Interpretable Equivariant Marks for Contrastive Cosmological Inference

Federico Semenzato [1 2] **Benjamin D. Wandelt** [3 4] **Michele Liguori** [1 2 5] **Alvise Raccanelli** [1 2 5]

Abstract

Recovering cosmological information beyond the power spectrum is a central goal for upcoming cosmological surveys, since late-time non-Gaussian signal in the matter density cannot be accessed through two-point statistics alone. Marked statistics fold part of this information back into the two-point level by reweighting the field with non-linear functions. We propose a neural marking scheme to generalize this process through a set of interpretable, physically motivated transformations that directly allow to interpret the gain in cosmological information at the morphological level. We employ a contrastive learning objective to align learnable marked summaries with the underlying cosmological parameters. At $k_{\max} = 0.2\,h\mathrm{Mpc}^{-1}$, our neural mark tightens the marginalized constraint on σ_8 by $2.9\times$ and on Ω_m by $1.8\times$ compared to classical marks, breaking the $\Omega_m - \sigma_8$ degeneracy at the Fisher information level. It further reduces the parameter MSE across our cosmological parameter prior by $1.45\times$ over the best classical mark. The learned latent geometry aligns with the Ω_m and σ_8 directions in parameter space, indicating that the contrastive objective recovers the dominant axes of cosmological information. Our approach opens the door to more powerful, interpretable summary statistics for cosmological inference.

[1]Dipartimento di Fisica Galileo Galilei, Università di Padova, I-35131 Padova, Italy [2]INFN Sezione di Padova, I-35131 Padova, Italy [3]Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD, 21218, USA [4]Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD, 21218, USA [5]INAF-Osservatorio Astronomico di Padova, Italy. Correspondence to: Federico Semenzato <federico.semenzato.1@phd.unipd.it>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

1. Introduction

Large-scale structure (LSS) surveys provide three-dimensional maps of millions of galaxies, with the current and next generation of surveys expanding the scale and precision of these maps (e.g., SPHEREx (Doré et al., 2014), DESI (Aghamousa et al., 2016), Euclid (Mellier et al., 2024)). Their statistical analysis can in principle constrain some of the deepest questions in fundamental physics: the amplitude of primordial non-Gaussianity and the inflationary models it discriminates (Lucchin & Matarrese, 1988; Maldacena & Martin, 2003; Dalal et al., 2008; Matarrese & Verde, 2008), the dark energy equation of state (Adame et al., 2025a), and the absolute neutrino mass scale (Lesgourgues & Pastor, 2006; Elbers et al., 2025). Much of this signal lives in correlations of the late time matter density that the Gaussian two-point structure alone cannot capture (Bernardeau et al., 2002; Scoccimarro et al., 1999; Sefusatti & Komatsu, 2007); extracting it is crucial for the scientific return of these surveys.

Two-point summaries break in the non-linear regime. The power spectrum is the canonical summary of LSS clustering: optimal for Gaussian fields, modelable with perturbation theory on quasi-linear scales, and well understood at the level of covariance and systematics (Tegmark, 1997; Percival et al., 2004; Adame et al., 2025b). The information beyond the linear regime lives in higher-order correlations (Scoccimarro et al., 1999; Verde et al., 2000; Bernardeau et al., 2002; Bertacca et al., 2014) but climbing the ladder of n-point correlators directly is expensive in configuration space, perturbation theory breaks down in the non-linear regime, and the Gaussian-likelihood assumptions underlying the choice of these summaries become increasingly fragile.

Field-level and hybrid neural summaries. Simulation-based inference and field-level neural summaries sidestep explicit summaries by leveraging simulated data and mapping them directly to parameter posteriors (Charnock et al., 2018; Makinen et al., 2021; Lemos et al., 2024). These approaches require many high-fidelity simulations, and trade constraining power for interpretability of the features that drive cosmological parameter constraints. Hybrid strategies blend neural summaries with traditional statistics, capturing

information beyond analytically tractable summaries while leaving the neural component opaque (Makinen et al., 2025; Bairagi & Wandelt, 2026).

Marked statistics. A complementary route reweights the field by a spatial mark $M(\mathbf{x})$; the two-point spectra of the resulting marked field then mix higher-order correlators of the original field (Philcox et al., 2020; Marinucci et al., 2025; Ebina et al., 2026). The summary stays a power spectrum, while the choice of mark sets which higher-order structure is folded down to the two-point level (White, 2016; Massara et al., 2021). Existing constructions fix this transformation from a smoothed-density response (White, 2016; Massara et al., 2021; Ebina & White, 2025). Recent work extends classical mark functions by optimizing their functional shape for a fixed smoothing scale R through Fisher information maximization (Cowell et al., 2024) at the fiducial cosmology. Both restrict the mark to narrowly parametrized forms and keep it tied to a single point in cosmological parameter space.

