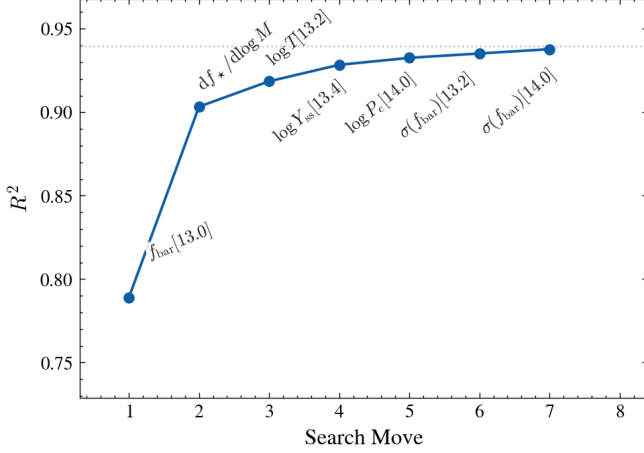


Figure 13. The path taken by the search to find halo quantities in Table 1 that map to the amplitudes of Eq. 17. The x-axis is the search step, and the y-axis is the mean value of R^2 over all WL statistics after adding a given member to the set. These members are listed next to the points along the lines. The search terminates at the dotted horizontal line, where the best remaining candidate would improve the R^2 below the 3σ acceptance threshold of 2×10^{-3} .

every remaining candidate, score it as a trial addition, and keep the best candidate if it clears the acceptance criterion. After each addition, we also allow the reverse move. If later picks make an earlier pick redundant (the group without it outscores the best group of that size found so far), we kick it out. This removal can temporarily lower the score, but it allows the search to grow from a better base. The search ends when no remaining candidate clears the threshold. We score it on its own five-fold grouping of simulations, independent of the folds used for every accuracy we report. This search score covers all the statistics we wish to emulate, namely, all the weak lensing statistics in § 3.2. So, latents accepted into the group must truly influence all associated statistics.

After this procedure, we find a set of seven accepted latent halo quantities (Fig. 13). In order, their choices tell a physical story in concordance with §§ 4.1–4.2. The first pick is the baryon fraction of low-mass groups ($13.0 \leq \log_{10} M < 13.2$). With no guidance and ninety candidates, the search rediscovers the same variable as M. P. van Daalen et al. (2020) as the single most informative halo property for the lensing statistics. This mass budget controls how much gas feedback has left R_{500c} at the group scale. This is also in agreement with where the redistribution families C3 and C4 move gas across and where the halo family C2 acts, as discussed in § 4.2 and in Fig. 10.

The second pick is the stellar fraction mass-trend slope, $df_{\star}/d\log M$. This quantity records whether feedback has preferentially stripped low-mass halos or high-

mass ones, which makes sense, as the shape of the stellar-to-halo mass function is controlled by how well stellar winds quench the lower mass range and the AGN sector the higher masses. Between these two numbers, f_{bar} and $df_{\star}/d\log M$ capture 0.90 of the variance in every weak lensing statistic.

The next three picks are thermodynamic quantities at increasing masses. The group temperature ($13.2 \leq \log_{10} M < 13.4$), the self-similar-scaled Compton Y one mass bin higher, and the cluster electron pressure ($14.0 \leq \log_{10} M < 14.3$) together add only +0.03 to the R^2 . Once the baryon fraction and the stellar partition are determined, the gas temperature remains, and these three thermodynamic quantities sample it from groups to clusters. $Y_{\text{ss}} = Y_{500c}/M_{500c}^{5/3}$ is the informative form here rather than Y itself, because dividing out the self-similar mass scaling leaves only the departure from gravitational heating, which is more reflective of the thermal energy that feedback has added.

The final two picks are the halo-to-halo scatter of the baryon fraction at fixed mass, $\sigma(f_{\text{bar}})$, at the group scale and the cluster scale, each adding an even smaller +0.003 to the R^2 score. The median gas content of halos does not capture the full picture, and apparently, how these quantities vary halo to halo at fixed mass also controls the suppression.

The final set (rightmost column of Table 1) then spans four physical axes, in order of importance: the baryon budget, the stellar partition across halo mass, the thermal state from groups to clusters, and the halo-to-halo scatter.

5.3. Halo properties, not parameters

At this point, it is fair to ask why we went through the latent search at all and why we have not just built a mapping between the 30 IllustrisTNG parameters and the amplitude vectors a_r . Nothing in the math prevents this, as Eq. 22 could take in the 30 simulation parameters for the latents (such that $\mathbf{\Lambda}$ becomes the 256×30 design matrix) and the model would become an emulator that goes from parameters as input to the statistics as an output.

We avoid this for a few reasons. First, the 30 parameters used in this work exist only inside IllustrisTNG. So a model that requires the user to define 30 subgrid parameters to generate a weak lensing statistic is a statement about one code and offers little physical insight, whereas halo quantities can be measured in any simulation and, in principle, in data. So, from the physical reasoning introduced with the family basis functions to the measured halo quantities for the amplitudes, the model presented here is directly connected to the physics driving changes

in the lensing statistics, rather than statements about the IllustrisTNG subgrid prescription (see § 6 for more details on our hypotheses of the underlying physics). The second, more pragmatic reason is that using the 30 parameters as latents will not work, and we tested this. A linear map from the parameters to the amplitudes following Eq. 21 reaches only $R^2 \simeq 0.6$ for $S(\ell)$, and replacing the least squares approach with progressively more flexible regressions (gradient-boosted trees, Gaussian processes, deep-network ensembles) moves it no further. In comparison, as shown in § 5.2 the mapping from the seven measured halo numbers to the amplitudes with least squares reaches $R^2 = 0.96$ under a purely linear map.

It is worth considering why we can achieve this result. The linear map between parameters and the amplitudes fails because the mapping between the 30 parameters and the statistics is non-linear. The attempt at building flexible regressions fails because training on 256 parameter locations in a 30-dimensional parameter space provides too few samples to learn a non-linear map. BIND is the tool that was created to capture this mapping, and it can do so because it learns it at the halo level, from roughly 3.6×10^5 projected halo maps drawn from over the set of SB35 simulations (M. E. Lee et al. 2026a), rather than from 256 lightcone summaries. The halo quantities are the output of that non-linear step, and what remains between them and the statistics is a seven-dimensional linear map, which is why ordinary least squares works at such high fidelity. Because λ is measured rather than looked up, the model carries no reference to the simulation parameters. Measure seven numbers in any simulation (or observations!) whose halos fall inside the design, and Eqs. 21–22 return every statistic. Constraining the same seven numbers from X-ray and SZ or other multiwavelength observations is a natural application (cf. A. Schneider et al. 2022; S. Grandis et al. 2024; S. Pandey et al. 2025), and we believe this is a promising direction, though real data would add hydrostatic mass bias, projection effects, and many other considerations that we save for future work.

5.4. Accuracy and validation

Table 2 collects the model’s accuracy for the twelve statistics, the seven weak-lensing statistics, and the five gas auto- and cross-spectra. The ceiling R^2 is the accuracy ceiling indicating how much of the 256-simulation variation the basis can absorb when Eq. 20 fits each run’s amplitudes freely. At 0.86–0.9997 for the weak-lensing statistics, the handful of family shapes is not the limitation for $S(\ell)$, the PDF, or the Minkowski functionals. The end-to-end CV R^2 corresponds to the models

Table 2. Family-basis model accuracy per statistic. K is the number of basis families and ceiling R^2 is the free-amplitude ceiling. The last two columns are end-to-end five-fold cross-validated accuracies, with the seven measured halo latents λ and, on the same folds, with the 30 simulation parameters θ in their place. Each statistic uses its own family basis, so K varies across statistics. The pooled rows are computed over all bins of the statistics in the block at once.

statistic	K	ceiling	end-to-end CV R^2	
		R^2	λ	θ
<i>weak lensing</i>				
$S(\ell)$	4	0.9997	0.963	0.528
PDF	4	0.985	0.939	0.498
peaks	5	0.857	0.787	0.385
minima	4	0.908	0.838	0.415
V_0	3	0.9994	0.960	0.535
V_1	2	0.986	0.946	0.457
V_2	1	0.985	0.964	0.481
pooled			0.914	0.471
<i>gas auto- and cross-spectra</i>				
C_ℓ^{yy}	2	0.991	0.862	0.494
$C_\ell^{\tau\tau}$	3	0.997	0.907	0.564
$C_\ell^{\kappa y}$	2	0.996	0.946	0.551
$C_\ell^{\kappa\tau}$	3	0.990	0.936	0.535
$C_\ell^{y\tau}$	2	0.992	0.865	0.526
pooled			0.903	0.534

where the amplitudes are computed by fitting to the latents following Eq. 22. This procedure involves measuring λ from the halos themselves, applying the $\lambda \rightarrow \mathbf{a}$ map fit with Eq. 22, and comparing the predicted curve to the held-out test set of Sobol nodes. For the weak-lensing statistics, this reaches 0.79–0.96, or 92–98% of each statistic’s ceiling, so the gap between those two columns corresponding to the information the $\lambda \rightarrow \mathbf{a}$ map fails to capture is small everywhere. The count statistics sit at the bottom of both columns, so their limitation lies in the family basis itself: some residual nonlinearity or rare-pixel sensitivity that the basis does not span. The latents are not the bottleneck there. The last column repeats the end-to-end fit with the 30 simulation parameters in place of the halo quantity latents. It reaches $R^2 = 0.39$ –0.54 for the weak-lensing statistics, roughly half of the halo quantity latent column for every statistic, and peaks and minima sit at the bottom of it as well.

Fig. 14 shows the model directly against held-out measurements for all twelve statistics. For seven Sobol simulations spanning the design space, including the strongest suppression and strongest enhancement, we refit the $\lambda \rightarrow \mathbf{a}$ map excluding the shown simulations

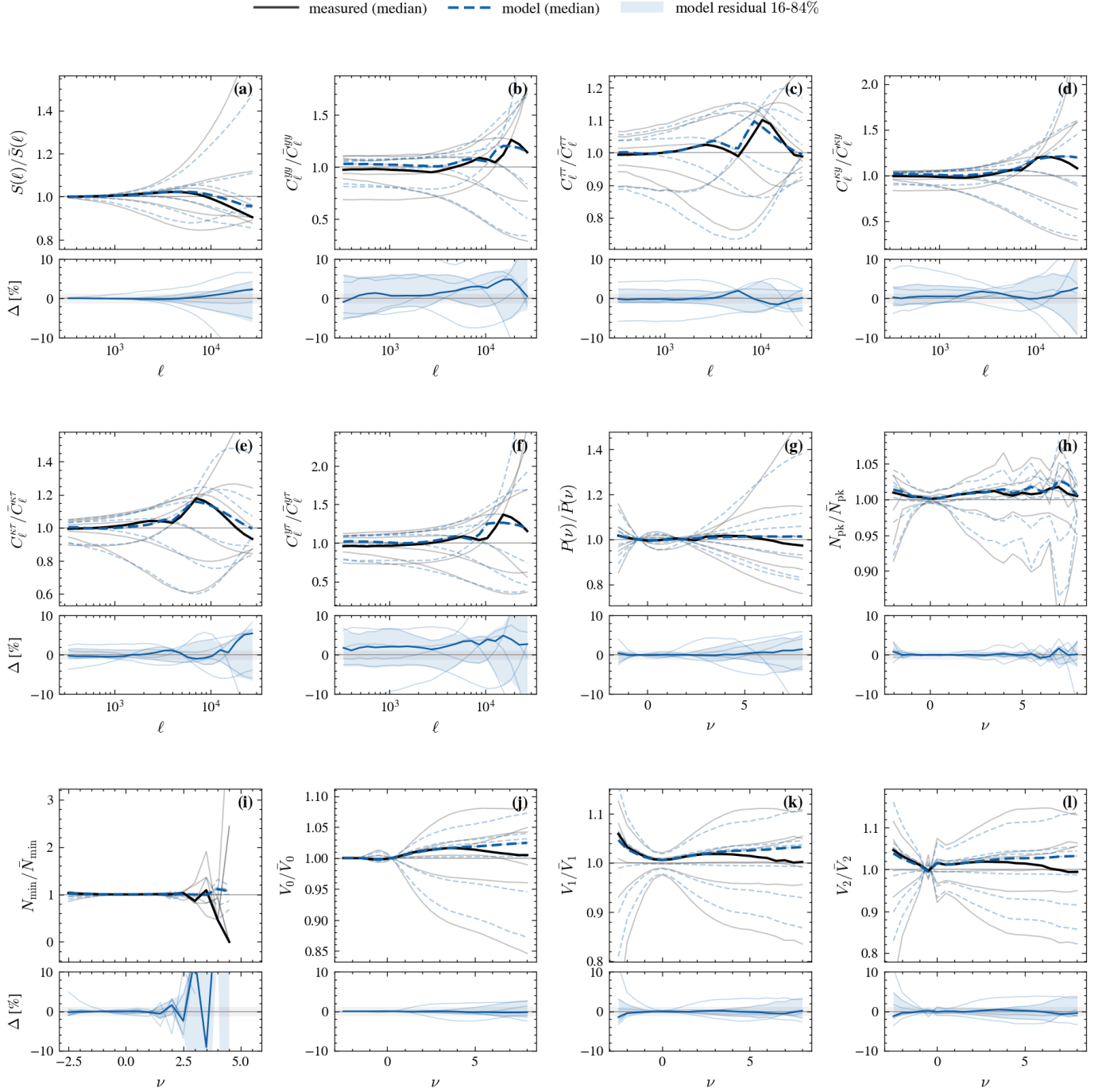


Figure 14. The linear model of Eq. 17 with halo quantities of Table 1 (rightmost column) used to compute the linear amplitudes for all twelve statistics. We compare the model predictions against a set of seven held-out simulations normalized by the mean to highlight typical deviations for a given statistic. Thin lines are the individual held-out nodes (gray, measured; blue dashed, model), heavy lines are the median over the seven nodes (black, measured; blue dashed, model), and the shaded band in each strip is the 16th–84th percentile range of the per-node model residuals. (a) $S(\ell)$; (b)–(f) the gas auto- and cross-spectra C_ℓ^{yy} , $C_\ell^{\tau\tau}$, $C_\ell^{\kappa\tau}$, C_ℓ^{xy} , and $C_\ell^{y\tau}$; (g) the convergence PDF; (h) peak counts; (i) minimum counts; (j)–(l) the Minkowski functionals V_0 , V_1 , and V_2 .

and predict every statistic using only their measured halo numbers. Because the raw statistics span several decades and a few-percent feedback response would be invisible, every panel divides both the measurement and the prediction by a common reference, the Sobol design mean, so the panels display the response each node carries and the residual strips read in the same units. The thin lines are the individual held-out nodes, the heavy lines are the median over the seven, and the band in each residual strip is the 16th–84th percentile range of the per-node residuals. The five gas spectra were not part of the search objective in § 5.2, so their panels are truly a test of the latents rather than a fit to them.

Fig. 14 shows that the model tracks the sign, the shape, and the ordering of the held-out responses in every panel. For the weak-lensing statistics, the residual band sits within ± 1 –3% wherever the reference is well measured, against node responses of up to 10–50%, and grows only at the edges: to about $\pm 5\%$ for $S(\ell)$ at $\ell \gtrsim 10^4$, to a few percent at $\nu \gtrsim 5$ for the PDF and the Minkowski functionals, and at $\nu \lesssim -2$ for V_1 and V_2 , where the reference itself is small. The minima are the one place the comparison breaks down, beyond $\nu \simeq 2.5$, where the reference falls to a handful of counts per map and the ratio stops being meaningful. The growth of the $S(\ell)$ residual at $\ell \gtrsim 10^4$ is where b_3 and b_4 finally separate (Fig. 11); over the rest of the range the two are nearly degenerate, so the combination of amplitudes that distinguishes them is set by a handful of small-scale bins and is the hardest thing in the model to predict.

The gas spectra are a harder test. Their responses across the design are of order unity rather than tens of percent (§ 4), and the residual band is correspondingly two to three times wider, ± 2 –5% at $\ell \lesssim 3 \times 10^3$ and ± 5 –10% at $\ell \gtrsim 10^4$, widest for the three y spectra (C_ℓ^{yy} , $C_\ell^{\kappa y}$, $C_\ell^{y\tau}$) that the rarest, most massive halos dominate. The optical-depth spectra carry the most scale structure in the set, with a rise to the one-halo peak near $\ell \simeq 10^4$ and the return to the mean beyond it, and the model reproduces that shape in $C_\ell^{\tau\tau}$ and $C_\ell^{\kappa\tau}$ with the peak height recovered to within 2–5%. Across all twelve panels, the largest residuals occur at the most extreme nodes, and in nearly every case the model pulls the extreme node toward the design mean.