Our contribution. We replace the hand-designed mark with a learned, cosmology-agnostic one that preserves interpretability through physically motivated architectural constraints. The mark is built from three, locally SO(3)-equivariant spherical harmonic filters ($\ell \in \{0, 1, 2\}$) reduced to four rotation-invariant scalar channels and processed by independent MLPs that additively combine into the mark function. Rather than assuming which environmental features matter the most, we can learn which aspects of the local cosmic web geometry carry the most cosmological information. The result is a trained marking function that remains inspectable in configuration space and encodes which local morphological features carry the richest cosmological signal.

To search this enlarged mark space, we borrow a tool from multimodal representation learning. We align marked summaries with cosmological parameters $\boldsymbol{\theta}$ in a shared latent space using a contrastive InfoNCE objective (van den Oord et al., 2018; Radford et al., 2021; Chen et al., 2020) with a learned Mahalanobis metric. We residualize the marked-summary embedding against the unmarked embedding so the mark is rewarded only for complementary information. On Quijote N-body simulations, the resulting summary tightens marginalized Fisher constraints by $2.9\times$ on σ_8 and $1.8\times$ on Ω_m at $k_{\max} = 0.2\,h\mathrm{Mpc}^{-1}$, breaks the canonical Ω_m–σ_8 degeneracy, and reduces held-out parameter MSE across the parameter space by up to $\sim 1.45\times$ over the best classical marks.

2. Methodology

Our framework is built of two main components. A learnable mark architecture extracts rotation-invariant local descriptors of the density field (Section 2.2), and a contrastive objective aligns marked summaries to cosmological parameters in a shared latent space (Section 2.3). We introduce an explicit residualization of the marked-statistic embeddings against the unmarked-statistic embeddings so the mark is rewarded only for information it adds beyond the standard power spectrum.

2.1. Marked Two-Point Summaries

We define a mark field $M(\mathbf{x})$ as a scalar function over the simulation volume, which weights different regions of the matter density contrast $\delta(\mathbf{x})$ according to their local properties. The (mean-centered) marked density field then reads

$$\Delta(\mathbf{x}) = M(\mathbf{x})\,[1 + \delta(\mathbf{x})] - \langle M(1+\delta)\rangle \ . \tag{1}$$

To map the marked field to cosmological parameters, we use the set of two-point power spectra of the marked and unmarked fields, that is

$$\mathbf{S}(k) = \{P_{\delta\delta}(k),\ P_{\delta\Delta}(k),\ P_{\Delta\Delta}(k)\}\ , \tag{2}$$

$$\langle a(\mathbf{k}) b^*(\mathbf{k}')\rangle = (2\pi)^3 \delta_D(\mathbf{k}-\mathbf{k}') P_{ab}(k)\,, \tag{3}$$

with $a, b \in \{\delta, \Delta\}$. Because Δ is nonlinear in δ, $P_{\Delta\Delta}$ and $P_{\delta\Delta}$ include higher-order correlations of the original field (Philcox et al., 2020; Marinucci et al., 2025). The non-linear transformation $M(\delta_R)$ effectively folds higher-order clustering into a modified two-point function. Crucially, (Cowell et al., 2024) showed that the constraints on cosmological parameters derived from the combination of auto- and cross-spectra $\{P_{\delta\delta}, P_{\delta\Delta}, P_{\Delta\Delta}\}$ are invariant for affine transformations of the mark $M' \to aM + b$.

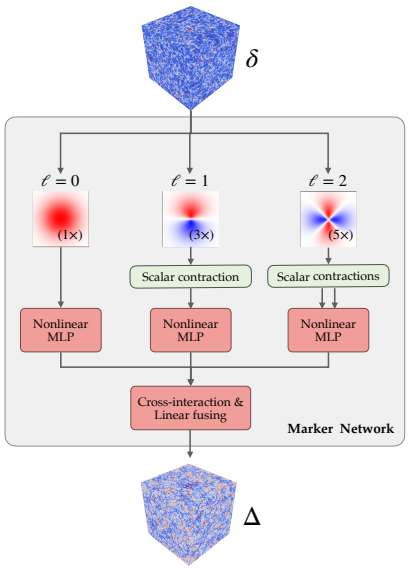
Classical marks usually depend on a smoothed density δ_R,

$$M(\mathbf{x}) = \left(\frac{1+\delta_s}{1+\delta_s+\delta_R(\mathbf{x})}\right)^p\,, \tag{4}$$

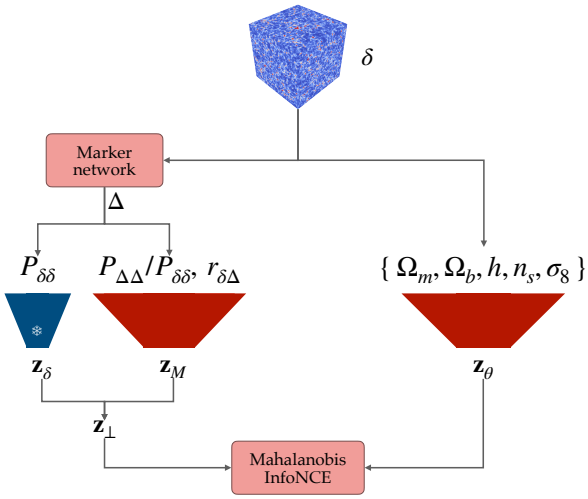
with smoothing scale R and saturation parameter δ_s (White, 2016; Massara et al., 2021). Positive p up-weights underdense regions; negative p emphasizes overdense or nonlinear regions. Learned marks expand the transformation space while preserving the two-point output in Equation (3).

2.2. Interpretable Learned Marks

We make use of matter density fields at $z = 0$ produced by the Quijote simulation suite of N-body simulations (Villaescusa-Navarro et al., 2018). Particles are placed in periodic boxes of side length $1\ h^{-1}\mathrm{Gpc}$, and assigned to 128^3 grids. The mark is therefore a translation-equivariant local rule applied to every voxel.



(a) Interpretable mark module.



(b) Summary alignment and contrastive setup.

Figure 1. The learned mark aims to extract interpretable features from the density field. (a) Spherical-harmonic filters ($\ell = 0, 1, 2$) produce density-like, gradient-like, and quadrupole-like responses. Scalar contractions form rotation invariant fields which are independently processed by MLPs and linearly combined with a small cross interaction. The resulting mark M produces a marked field Δ. (b) The $P_{\delta\delta}$ embedder is pretrained against parameters and frozen; the marked summary embedding $\mathbf{z}_M$ is residualized against $\mathbf{z}_\delta$ to give $\mathbf{z}_\perp$, which is aligned with $\mathbf{z}_\theta$ through a learned Mahalanobis-form InfoNCE loss.

We decompose the density field into spherical-harmonic-aware filtered channels and then combine them into scalar invariants. The filtered maps are

$$F_{\ell m}(\mathbf{x}) = \mathcal{F}^{-1}\big[\tilde{\delta}(\mathbf{k}) W_{\rm MAS}^{-1}(\mathbf{k}) T_{\rm tap}(k) \times G_\ell(k) i^\ell Y_{\ell m}(\hat{\mathbf{k}})\big](\mathbf{x}), \tag{5}$$

for $\ell \leq 2$. Here $G_\ell(k)$ is a learned radial profile in Fourier space and $Y_{\ell m}$ is a real spherical harmonic. The factor $W_{\rm MAS}^{-1}$ removes the mass-assignment window used to place particles on the grid (Hand et al., 2018). The spherical Nyquist cutoff $T_{\rm tap}(k)$ suppresses high-k modes amplified by the deconvolution that cannot be reliably modeled because of aliasing, with a mild Tukey cosine roll-off to control cubic-grid artifacts. Each $G_\ell(k)$ is parameterised as a learnable Gaussian band-pass with shared center r_0 and width σ plus a per-ℓ residual MLP on log-spaced Fourier features, zero-initialized (Mildenhall et al., 2020).

The $\ell = 0$ block is a signed monopole response, $E_0 = F_{00}$, while the $\ell = 1$ block gives a vector-like response whose invariant amplitude is

$$E_1 = \left(\sum_{m=-1}^{1} F_{1m}^2 + \epsilon \right)^{1/2}. \tag{6}$$

For $\ell = 2$, the five real spherical-harmonic components F_{2m} map to the five independent entries of a traceless symmetric tensor Q, from which we extract two natural scalar invariants:

$$E_2 = \left[\mathrm{Tr}(Q^2) + \epsilon\right]^{1/2}, \qquad I_3 = \frac{\mathrm{Tr}(Q^3)}{E_2^3 + \epsilon}. \tag{7}$$

E_2 measures anisotropy strength, while I_3 distinguishes prolate and oblate quadrupole shape.

The invariant channels are processed by independent scalar MLPs,

$$M(\mathbf{x}) = \mathrm{softplus}\,[\eta(\mathbf{x})], \tag{8}$$

$$\eta(\mathbf{x}) = \sum_a h_a(f_a(\mathbf{x})) + h_\times(E_0(\mathbf{x}), E_2(\mathbf{x}), I_3(\mathbf{x})), \tag{9}$$

where $f_a \in \{E_0, E_1, E_2, I_3\}$. The cross term $h_\times(E_0, E_2, I_3)$ couples density to tidal-like features E_2 and I_3 to provide a probe of how density modulates anisotropic structure. Richer cross-interactions and full equivariant mixing of the ℓ-channels are natural extensions we leave to future work.

The output is initialized at $M(\mathbf{x}) \simeq 1$, so training begins at the unmarked field. Because each h_a is separate, the learned mark can be decomposed into distinct contributions before spectra are evaluated.