The model mapping is linear, and the design is the SB35 astrophysical prior (the CAMELS-IllustrisTNG prior of S. Genel et al. 2026, for which we vary the 30 galaxy formation parameters), so λ values far outside the design ranges extrapolate unphysically. The cosmology is fixed to that of TNG300, and the basis inherits the noise of the 1P suite from which it is built. We ship the model as a single table of curves and matrices ($\bar{S}$,

$\hat{B}_k$, $\mathbf{M}$, $\bar{\lambda}$, and the per-bin predictive scatter σ_{pred}) with a numpy-only loader.¹³

6. THE STORY OF THE STATISTICS AS TOLD BY THE PARAMETERS

The preceding sections presented three distinct descriptions of the lensing and CMB maps, along with their dependence on the IllustrisTNG model parameters, by posing three questions.

1. § 4.1 asked which parameters move which statistics. The peak-correlation grid in Fig. 9 answered that the ℓ -domain statistics are more sensitive to the wind parameters, while the ν -domain morphological statistics are more sensitive to the IMF slope and the AGN parameters.
2. § 4.2 asked what shape each statistic’s response to the parameters takes, and found four families of shapes describing these responses in Fig. 10.
3. § 5.2 asked which halo properties carry the amplitudes of the response, and found that seven halo quantities, chosen by a search that knew nothing about the parameters, suffice.

In this section, we attempt to tell a cohesive story that ties these three descriptions together and propose a physical narrative to connect them.

We want to be upfront about these next sections. The physics of galaxy formation, and the IllustrisTNG model used in this work, is complex, and its effects on the summary statistics are nonlinear and coupled (as shown by the R^2 score in the rightmost column of Table 2). A complete account of the physical effects and statistical intricacies lies beyond the scope of this work and will require targeted investigation. But the three constructions above provide empirical constraints on which physical processes could drive the effects. The following subsections present a physical narrative we have found that satisfies these empirical constraints.

6.1. A gas budget and a thermal state

We start by examining the gas fields, τ and Compton- y , because they respond differently to gas density. τ maps are formed simply by the electron density traced along the line of sight (Eq. 6), but pixels in the Compton- y maps are proportional to the temperature-weighted density (the integrated pressure, Eq. 7). Therefore, a parameter that changes only the *amount* of gas would shift the two spectra in the same direction by moving the density up or down.

¹³ Release DOI to be inserted at acceptance; see § 9.

In Fig. 9, many parameters do exactly this. For example, the wind energy $\bar{e}_w$ (WindEnergyIn1e51erg) moves both spectra in the same direction and with roughly the same magnitude, $\rho = +0.54$ and $\rho = +0.42$. The wind velocity factor κ_w (VarWindVelFactor) does not. Instead, it produces the strongest peak correlation in the grid, $\rho = -0.68$ for $C_\ell^{\tau\tau}$, and a nearly equal and opposite peak correlation for C_ℓ^{yy} , $\rho = +0.60$. The only thermodynamic quantity that can separate a density-weighted field from a pressure-weighted one is the temperature, so we take this to mean that $\bar{e}_w$ sets how much gas the halo holds while κ_w sets how hot that gas is.

This is what we would expect from the IllustrisTNG wind model, in which the mass loading factor scales as (A. Pillepich et al. 2018)

$$\eta_w \propto \bar{e}_w / \kappa_w^2. \quad (23)$$

So the two wind parameters change the mass loading in different ways. Raising $\bar{e}_w$ loads more gas into winds at the same speed, lifting gas out of the galaxy but not out of the halo. As a result, the reservoir feeding the central black hole shrinks, so the black hole never grows large enough to drive feedback that escapes the halo, and the halo keeps its gas. Both τ and y grow, which is reflected in the pair of positive correlation cells in Fig. 9.

Raising κ_w instead lowers the mass loading and makes the winds faster and lighter. What they launch is carried out of the halo core to large radii, and when the wind recouples to the gas, it thermalizes. The density-weighted τ loses power where the gas used to be, while the pressure-weighted y , which is also sensitive to the thermal state of the gas, gains power wherever the gas ends up. The gas-spectrum families in Appendix A show the same thing as a function of ℓ : the Compton- y response to κ_w is effectively scale-independent, because the gas moved out of the core has been balanced by heating (Fig. 16).

The recoupling density $n_{w,\text{rec}}$ (WindFreeTravelDens) offers a similar story. A wind particle recouples with the medium once the gas density falls below the threshold set by $n_{w,\text{rec}}$, so raising it makes the wind recouple sooner and deposit its mass closer to where it was launched. Every peak correlation between the statistics and $n_{w,\text{rec}}$ in Fig. 9 is positive, indicating that gas is retained as the recoupling density is raised.

In this interpretation, the halo properties needed to predict all the statistics should include quantities that reflect the gas budget and thermal state of halos. Indeed, this is exactly what the search in § 5.2 chooses. The first pick was the baryon fraction of low-mass groups, the gas budget, which alone accounts for $R^2 = 0.79$ of the pooled variance of the statistics and

rediscovers the quantity M. P. van Daalen et al. (2020) identified for the power spectrum. The search’s third, fourth, and fifth picks were thermal-state quantities: the group temperature, the self-similar-scaled Compton Y , and the cluster electron pressure.

6.2. Where the gas ends up

The gas budget and its thermal state tell us how much gas there is and how hot it is, but the response families in Fig. 10 show where this gas has been redistributed. Viewed this way, the wind-sector parameters split into three families.

The wind energy $\bar{e}_w$ and every parameter that modulates it (the specific momentum and the three energy-reduction parameters) form C3 by themselves. C3 peaks at the halo scale, crosses zero near $\ell \simeq 1.5 \times 10^4$, and is negative at the smallest scales. With $\rho = +0.32$ for $\bar{e}_w$ on $S(\ell)$ in Fig. 9, we read this to mean that gas is lifted out of the galaxy and deposited in the circumgalactic medium, and the zero crossing in the C3 panel marks the radius that separates where the gas is taken from and where it is put.

The recoupling density $n_{w,\text{rec}}$ sits in C4 with the thermal wind fraction τ_w and the SNII minimum mass $M_{\text{SNII},\text{min}}$. C4 is nearly the inverse of C3: its peak is at the smallest scales rather than the halo scale, and it crosses zero near $\ell \simeq 10^4$. Winds that recouple at higher density deposit their mass closer to where they were launched, so power moves from the halo scale back toward the core. A larger thermal fraction makes the wind slower and more extended, so it recouples sooner; a higher SNII minimum mass means fewer supernovae per unit stellar mass and less wind energy, the inverse of raising $\bar{e}_w$. So C3 acts mainly on the halo scale with a smaller, opposite effect in the core, while C4 acts mainly on the core with a smaller, opposite effect at the halo scale.

The wind velocity factor belongs to C1, whose response vanishes at $\ell \lesssim 10^3$, grows through the halo scales, and continues to grow up to the Nyquist limit. So the three wind mechanisms are: how much gas is moved (C3), how far it gets (C4), and how fast it leaves the core (C1). Each has its own characteristic shape, and the family that sets the thermal state shares its shape with the sector that heats the same gas from the inside. That is the family-space version of what the (y, τ) pair showed in the grid. The winds are not the full story, however; next, we turn to the AGN parameters and how they fit into this interpretation.

6.3. The engine in the core

We find that the AGN parameters act on the core in two distinct ways. The first is the quasar-mode thresh-

old ξ_0 (QuasarThreshold), which moves the peaks and the Minkowski functionals in Fig. 9 at $\rho \simeq -0.3$ and $S(\ell)$ at -0.26 and -0.28 , while sitting at or below the null for every gas spectrum. In IllustrisTNG, this parameter sets the Eddington ratio below which a black hole switches from the thermal quasar mode to the kinetic mode, $\chi_{\text{thresh}} = \min[\xi_0 (M_{\text{BH}}/10^8 M_\odot)^{\beta_x}, 0.1]$ (R. Weinberger et al. 2017), and it is the kinetic mode that quenches star formation and ejects gas from halo centers (E. Zinger et al. 2020). Raising the threshold shifts the transition to kinetic-mode feedback to lower black hole masses, allowing more efficient feedback from a halo core at any given time. The expectation is that the stronger feedback suppresses power, peaks, and the other statistics, which is what the negative peak correlations in Fig. 9 show.

The second lever is black hole growth, for example through the radiative efficiency ϵ_r , which lowers the Eddington limit and slows black hole mass growth. The upper bound of ϵ_r corresponds to smaller black holes in massive halos and weaker integrated AGN feedback. The ξ_0 threshold and the ϵ_r efficiency therefore have nearly opposite effects on the feedback, and they correlate in opposite directions with the statistics in Fig. 9.

The families do not show the sign of the effects, because in defining them in § 4.2, each response is normalized by its extremal value so that every curve reaches +1 at its peak regardless of sign (Eq. 15), and the opposite effects of ξ_0 and ϵ_r have the same shape in Fig. 10. Both, together with β_x , reside in C1 alongside κ_w , with the black hole growth and SNIa parameters as low-amplitude companions. So all four significant members of the C1 core family affect clustering only on the smallest scales, not outside the halo regions.

The two radio-mode AGN parameters, A_{AGN1} and f_{re} , land in C2, the halo family, to which the clustering responds most strongly at the halo scale and at a somewhat lower level toward smaller scales.

Parameters that determine whether and how strongly the kinetic mode fires act on the core in C1, while the mode itself reshapes the gas out to larger scales in C2. We think this is also why the thermal latents of § 5.2 climb the mass range, from the group temperature at $10^{13.2} M_\odot$ to the cluster electron pressure at $10^{14} M_\odot$: winds set the thermal state of the groups, and black holes set it for the clusters. The same mass split is, we believe, what separates the rows of Fig. 9. The ℓ -domain statistics integrate over every halo above $10^{13} M_\odot h^{-1}$ and are dominated by the numerous groups just above that threshold, where the winds set the budget. The high- ν peaks and the deepest minima are produced by rare massive halos, whose cores, as M. E. Lee et al. 2026b

showed, source the baryonic response in the morphological statistics, and those cores are where the black hole engine acts. Different statistics listen to different sectors because different sectors act on different halos. The one parameter that shares the radio-mode shape but is not an AGN parameter is the IMF slope, which we discuss next.

6.4. The stellar mass

The IllustrisTNG model citep Lee-2024 has previously shown that the IMF slope behaves like an AGN parameter (M. E. Lee et al. 2024). The IMF slope controls the number of massive stars, such that a shallower IMF contains more massive stars and produces more supernovae, and one might expect it to act as another wind parameter, excising the core or redistributing matter out to the halo scale. While this is partly true, the IMF slope is not just another wind-energy dial, because it also sets the metallicity of the halo gas, which, through enrichment and cooling, drives the growth of the central black holes (A. Pillepich et al. 2018; M. E. Lee et al. 2024).

The grid hints at the same thing with the peak correlations of b_{IMF} with the WL peak and minimum counts carrying opposite signs, -0.53 and $+0.35$, both above the whole-grid null, whereas the wind parameters move the two counts together, though the minima are the noisiest statistic in the suite, and we do not rest the argument on this.

The C2 family in Fig. 10 agrees that the IMF slope acts like the radio-mode AGN parameters, on the halo scale and inward, while staying responsive at the smallest scales like the wind parameters. Ahead of every thermodynamic quantity, the latent search of § 5.2 chose the slope of the stellar fraction across halo mass, which records whether feedback has quenched the low-mass halos or the high-mass ones. A parameter that moves the stellar mass of cores through a channel that switches on with black hole mass would show up first in exactly this quantity.

6.5. What this interpretation explains

Taken together, the four families in Fig. 10 sort the galaxy formation parameters by the radius at which they act: the core (C1), the halo (C2), and the two directions of transfer between them, outward from the core to the halo scale (C3) and back from the halo scale toward the core (C4).

The wind sector supplies one family that determines how much gas moves and one that sets how far it travels before it stops. The AGN sector sets when its engine switches on and how large it grows, and the IMF slope

reaches the same core through stellar mass. This is the same structure the analytic model of § 5 exploits, and it is why the seven halo-quantity latents in Table 2 present themselves as four physical axes, a baryon budget, a stellar partition across mass, a thermal state, and a scatter, rather than as an arbitrary list.

We emphasize that this reading is a hypothesis about IllustrisTNG, assembled from rank correlations, response shapes, and a halo-variable search, and that we have not traced any of its mechanisms in the simulations themselves; doing so is a natural next step. This interpretation explains why a single suppression amplitude cannot describe the parameter space. Two parameters can produce the same suppression while moving the core in opposite directions, and the peaks and Minkowski functionals, whose baryonic response M. E. Lee et al. (2026b) showed is dominated by the cores of massive halos, tell them apart where the power spectrum cannot.

7. CAVEATS

Throughout this manuscript, we make several approximations, which we state explicitly here. We group these approximations and caveats by whether they concern what the maps physically contain, what the training covers, or what the validation actually establishes.

7.1. What the maps contain

The most important approximation is that we replace halos via our generated model rather than an entire volume. So only regions centered on halos with $M_{200c} \geq 10^{13} M_{\odot} h^{-1}$ carry generated baryonic fields, and below that threshold, and in the space between halos, the matter field is dark-matter-only from TNG300-Dark. The construction also omits the back-reaction of baryons on the dark matter outside of the painted regions, which can alter the dark matter power spectrum by up to tens of percent at $k \sim 10 h \text{ Mpc}^{-1}$ in extreme feedback models (M. Gebhardt et al. 2026). As noted in § 3.2, M. E. Lee et al. (2026b) found that replacing halos at and above our threshold of $10^{13} M_{\odot} h^{-1}$ accounts for roughly 60–65% of the power-spectrum response and for virtually all of the response in the peak counts and Minkowski functionals. By construction, our maps lack the missing ~ 35 –40% of the two-point response carried by halos below the threshold and by the diffuse field. This is why we compare every claim in this paper against the hydro-pasted control rather than a hydrodynamical simulation. The control shares the same missing physics, so the comparison isolates the generative model’s error rather than the architecture’s. It is also why the τ and y maps should be used for responses and aperture measurements centered on halos, rather than for absolute

amplitudes on large scales, where the diffuse component we do not model contributes a significant fraction of the signal.

Two smaller geometric approximations follow from the same architecture. Halos whose centers fall near a slab boundary are, in essence, cut off, so part of their gas is assigned to the neighboring shell. This affects a small fraction of halos and enters as noise, since the assignment is uncorrelated with the parameters. Second, BIND is trained on projections of depth $50 \text{ Mpc } h^{-1}$ while the TNG300-Dark slabs are $51.25 \text{ Mpc } h^{-1}$, so the conditioning fields carry about $1 \text{ Mpc } h^{-1}$ more line-of-sight material than the model saw in training. We show in § 3.1 that this does not measurably degrade the generated halos, but it is a mismatch and implies an out-of-distribution usage in this paper.

Finally, we model no velocities. The maps are of the optical depth itself, not of the kinematic Sunyaev-Zel’dovich signal, and converting between the two requires a peculiar velocity field that our pipeline does not supply. Users who need either must supply the velocity assumptions when reducing their data.

7.2. What the training covers

This paper only uses the IllustrisTNG galaxy formation model and its parameter space. The four response families, the seven latents, and the linear map between them are properties of that TNG model. A different subgrid family, such as SIMBA (R. Davé et al. 2019) or ASTRID, can potentially have a completely different response family and different latents (A. M. Delgado et al. 2023; Y. Yang et al. 2026). We are not claiming a universal law, but instead a model for IllustrisTNG. Expanding BIND to other subgrid parameterizations is an interesting open question, because a compression that survives changes in subgrid prescriptions would be a statement about feedback, whereas in this work we can only make a statement about IllustrisTNG.

The cosmology is also fixed at cosmological parameter constraints from Planck Collaboration et al. 2016. BIND itself is conditioned on cosmological parameters through the SB35 training suite, but every lightcone here is built on TNG300-Dark, which was run at a single cosmology. This is what makes the responses in § 4 independent of cosmology and unambiguous, but we cannot say anything about degeneracies between feedback and cosmological parameters. Those degeneracies are the reason feedback is a systematic in the first place, and quantifying them requires painting across a suite of cosmologies rather than one.

The analytic model inherits a related limit. The map from latents to amplitudes is linear and is fit to the

SB35 design, so it is valid only within the prior bounds and nowhere else. A simulation whose measured λ falls outside the design ranges will extrapolate out of distribution. However, since the Sobol sequence used in this work extends over the full reasonable prior space of IllustrisTNG’s model parameter space, we believe extrapolation is already unphysical.

7.3. What the validation establishes

Our fiducial validation in § 3 compares generated halos with hydro-pasted halos and ray-traced maps, measuring BIND’s error rather than the halo-replacement architecture itself, which we characterize separately in M. E. Lee et al. 2026b. A full-hydro anchor, in which the same rays are traced through the full TNG300 hydrodynamical simulation, would fold in the halo-replacement error as well. M. E. Lee et al. 2026b characterized that error, so we omit the full-hydro comparison here.