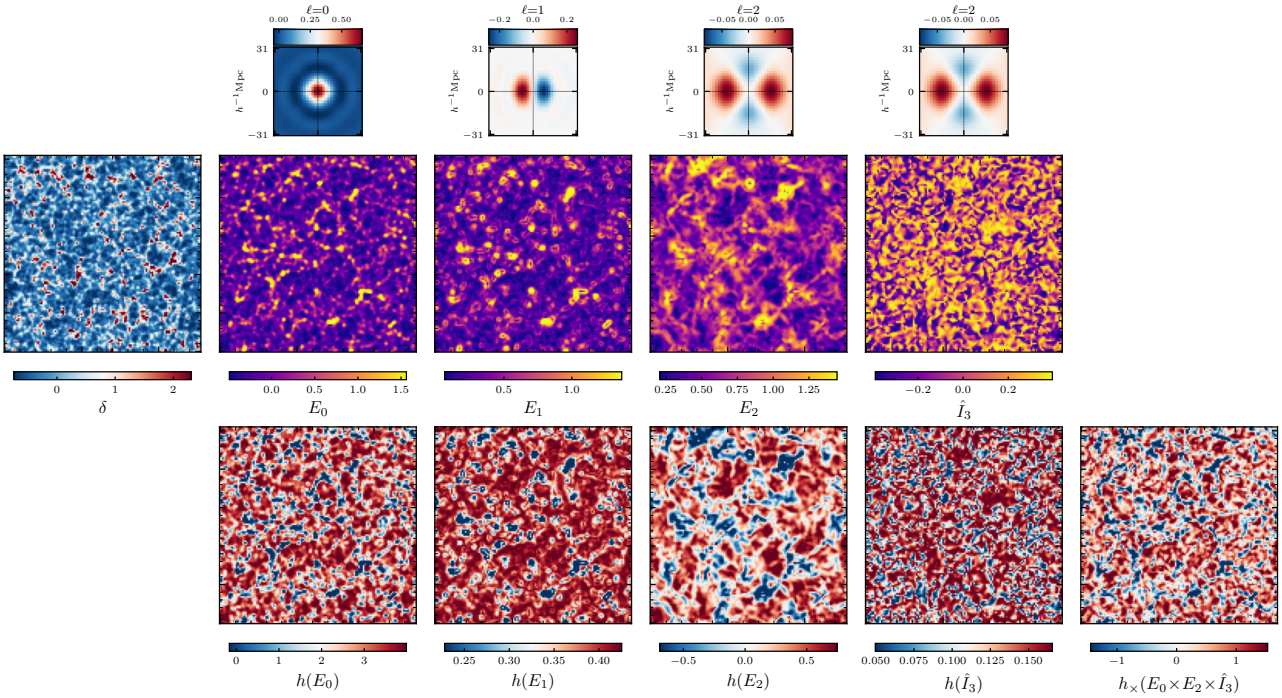


Figure 2. End-to-end introspection of the learned mark on a representative held-out simulation. **Top row**: real-space projection of the per-ℓ kernels reconstructed from the trained $G_\ell(k)$, for $\ell = 0, 1, 2$. **Middle row**: the input density field δ and the four rotation-invariant maps E_0, E_1, E_2, I_3 extracted from the spherical-harmonic-filtered field. **Bottom row**: the learned scalar response contributions $h_a(E_0)$, $h_a(E_1)$, $h_a(E_2)$, $h_a(I_3)$, and the cross-interaction term $h_\times(E_0, E_2, I_3)$ that combine into the mark M. The isotropic ($\ell = 0$) channel dominates the response on these scales; the anisotropic channels are present but less dominant.

2.3. Contrastive Training

The contrastive objective operates on the summary vectors for each simulation. The unmarked summary is

$$\mathbf{S}_\delta = \log_{10} P_{\delta\delta}(k)\,, \tag{10}$$

while the marked summary is built as

$$\mathbf{S}_M(k) = \left[\log_{10} \frac{P_{\Delta\Delta}(k)}{P_{\delta\delta}(k)},\ r(k)\right], \tag{11}$$

$$r(k) = \frac{P_{\delta\Delta}(k)}{\sqrt{P_{\delta\delta}(k) P_{\Delta\Delta}(k)}}. \tag{12}$$

For simulation i, the positive pair is therefore $(\mathbf{S}_{M,i}, \boldsymbol{\theta}_i)$. We project the two summary vectors and the cosmological parameters into a shared D-dimensional embedding space, where the contrastive loss is applied. The parameter projector is chosen to be linear.