Finally, the validation and the detectability comparison in § 3.2 are noiseless. We apply no shape noise, survey mask, photometric redshift uncertainty, or intrinsic alignment contamination, and we measure the covariance across realizations that are rotations and translations of a single lightcone rather than independent volumes. The precision envelopes of § 4 are then statements about statistical power, not forecasts from a systematics-marginalized analysis.

8. FUTURE OUTLOOK

The limits above sort naturally into three groups, distinguished by whether they require new analysis, a new model, or new science.

The first group needs neither a new model nor new science. We can add survey realism to the maps we have already released by drawing shape-noise realizations onto the convergence maps, applying a mask, and repeating the detectability comparison in § 4 under tomography. This would convert the statistical-power statement into a forecast, which is a plan for upcoming work.

The second group requires retraining BIND. The most important next step for BIND is to allow lower masses, which M. E. Lee et al. (2026b) show are relevant for power spectrum measurements. Lowering the mass threshold toward $10^{12} M_{\odot} h^{-1}$ would recover most of the suppression that our maps currently miss, but requires thought on how to apply this to N -body simulations, since the halo mass function requires a significantly larger population of halos to be generated per snapshot. We envision BIND applied to lower-mass halos as the next iteration. Training the model on alternative subgrid parameter sets individually, such as those from SIMBA or ASTRID, would test whether the

family decomposition and the seven latents are properties of feedback in general or are just specific to IllustrisTNG. This requires a training set of halos from CAMELS at the $50 \text{ Mpc } h^{-1}$ resolution for these subgrid prescriptions. Once available, we will use them to train BIND in this way. Painting onto dark-matter-only simulations run at varied cosmologies would open the feedback-cosmology degeneracies that this work holds fixed and allow for use in cosmological parameter analyses. Finally, moving from projected patches to three-dimensional volumes would remove the projection approximation entirely, allowing for 3D halo shapes, capturing the anisotropic ejection of matter around halos, and enabling a higher-fidelity model, but at a cost of memory and generation time.

The third group is the science the maps were built for. The compression result of § 5.2 identifies seven halo properties that carry feedback’s imprint on the lensing sky, and each of them is a measurable quantity from X-ray, SZ, and spectroscopic campaigns. The natural next step is to try to constrain those halo measurables from the existing multiwavelength observations of gas, and propagate them through the analytic model into a prediction for weak lensing statistics (for steps in this direction, see A. Schneider et al. 2022; S. Pandey et al. 2023; A. Ganguly et al. 2026). Doing this correctly requires many observational considerations, such as hydrostatic mass bias, projection effects, and the possibility that the real universe’s subgrid calibration does not match TNG’s. For these reasons, we believe this is out of the scope of the current work. However, the machinery is in place, and the maps, statistics, and model tables have been released, so it can be attempted.

9. CONCLUSIONS

We set out to ask what a full galaxy formation parameter space does to the Stage-IV weak lensing and SZ statistics, and if that response could be compressed into measurable quantities. Using BIND, we generated 314 lightcones spanning the 30-dimensional IllustrisTNG astrophysical prior, comprising the 256-node Sobol sequence, the 57-run 1P atlas, and the fiducial, alongside the DMO and hydro-pasted references. Every lightcone carries κ , Compton- y , and τ maps at five source redshifts, built from the same halos with the same flow-matching seeds, so that any difference between two lightcones is driven by the parameters alone. To our knowledge, this is the first set of lightcones in which the lensing and the gas observables of every halo respond consistently to a full galaxy formation parameter space, and it is what lets us study feedback effects in a con-

trolled way, and with respect to observational quantities such as lensing power spectra and peak counts.

At the fiducial parameters, the BINDed maps reproduce the hydro-pasted control set to within the statistical precision of an LSST-Y10-like survey for every weak lensing statistic we measure. The gas spectra are more biased. The optical-depth spectra agree with the control to within about 5%, while the spectra that involve Compton- y run from 5–20% low at $\ell \lesssim 10^3$ to about 25% high at the smallest scales. The weak lensing maps are therefore usable as they stand, and the gas maps are usable for responses rather than for absolute amplitudes.

Across the prior, Stage-IV surveys can detect feedback, but only on small scales. Over $\ell \simeq 2.2 \times 10^3 - 1.8 \times 10^4$ the spread in $S(\ell)$ across the Sobol nodes exceeds the noiseless LSST-Y10 statistical precision by more than 10σ , peaking at 22σ near $\ell \simeq 5 \times 10^3$, and 84% of nodes depart from the fiducial by more than 5σ (73% for *Euclid*). We also find that the statistics do not all respond in the same ways to the same galaxy formation model parameters. Eleven of the thirty parameters clear their look-elsewhere null for at least one statistic, and the row structure separates, with the ℓ -domain statistics and every gas spectrum dominated by the galactic wind parameters, while the ν -domain morphological statistics respond instead to the IMF slope and the AGN parameters. Within the wind sector, the wind velocity sets the thermal state of the gas and moves τ and y in opposite directions, while the wind energy sets the gas budget and moves them together. The thirty 1P response curves of the power spectrum collapse into four families, sorted by the radius at which they act: a core family, a halo family, and two redistribution families that move gas outward or inward across the halo scale.

The central result of this work is that the response is compressible. Each statistic can be modeled as a linear combination of its response families, and a data-driven search over ninety candidate halo summaries then selects seven measured halo properties as the latent variables that set the amplitudes. They sort into four physical categories in order of importance: the baryon fraction of low-mass groups, its partition across halo mass, the thermal state of the gas from groups to clusters, and its halo-to-halo scatter. The search was unguided, and its first pick, the baryon fraction of low-mass groups, essentially rediscovers the [M. P. van Daalen et al. \(2020\)](#) conclusion, which also anchors observationally calibrated

suppression models ([J. Salcido et al. 2023](#); [M. L. van Loon & M. P. van Daalen 2024](#); [M. P. van Daalen et al. 2025](#)). Under a purely linear map, these seven numbers predict every held-out statistic – the seven weak lensing statistics and the five gas spectra alike – at $\text{CV } R^2 = 0.79\text{--}0.96$ (Table 2). The same linear map from the thirty IllustrisTNG parameters reaches only 0.39–0.56, and more flexible regressions do no better, because 256 nodes in thirty dimensions cannot resolve a non-linear map. BIND is what absorbs that non-linearity, and once it has, what remains is linear in quantities that X-ray and SZ campaigns already measure.

We publicly release the lightcones, the measured statistics, and the model tables, the latter as a single set of curves and matrices ($\bar{S}$, $\hat{B}_k$, $\mathbf{M}$, $\bar{\lambda}$, and the per-bin predictive scatter σ_{pred}) with a numpy-only loader. We intend for them to be used to test gas-calibrated baryon corrections against a field-level reference, to forecast joint lensing and SZ analyses, and to ask, for any weak lensing statistic, which halo property it is really measuring.

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. 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 and BH.

Software: astropy ([Astropy Collaboration et al. 2013, 2018, 2022](#)), LensTools ([A. Petri 2016](#))