In latent space, we remove the component of the marked embedding parallel to the unmarked embedding,

$$\mathbf{z}_\perp = \mathbf{z}_M - \frac{\mathbf{z}_M \cdot \mathbf{z}_\delta}{\mathbf{z}_\delta \cdot \mathbf{z}_\delta + \epsilon}\, \mathbf{z}_\delta. \tag{13}$$

This projection discourages the learned mark from simply reproducing information already available in $P_{\delta\delta}$. The contrastive loss then aligns each marked summary embedding $\mathbf{z}_{\perp,i}$ with the corresponding parameter embedding vector $\mathbf{z}_{\theta,i}$, while separating it from a set $\mathcal{N}_i$ of specific negative candidates $\mathbf{z}_{\theta,j}$. We use an InfoNCE loss (van den Oord et al., 2018):

$$\mathcal{L}_i = -\log \frac{\exp s(\mathbf{z}_{\perp,i}, \mathbf{z}_{\theta,i})}{\sum_{j \in \mathcal{N}_i^+} \exp s(\mathbf{z}_{\perp,i}, \mathbf{z}_{\theta,j})}, \tag{14}$$

where $\mathcal{N}_i^+$ includes both the negatives $\mathcal{N}_i$ and the true parameter embedding. Since generating $\mathbf{z}_{\theta,j}$ only requires a cosmological parameter vector without the respective $\mathbf{z}_{\perp,j}$ (and thus a paired density field), we can efficiently design the negative sampling scheme. We extract these contrastive negatives from three regions of the Quijote prior: real negatives are drawn from the training set and shared across elements of a batch; global synthetic negatives sample the full prior volume; and local synthetic negatives sample a shell around the anchor cosmology. The local candidates force the summary to distinguish nearby cosmologies without requiring additional simulations. The score is a Mahalanobis distance,

$$s(\mathbf{z}_\perp, \mathbf{z}_\theta) = -\tau^{-1} (\mathbf{z}_\perp - \mathbf{z}_\theta)^T L L^T (\mathbf{z}_\perp - \mathbf{z}_\theta), \tag{15}$$

with a learned lower-triangular factor L and fixed temperature τ. This metric allows rotations and stretchings of latent

space to better align the summary geometry with the parameter geometry, while still penalizing large distances in any direction.

We first pretrain the unmarked branch to align $P_{\delta\delta}$ with parameter embedding. We then freeze the $P_{\delta\delta}$-embedder weights when training the marker module. The parameter embedder is initialized to the configuration learned for $P_{\delta\delta}$ and allowed to fine-tune during mark training, so the final geometry is not fixed by the unmarked summary. Without this bootstrap, the unmarked embedder could collapse and make the complementarity constraint meaningless. Loss-function hyperparameters and optimizer settings are reported in Section B.1.

3. Experiments

We train our contrastive framework on a set of $5,000$ simulations from the Quijote BSQ suite (Bairagi et al., 2025). The dataset comprises simulations with 5 varying cosmological parameters $\boldsymbol{\theta} = (\Omega_m, \Omega_b, h, n_s, \sigma_8)$, with $\Omega_m \in [0.1, 0.5]$, $\Omega_b \in [0.02, 0.08]$, $h \in [0.5, 0.9]$, $n_s \in [0.8, 1.2]$, and $\sigma_8 \in [0.6, 1.0]$. Density fields are assigned to 128^3 grids, giving a cell size of $7.8\ h^{-1}\mathrm{Mpc}$ and $k_{\mathrm{Nyq}} =\simeq 0.4\ h\,\mathrm{Mpc}^{-1}$. The contrastive training uses spectra through $k \simeq 0.3\ h\,\mathrm{Mpc}^{-1}$, to avoid artifacts close to k_{Nyq}. Once the contrastive training has converged, the learned mark is a fixed, interpretable transform of the density field. We can then evaluate the per-channel response inside the mark itself, and extract marked spectra for regular downstream cosmological constraints. The code developed for this analysis can be found here.

3.1. Mark Introspection

Figure 2 traces the trained mark from the input density down to the per-channel scalar contributions that combine to give $M(\mathbf{x})$. The top row shows the per-ℓ kernels reconstructed from the trained $G_\ell(k)$: the monopole is broad and approximately isotropic, the dipole resembles a derivative-of-Gaussian profile, and the quadrupole exhibits the expected oriented six-lobe pattern. The middle row shows the bare density δ alongside the rotation invariants E_0, E_1, E_2 and the cubic I_3: E_1 peaks at the boundaries between voids and filaments, E_2 in elongated filamentary regions, and I_3 separates oblate from prolate quadrupole configurations at fixed E_2. The bottom row shows the per-channel scalar responses $h_a(E_a)$ together with the cubic cross-interaction $h_\times(E_0, E_2, I_3)$, which sum (modulo the output activation) to the scalar mark M. At this resolution and $k_{\max}$, the response is dominated by the isotropic E_0 channel: the trained $G_0(k)$ acts as a high-pass filter, so the network effectively recovers a non-linearly reweighted version of the density. The dipole and quadrupole responses contribute subdominant corrections at structure boundaries