APPENDIX

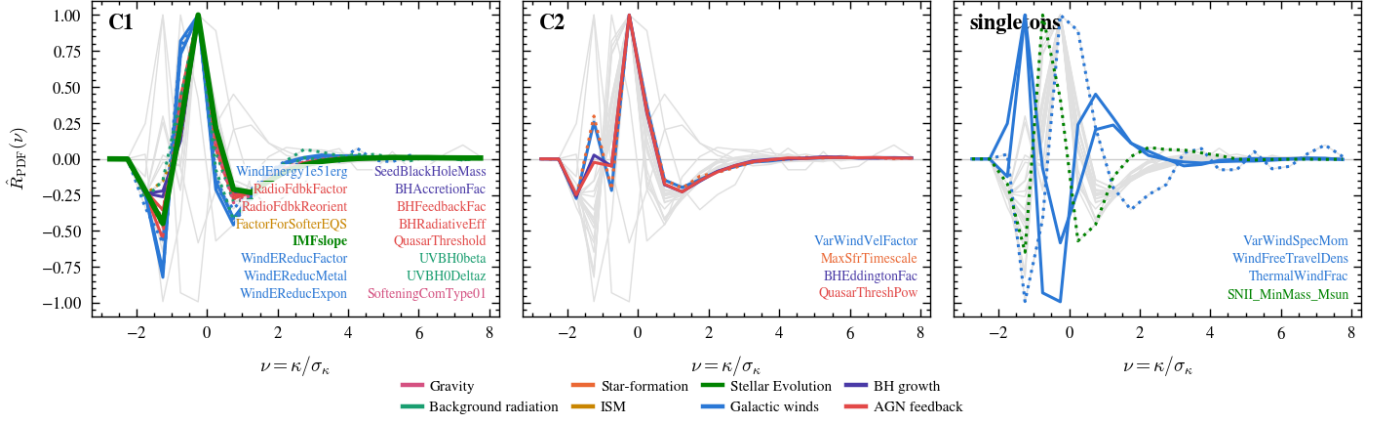


Figure 15. Response families of the convergence PDF, constructed exactly as in § 4.2, where the thirty normalized 1P responses $\hat{R}_{\text{PDF}}(\nu)$ (gray in every panel), clustered at $r \geq 0.85$, with each panel’s family members colored by their astrophysical sector and solid lines marking parameters that pass the statistic-level noise floor of Fig. 9. The PDF collapses into two families and four singletons. C1 gathers sixteen parameters spanning AGN feedback, black hole growth, the IMF slope, and the wind energy-reduction sector into a single shape, which provides a sharpening of the peak just below $\nu = 0$ fed by a deficit near $\nu \simeq -1.3$, and C2, containing VarWindVelFactor among four members, carries a variant with a second node in the underdense wing. The singletons are the momentum-driven wind parameters and SNII_MinMass_Msun, neither of which clears its noise floor with a coherent shape. That thirty-parameter responses reduce to essentially two shapes is what makes the $K = 4$ PDF basis of Table 2 sufficient.

A. PER-STATISTIC RESPONSE FAMILIES AND MODEL OVERLAYS

The main text develops the response-family construction for $C_\ell^{\kappa\kappa}$ (§ 4.2) and validates model in Fig. 14. This appendix shows alternative per-statistic counterparts for completeness. Showing all for each statistic would be overwhelming for the text, but are available upon request.

Here we show the family decomposition of the convergence PDF and the response families of the gas spectra, which provide the basis for the model of § 5 and the held-out model overlays for the six ℓ -spectra.

Fig. 15 shows the families of the convergence PDF. The procedure is identical to that of Fig. 10 where the responses of Eq. 14 on the 1P set, normalized through Eq. 15 and clustered at $r \geq 0.85$. The only difference here is the change in statistics. The membership rearranges relative to the power spectrum in the way Fig. 9 anticipates. The AGN, black hole growth, and IMF-slope parameters, which dominate the ν -domain rows of the correlation grid, merge into one large family rather than splitting across several, and the wind sector contributes the second family and the singletons. The shapes say that essentially all of the effective parameters act on the PDF the same way, by narrowing the distribution about its peak, and differ mainly in where in the underdense wing the compensating probability is drawn from.

Fig. 16 shows the response families of the two gas auto-spectra that enter the model of § 5. Unlike the lensing statistics, these are built in the log, since the responses are multiplicative and of order unity. In all other ways, the construction is that of § 4.2. The family structure again follows the correlation grid, with the optical-depth spectrum cleanly separating into an AGN family and two wind families, while the Compton- y spectrum promotes the wind-velocity pair that dominates its correlations into a family of its own.

REFERENCES

- Amodeo, S., Battaglia, N., Schaan, E., et al. 2021, PhRvD, 103, 063514, doi: [10.1103/PhysRevD.103.063514](https://doi.org/10.1103/PhysRevD.103.063514)
- Amon, A., & Efstathiou, G. 2022, MNRAS, 516, 5355, doi: [10.1093/mnras/stac2429](https://doi.org/10.1093/mnras/stac2429)
- 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)
- Angulo, R. E., & White, S. D. M. 2010, MNRAS, 405, 143, doi: [10.1111/j.1365-2966.2010.16459.x](https://doi.org/10.1111/j.1365-2966.2010.16459.x)
- Aricò, G., Angulo, R. E., Contreras, S., et al. 2021, MNRAS, 506, 4070, doi: [10.1093/mnras/stab1911](https://doi.org/10.1093/mnras/stab1911)
- 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)

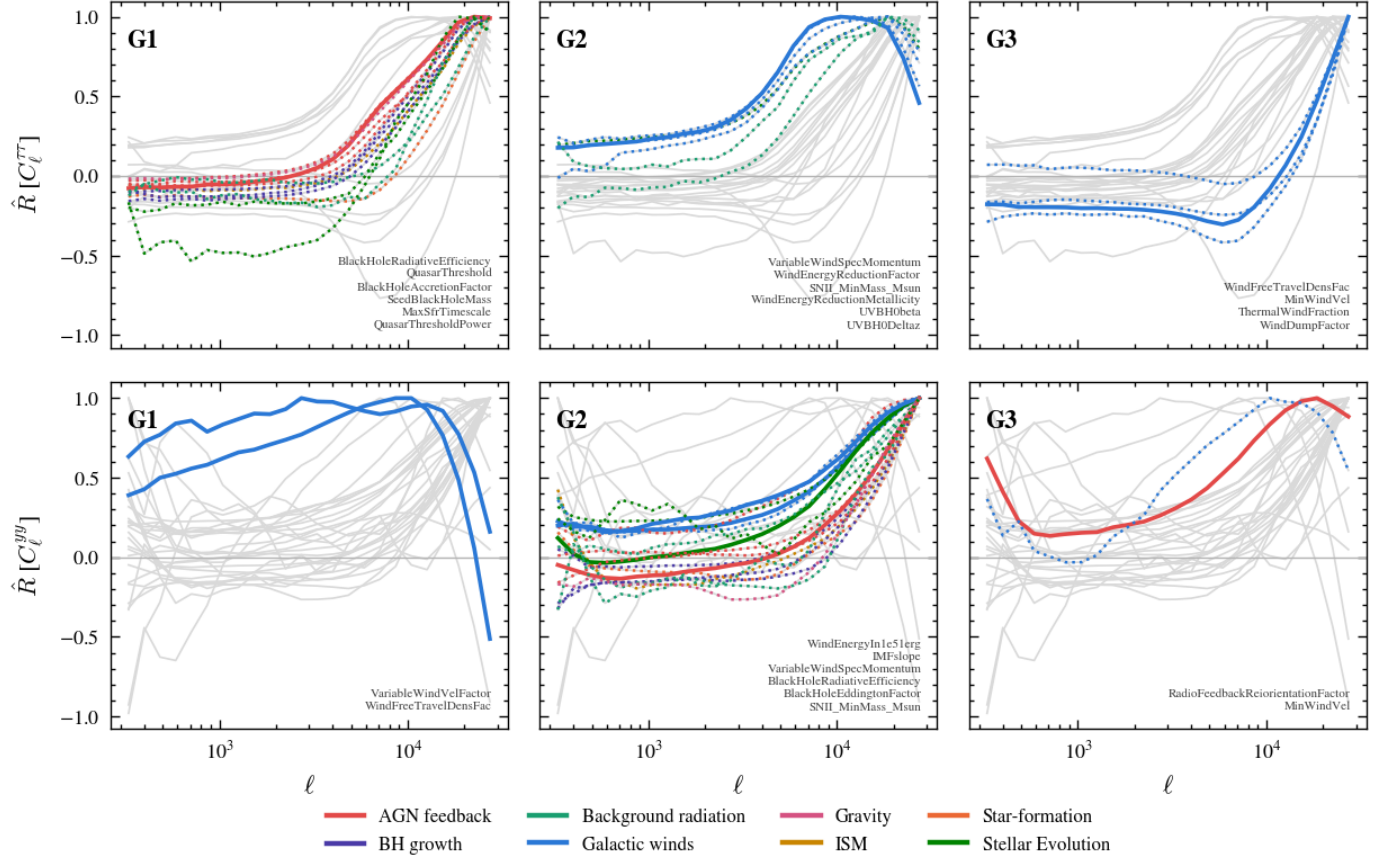


Figure 16. Response families of the gas auto-spectra, built in $\ln C_\ell$ from the 1P atlas as described in the text of this appendix. $C_\ell^{\tau\tau}$ (top row) and C_ℓ^{yy} (bottom row), with the same normalization, clustering threshold, sector colors, and line conventions as Fig. 10. The $\tau\tau$ families separate by feedback channel. G1 is the black hole growth and AGN sector, G2 is the wind energy-reduction and stellar-evolution parameters together with the UV background, and G3 is the wind travel and velocity parameters, whose response stays flat to $\ell \simeq 10^4$ before rising steeply. For C_ℓ^{yy} , the two strongest correlates of Fig. 9, VarWindVelFactor and $\text{WindFreeTravelDensFac}$, form their own family (G1), rising from large scales onward rather than only at high ℓ . The $\tau\tau$ families are built from the 26 parameters whose 1P τ planes exist (see text). The cross-spectrum families are constructed identically and are not shown.