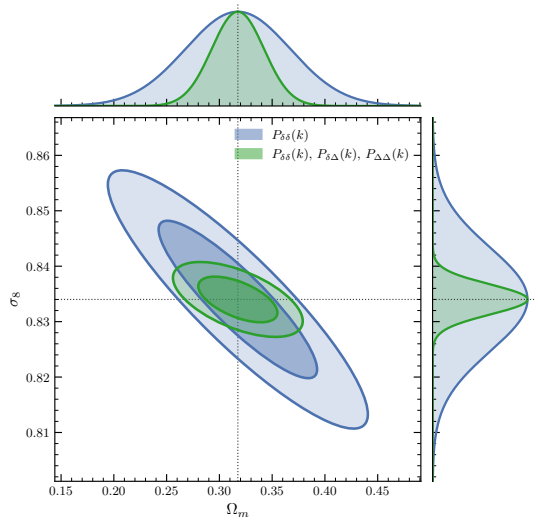


Figure 3. Marginalized posterior contours from the Fisher information matrix at the fiducial cosmology, for $P_{\delta\delta}$ alone (blue) and the learned mark $\{P_{\delta\delta}, P_{\delta\Delta}, P_{\Delta\Delta}\}$ (green). The learned mark tightens the one-dimensional posteriors on every parameter and rotates the Ω_m–σ_8 contour, breaking the canonical degeneracy and enabling simultaneous improvement on both parameters.

and in elongated regions, and I_3 adds only a weak modulation. In Section A we quantify the relative contribution of the anisotropic channels in terms of Fisher constraints, comparing the full mark to an isotropic $\ell_{\max} = 0$ ablation. E_0 carries the largest fraction of the gain in mrginalized confidence level, with the anisotropic channels refining the constraints in this setup. The additive form of Equation (9) makes this channel decomposition exact rather than a saliency proxy: the panels in the bottom row are literally the terms that combine into $M(\mathbf{x})$, so the dominance of E_0 is a property of the trained mark itself, not an artifact of visualization. The same observation motivates the morphology-aware extensions discussed in Section 4.

3.2. Summary-Level Constraints

We evaluate the learned summary on two complementary tasks at $k_{\max} = 0.20\ h\,\mathrm{Mpc}^{-1}$, well inside the regime where mass-assignment artifacts are negligible and the learned mark is active. At higher $k_{\max} \geq 0.3\,h\mathrm{Mpc}^{-1}$, the anti-aliasing taper required by the spherical-harmonic filtering on the cubic grid removes power that classical pixel-space marks retain, and some classical marks become competitive (see Section A). Extensions of this framework to morphology-based filters, which we will report in follow-up work, restore uniform superiority.

The Fisher information matrix (Tegmark, 1997; Heavens & Taylor, 1995) provides the standard forecasting tool at the fiducial cosmology. Under a Gaussian likelihood with covariance C and parameter-dependent mean $\mu(\boldsymbol{\theta})$, $F_{\alpha\beta} =$

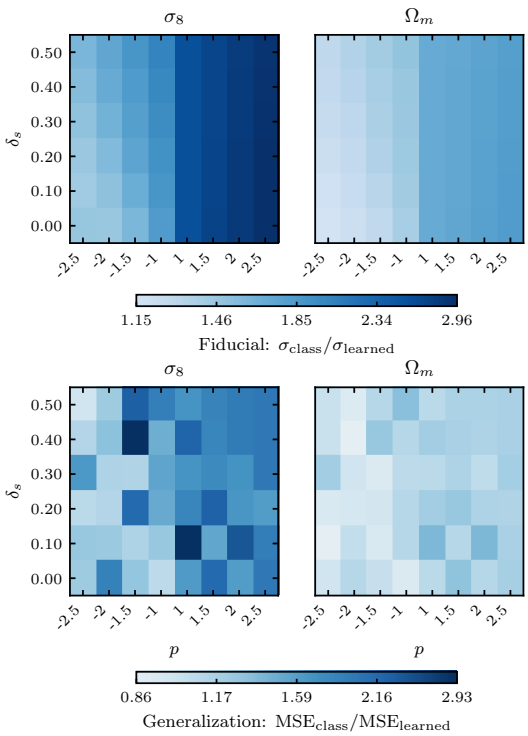


Figure 4. Comparison of learned and classical marks on both Fisher and held-out generalization at $k_{\max} = 0.20\,h\,\mathrm{Mpc}^{-1}$. **Top**: Improvement in the marginalized confidence intervals relative to a set of classical marks from (Massara et al., 2021) with $R = 10\,h^{-1}\mathrm{Mpc}$. **Bottom**: Mean-squared error from a MLP regressor on Latin-hypercube simulations unseen during training; the learned mark beats the unmarked baseline and the best classical mark, with the largest gains on σ_8.

$\nabla_\alpha \mu^T C^{-1} \nabla_\beta \mu$, and the Cramér–Rao inequality bounds the achievable variance on any unbiased estimator by F^{-1}. We compute the covariance by evaluating the spectra on $10,000$ fiducial simulations and the derivatives by finite differences on 500 simulations per parameter. These fiducial and derivative simulations are disjoint from the BSQ set used for contrastive training and from the Latin-hypercube set used for the held-out tests.

In Figure 3 we show the marginalized posterior contours for $P_{\delta\delta}$ alone and the learned mark summaries $\{P_{\delta\delta}, P_{\delta\Delta}, P_{\Delta\Delta}\}$. The learned mark tightens the one-dimensional posteriors on every parameter. Importantly, the additional marked spectra also allow the contour to rotate, breaking the canonical Ω_m–σ_8 degeneracy and enabling simultaneous improvement on both parameters. In the top row of Figure 4 we compare the improvement in the marginalized confidence intervals relative to a set of classical marks from (Massara et al., 2021) with $R = 10\,h^{-1}\mathrm{Mpc}$, which includes the best mark parameters found in that work. The learned mark improves on the best classical mark for every parameter, with the largest gains on σ_8 and Ω_m.

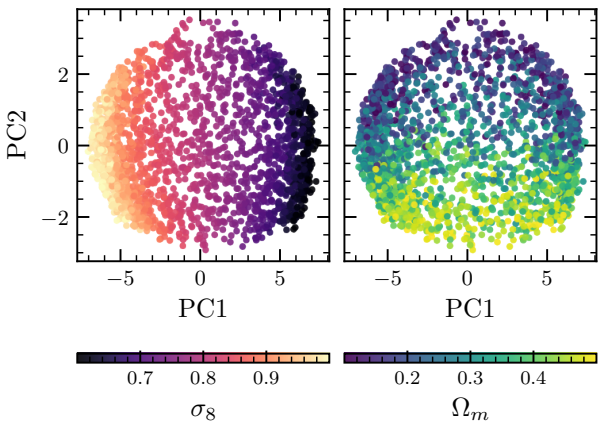


Figure 5. Two-dimensional PCA projection of the complementary query embedding $\mathbf{z}_\perp$ for held-out simulations, colored by σ_8 (left) and Ω_m (right). σ_8 varies smoothly along the first principal direction and Ω_m along an approximately orthogonal one. The first two PCs therefore align with the most well-constrained parameters of the problem.

On the other hand, with a further held-out test set spanning the entire Hypercube, we train an MLP from each summary to the five-parameter vector to evaluate generalization performance across the full parameter volume. In Section B.2 we provide details of the MLP architecture and training procedure. In the bottom row of Figure 4 we show the mean-squared error (MSE) from this regressor on Latin-hypercube simulations unseen during training. The learned mark presents good generalization performance with respect to the configurations tested, with particular gains on σ_8. For a small number of (p, δ_s) configurations, the per-parameter MSE ratio dips marginally below unity on Ω_m, so the best classical marks remain slightly more stable on Ω_m. Conversely, the learned density modulation efficiently generalizes on σ_8 across all configurations tested.

Adding the marked spectra to $P_{\delta\delta}$ tightens the marginalized Fisher posteriors on every cosmological parameter and the learned summary reaches a lower error than any classical (p, δ_s, R) choice on the same grid, with the largest gains on σ_8 and Ω_m. The held-out MLP delivers the same conclusion in the generalization regime. Across both tasks, the learned mark outperforms classical mark configurations in this setting.

3.3. Latent-Space Geometry

We further explore the learned geometry of the marked summary embedding $\mathbf{z}_\perp$. The contrastive objective shapes the marked-summary embedding $\mathbf{z}_\perp$ into a $D = 16$ space, but the trained representation has effective rank ≈ 3.4, with the first two principal components carrying $\sim 85\%$ of the vari-

ance. Most of the signal therefore lives on a small subspace, which is consistent with the intrinsic dimensionality of the cosmological-parameter target.

Figure 5 shows that the first two principal components of $\mathbf{z}_\perp$ on held-out simulations align almost perfectly with σ_8 and Ω_m, exactly the two parameters that dominate the constraint gain in Figure 4. This alignment is a representation-level signature consistent with the breaking of the Ω_m–σ_8 degeneracy at the parameter constraint level.

Using the same held-out simulations as in Section 3.2, we evaluate the fraction $R@k$ of simulations whose true parameters fall among the k nearest candidates. For each of the 2000 Latin-hypercube simulations i, we compute the metric distance (see Equation (15)) between the orthogonalized mark embedding $\mathbf{z}_{q,i}$ and all the other parameter embeddings $\mathbf{z}_{\theta,j}$ in the set, with the matched index $j=i$ as the only correct candidate. In terms of recall-at-k, we obtain $R@1=66.5\%$, $R@5=96.1\%$, against a random baseline $R@1=1/N=0.05\%$. This test does not include the intrinsic cosmic-variance scatter of the summaries, which sets a floor on the achievable retrieval performance from a single realization. Unlike the Fisher forecast in Section 3.2, which measures *local* sensitivity at the fiducial point, this test probes *global* discriminability across the entire hypercube.