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)

Battaglia, N. 2016, *JCAP*, 2016, 058,

doi: [10.1088/1475-7516/2016/08/058](https://doi.org/10.1088/1475-7516/2016/08/058)

Bigwood, L., Amon, A., McCarthy, I. G., et al. 2025a,

arXiv e-prints, arXiv:2510.15822,

doi: [10.48550/arXiv.2510.15822](https://doi.org/10.48550/arXiv.2510.15822)

Bigwood, L., Amon, A., Schneider, A., et al. 2024,

MNRAS, 534, 655, doi: [10.1093/mnras/stae2100](https://doi.org/10.1093/mnras/stae2100)

Bigwood, L., McCullough, J., Siegel, J., et al. 2025b, arXiv e-prints, arXiv:2512.04209,

doi: [10.48550/arXiv.2512.04209](https://doi.org/10.48550/arXiv.2512.04209)

Broxterman, J. C., Schaller, M., Schaye, J., et al. 2024,

MNRAS, 529, 2309, doi: [10.1093/mnras/stae698](https://doi.org/10.1093/mnras/stae698)

Chadayammuri, U., Ntampaka, M., ZuHone, J., Bogdán,

Á., & Kraft, R. P. 2023, *MNRAS*,

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)

Coulton, W. R., Liu, J., McCarthy, I. G., & Osato, K.

2020, *MNRAS*, 495, 2531, doi: [10.1093/mnras/staa1098](https://doi.org/10.1093/mnras/staa1098)

Dai, B., & Seljak, U. 2021, *Proceedings of the National*

Academy of Sciences, 118, e2020324118,

doi: [10.1073/pnas.2020324118](https://doi.org/10.1073/pnas.2020324118)

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)

Delgado, A. M., Anglés-Alcázar, D., Thiele, L., et al. 2023,

MNRAS, 526, 5306, doi: [10.1093/mnras/stad2992](https://doi.org/10.1093/mnras/stad2992)

Eifler, T., Krause, E., Dodelson, S., et al. 2015, *MNRAS*,

454, 2451, doi: [10.1093/mnras/stv2000](https://doi.org/10.1093/mnras/stv2000)

- Euclid Collaboration: Ajani, V., Baldi, M., Barthelemy, A., Boyle, A., et al. 2023, *A&A*, 675, A120, doi: [10.1051/0004-6361/202346017](https://doi.org/10.1051/0004-6361/202346017)
- Ganguly, A., Schaan, E., Krause, E., et al. 2026, arXiv e-prints, arXiv:2607.29091, doi: [10.48550/arXiv.2607.29091](https://doi.org/10.48550/arXiv.2607.29091)
- Gatti, M., Jain, B., Chang, C., et al. 2022, *PhRvD*, 106, 083509, doi: [10.1103/PhysRevD.106.083509](https://doi.org/10.1103/PhysRevD.106.083509)
- 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., 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)
- Giri, S. K., & Schneider, A. 2021, *JCAP*, 2021, 046, doi: [10.1088/1475-7516/2021/12/046](https://doi.org/10.1088/1475-7516/2021/12/046)
- Grandis, S., Aricò, G., Schneider, A., & Linke, L. 2024, *MNRAS*, 528, 4379, doi: [10.1093/mnras/stae259](https://doi.org/10.1093/mnras/stae259)
- Grandón, D., Marques, G. A., Thiele, L., et al. 2024, *PhRvD*, 110, 103539, doi: [10.1103/PhysRevD.110.103539](https://doi.org/10.1103/PhysRevD.110.103539)
- Hadzhiyska, B. 2026, arXiv e-prints, arXiv:2607.27149, doi: [10.48550/arXiv.2607.27149](https://doi.org/10.48550/arXiv.2607.27149)
- 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., Qu, F. J., et al. 2026, arXiv e-prints, arXiv:2604.19745, doi: [10.48550/arXiv.2604.19745](https://doi.org/10.48550/arXiv.2604.19745)
- 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)
- Harnois-Déraps, J., Martinet, N., Castro, T., et al. 2021, *MNRAS*, 506, 1623, doi: [10.1093/mnras/stab1623](https://doi.org/10.1093/mnras/stab1623)
- Hartlap, J., Simon, P., & Schneider, P. 2007, *A&A*, 464, 399, doi: [10.1051/0004-6361:20066170](https://doi.org/10.1051/0004-6361:20066170)
- Hearin, A. P., Zentner, A. R., & Ma, Z. 2012, *JCAP*, 2012, 034, doi: [10.1088/1475-7516/2012/04/034](https://doi.org/10.1088/1475-7516/2012/04/034)
- Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, *A&A*, 499, 31, doi: [10.1051/0004-6361/200811054](https://doi.org/10.1051/0004-6361/200811054)
- Hockney, R. W., & Eastwood, J. W. 1988, *Computer Simulation Using Particles* (Bristol: Hilger)
- Huang, H.-J., Eifler, T., Mandelbaum, R., & Dodelson, S. 2019, *MNRAS*, 488, 1652, doi: [10.1093/mnras/stz1714](https://doi.org/10.1093/mnras/stz1714)
- Huterer, D., & Takada, M. 2005, *Astroparticle Physics*, 23, 369, doi: [10.1016/j.astropartphys.2005.02.006](https://doi.org/10.1016/j.astropartphys.2005.02.006)
- Jain, B., Seljak, U., & White, S. 2000, *ApJ*, 530, 547, doi: [10.1086/308384](https://doi.org/10.1086/308384)
- Kacprzak, T., Kirk, D., Friedrich, O., et al. 2016, *MNRAS*, 463, 3653, doi: [10.1093/mnras/stw2070](https://doi.org/10.1093/mnras/stw2070)
- Kannan, S., Qiu, T., Cuesta-Lazaro, C., & Jeong, H. 2025, arXiv e-prints, arXiv:2507.11842, doi: [10.48550/arXiv.2507.11842](https://doi.org/10.48550/arXiv.2507.11842)
- Kratochvil, J. M., Haiman, Z., & May, M. 2010, *PhRvD*, 81, 043519, doi: [10.1103/PhysRevD.81.043519](https://doi.org/10.1103/PhysRevD.81.043519)
- Kratochvil, J. M., Lim, E. A., Wang, S., et al. 2012, *PhRvD*, 85, 103513, doi: [10.1103/PhysRevD.85.103513](https://doi.org/10.1103/PhysRevD.85.103513)
- 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)
- 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)
- Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. 2022, arXiv e-prints, arXiv:2210.02747, doi: [10.48550/arXiv.2210.02747](https://doi.org/10.48550/arXiv.2210.02747)
- Lipman, Y., Havasi, M., Holderrieth, P., et al. 2024, arXiv e-prints, arXiv:2412.06264, doi: [10.48550/arXiv.2412.06264](https://doi.org/10.48550/arXiv.2412.06264)
- Liu, J., & Haiman, Z. 2016, *PhRvD*, 94, 043533, doi: [10.1103/PhysRevD.94.043533](https://doi.org/10.1103/PhysRevD.94.043533)
- Liu, J., Petri, A., Haiman, Z., et al. 2015, *PhRvD*, 91, 063507, doi: [10.1103/PhysRevD.91.063507](https://doi.org/10.1103/PhysRevD.91.063507)
- Liu, R. H., Hadzhiyska, B., Ferraro, S., et al. 2025, arXiv e-prints, arXiv:2504.11794, doi: [10.48550/arXiv.2504.11794](https://doi.org/10.48550/arXiv.2504.11794)
- Lu, T., Haiman, Z., & Li, X. 2023, *MNRAS*, 521, 2050, doi: [10.1093/mnras/stad686](https://doi.org/10.1093/mnras/stad686)
- Lu, T., Haiman, Z., & Zorrilla Matilla, J. M. 2022, *MNRAS*, 511, 1518, doi: [10.1093/mnras/stac161](https://doi.org/10.1093/mnras/stac161)
- Marinichenko, M., van Daalen, M. P., & Sellentin, E. 2025, arXiv e-prints, arXiv:2510.09761, doi: [10.48550/arXiv.2510.09761](https://doi.org/10.48550/arXiv.2510.09761)
- Marques, G. A., Liu, J., Shirasaki, M., et al. 2024, *MNRAS*, 528, 4513, doi: [10.1093/mnras/stae098](https://doi.org/10.1093/mnras/stae098)
- Martinet, N., Castro, T., Harnois-Déraps, J., et al. 2021, *A&A*, 646, A62, doi: [10.1051/0004-6361/202039679](https://doi.org/10.1051/0004-6361/202039679)
- Martinet, N., Schneider, P., Hildebrandt, H., et al. 2018, *MNRAS*, 474, 712, doi: [10.1093/mnras/stx2793](https://doi.org/10.1093/mnras/stx2793)
- Martizzi, D., Vogelsberger, M., Artale, M. C., et al. 2019, *MNRAS*, 486, 3766, doi: [10.1093/mnras/stz1106](https://doi.org/10.1093/mnras/stz1106)
- Maturi, M., Angrick, C., Pace, F., & Bartelmann, M. 2010, *A&A*, 519, A23, doi: [10.1051/0004-6361/200912866](https://doi.org/10.1051/0004-6361/200912866)

- McCarthy, I. G., Amon, A., Schaye, J., et al. 2024, arXiv e-prints, arXiv:2410.19905, doi: [10.48550/arXiv.2410.19905](https://doi.org/10.48550/arXiv.2410.19905)
- McQuinn, M. 2016, *ARA&A*, 54, 313, doi: [10.1146/annurev-astro-082214-122355](https://doi.org/10.1146/annurev-astro-082214-122355)
- 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)
- 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)
- Osato, K., Liu, J., & Haiman, Z. 2021, *MNRAS*, 502, 5593, doi: [10.1093/mnras/stab395](https://doi.org/10.1093/mnras/stab395)
- Pandey, S., Gatti, M., Baxter, E., et al. 2022, *PhRvD*, 105, 123526, doi: [10.1103/PhysRevD.105.123526](https://doi.org/10.1103/PhysRevD.105.123526)
- Pandey, S., Hill, J. C., Alarcon, A., et al. 2025, arXiv e-prints, arXiv:2506.07432, doi: [10.48550/arXiv.2506.07432](https://doi.org/10.48550/arXiv.2506.07432)
- Pandey, S., Lehman, K., Baxter, E. J., et al. 2023, *MNRAS*, doi: [10.1093/mnras/stad2268](https://doi.org/10.1093/mnras/stad2268)
- Petri, A. 2016, *Astronomy and Computing*, 17, 73, doi: [10.1016/j.ascom.2016.06.001](https://doi.org/10.1016/j.ascom.2016.06.001)
- Petri, A., Haiman, Z., Hui, L., May, M., & Kratochvil, J. M. 2013, *PhRvD*, 88, 123002, doi: [10.1103/PhysRevD.88.123002](https://doi.org/10.1103/PhysRevD.88.123002)
- Petri, A., Haiman, Z., & May, M. 2016, *PhRvD*, 93, 063524, doi: [10.1103/PhysRevD.93.063524](https://doi.org/10.1103/PhysRevD.93.063524)
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077, doi: [10.1093/mnras/stx2656](https://doi.org/10.1093/mnras/stx2656)
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13, doi: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830)
- Popesso, P., Biviano, A., Marini, I., Dolag, K., et al. 2026, *A&A*, 707, A362, doi: [10.1051/0004-6361/202453256](https://doi.org/10.1051/0004-6361/202453256)
- Preston, C., Amon, A., & Efstathiou, G. 2023, *MNRAS*, 525, 5554, doi: [10.1093/mnras/stad2573](https://doi.org/10.1093/mnras/stad2573)
- Qu, F. J., Ried Guachalla, B., Schaan, E., Hadzhiyska, B., et al. 2026, arXiv e-prints, arXiv:2604.19744, doi: [10.48550/arXiv.2604.19744](https://doi.org/10.48550/arXiv.2604.19744)
- Salcido, J., McCarthy, I. G., Kwan, J., Upadhye, A., & Font, A. S. 2023, *MNRAS*, 523, 2247, doi: [10.1093/mnras/stad1474](https://doi.org/10.1093/mnras/stad1474)
- Schaan, E., Ferraro, S., Amodeo, S., et al. 2021, *PhRvD*, 103, 063513, doi: [10.1103/PhysRevD.103.063513](https://doi.org/10.1103/PhysRevD.103.063513)
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, *MNRAS*, 526, 4978, doi: [10.1093/mnras/stad2419](https://doi.org/10.1093/mnras/stad2419)
- Schneider, A., Giri, S. K., Amodeo, S., & Refregier, A. 2022, *MNRAS*, 514, 3802, doi: [10.1093/mnras/stac1493](https://doi.org/10.1093/mnras/stac1493)
- Schneider, A., Stoira, N., Refregier, A., et al. 2020, *JCAP*, 2020, 019, doi: [10.1088/1475-7516/2020/04/019](https://doi.org/10.1088/1475-7516/2020/04/019)
- 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)
- Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, *MNRAS*, 417, 2020, doi: [10.1111/j.1365-2966.2011.19385.x](https://doi.org/10.1111/j.1365-2966.2011.19385.x)
- Shan, H., Liu, X., Hildebrandt, H., et al. 2018, *MNRAS*, 474, 1116, doi: [10.1093/mnras/stx2837](https://doi.org/10.1093/mnras/stx2837)
- 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. 2025a, arXiv e-prints, arXiv:2509.10455, doi: [10.48550/arXiv.2509.10455](https://doi.org/10.48550/arXiv.2509.10455)
- Siegel, J., Bigwood, L., Amon, A., et al. 2025b, arXiv e-prints, arXiv:2512.02954, doi: [10.48550/arXiv.2512.02954](https://doi.org/10.48550/arXiv.2512.02954)
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, 475, 676, doi: [10.1093/mnras/stx3304](https://doi.org/10.1093/mnras/stx3304)
- Thiele, L., Hill, J. C., & Smith, K. M. 2020a, *PhRvD*, 102, 123545, doi: [10.1103/PhysRevD.102.123545](https://doi.org/10.1103/PhysRevD.102.123545)
- Thiele, L., Villaescusa-Navarro, F., Spergel, D. N., Nelson, D., & Pillepich, A. 2020b, *ApJ*, 902, 129, doi: [10.3847/1538-4357/abb80f](https://doi.org/10.3847/1538-4357/abb80f)
- Tröster, T., Ferguson, C., Harnois-Déraps, J., & McCarthy, I. G. 2019, *MNRAS*, 487, L24, doi: [10.1093/mnras/rlz075](https://doi.org/10.1093/mnras/rlz075)
- van Daalen, M. P., Koutalios, I., Broxterman, J. C., et al. 2025, arXiv e-prints, arXiv:2509.04552, doi: [10.48550/arXiv.2509.04552](https://doi.org/10.48550/arXiv.2509.04552)
- 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)
- van Loon, M. L., & van Daalen, M. P. 2024, *MNRAS*, 528, 4623, doi: [10.1093/mnras/stae285](https://doi.org/10.1093/mnras/stae285)
- 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)
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, *MNRAS*, 465, 3291, doi: [10.1093/mnras/stw2944](https://doi.org/10.1093/mnras/stw2944)
- Wright, A. H., Stölzner, B., Asgari, M., et al. 2025, *A&A*, 703, A158, doi: [10.1051/0004-6361/202554908](https://doi.org/10.1051/0004-6361/202554908)
- Yang, X., Kratochvil, J. M., Wang, S., et al. 2011, *PhRvD*, 84, 043529, doi: [10.1103/PhysRevD.84.043529](https://doi.org/10.1103/PhysRevD.84.043529)
- Yang, Y., Bird, S., Zhou, Y., et al. 2026, arXiv e-prints, arXiv:2605.04176, doi: [10.48550/arXiv.2605.04176](https://doi.org/10.48550/arXiv.2605.04176)
- Zennaro, M., Angulo, R. E., Aricò, G., Contreras, S., & Pellejero-Ibañez, M. 2019, *MNRAS*, 489, 5938, doi: [10.1093/mnras/stz2612](https://doi.org/10.1093/mnras/stz2612)

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)

Zinger, E., Pillepich, A., Nelson, D., et al. 2020, MNRAS, 499, 768, doi: [10.1093/mnras/staa2607](https://doi.org/10.1093/mnras/staa2607)

Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., & Refregier, A. 2021, JCAP, 2021, 028, doi: [10.1088/1475-7516/2021/01/028](https://doi.org/10.1088/1475-7516/2021/01/028)

Zürcher, D., Fluri, J., Sgier, R., et al. 2022, MNRAS, 511,

2075, doi: [10.1093/mnras/stac078](https://doi.org/10.1093/mnras/stac078)