4. Conclusions

We have introduced a learnable, interpretable generalization of marked statistics for cosmological density fields. The mark is a translation-equivariant scalar response built from spherical-harmonic-filtered, rotation-invariant channels processed by independent MLPs, so the trained transformation can be opened up and read as a list of morphological preferences before any spectrum is computed. A contrastive objective aligns the marked summary with the cosmological parameters across the full Latin-hypercube and is residualized against the unmarked $P_{\delta\delta}$, so the mark is rewarded only for the complementary signal it adds.

On simulated density fields at $z=0$, the resulting summary improves both Fisher constraints at the fiducial cosmology and held-out parameter recovery over the unmarked $P_{\delta\delta}$ and classical density marks at $k_{\max}=0.20\,h\,\mathrm{Mpc}^{-1}$, with the largest gains in Ω_m and σ_8 and a rotation of the joint Ω_m–σ_8 contour that breaks the canonical degeneracy. Mark introspection traces the gain to a small-scale response that recovers signal lost by the unmarked $P(k)$. In the latent geometry, the first two principal components of the learned summary embeddings align almost perfectly with σ_8 and Ω_m.

Richer geometrically motivated invariants represent a natural extension to provide more refined interpretability, targeting specific morphological features to better understand the physical origin of the gain and to further improve the summary's constraining power, and we plan to explore this direction in future work. Furthermore, the same approach could be applied to other set of cosmological parameters such as primordial non-Gaussianity, which inherently requires higher-order statistics to be constrained.

This approach represents a step towards more flexible and interpretable summaries for cosmological inference.

Acknowledgements

The authors thank Marco Marinucci, Francesco Spezzati and Nicola Bellomo for useful discussions. The authors also thank the anonymous reviewers for their constructive feedback, which was useful in refining the presentation of our work. FS is partly supported by ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union - NextGenerationEU. ML acknowledges support by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2022 - grant 20228RMX4A. AR acknowledges funding from the Italian Ministry of University and Research (MIUR) through the "Dipartimenti di eccellenza" project "Science of the Universe".

Impact Statement

This work develops scientific machine-learning tools for cosmological inference. More informative and interpretable summaries can reduce simulation cost and improve the scientific return of galaxy survey data. We do not identify specific harmful applications.

References

Adame, A. G. et al. DESI 2024 VII: cosmological constraints from the full-shape modeling of clustering measurements. *JCAP*, 07:028, 2025a. doi: 10.1088/1475-7516/2025/07/028.

Adame, A. G. et al. DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations. *JCAP*, 02:021, 2025b. doi: 10.1088/1475-7516/2025/02/021.

Aghamousa, A. et al. The DESI Experiment Part I: Science,Targeting, and Survey Design. 10 2016.

Bairagi, A. and Wandelt, B. PatchNet: A hierarchical approach for neural field-level inference from Quijote simulations. *JCAP*, 03:028, 2026. doi: 10.1088/1475-7516/2026/03/028.

Bairagi, A., Wandelt, B., and Villaescusa-Navarro, F. The BIG SOBOL SEQUENCE: How many simulations do

we need for simulation-based inference in cosmology? *Astron. Astrophys.*, 703:A301, 2025. doi: 10.1051/0004-6361/202554602.

Bernardeau, F., Colombi, S., Gaztanaga, E., and Scoccimarro, R. Large scale structure of the universe and cosmological perturbation theory. *Phys. Rept.*, 367:1–248, 2002. doi: 10.1016/S0370-1573(02)00135-7.

Bertacca, D., Maartens, R., and Clarkson, C. Observed galaxy number counts on the lightcone up to second order: II. Derivation. *JCAP*, 11:013, 2014. doi: 10.1088/1475-7516/2014/11/013.

Charnock, T., Lavaux, G., and Wandelt, B. D. Automatic physical inference with information maximizing neural networks. *Phys. Rev. D*, 97(8):083004, April 2018. doi: 10.1103/PhysRevD.97.083004.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. *arXiv e-prints*, art. arXiv:2002.05709, feb 2020. doi: 10.48550/arXiv.2002.05709.

Cowell, J. A., Alonso, D., and Liu, J. Optimizing marked power spectra for cosmology. *Mon. Not. Roy. Astron. Soc.*, 535(4):3129–3140, 2024. doi: 10.1093/mnras/stae2492.

Dalal, N., Doré, O., Huterer, D., and Shirokov, A. The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects. *Phys. Rev. D*, 77:123514, 2008. doi: 10.1103/PhysRevD.77.123514.

Doré, O. et al. Cosmology with the SPHEREX All-Sky Spectral Survey. 12 2014.

Ebina, H. and White, M. An analytically tractable marked power spectrum. *JCAP*, 01:150, 2025. doi: 10.1088/1475-7516/2025/01/150.

Ebina, H., White, M., and Chaussidon, E. The Marked Power Spectrum as a Practical Bispectrum Measure for Galaxy Redshift Surveys. 3 2026.

Elbers, W. et al. Constraints on neutrino physics from DESI DR2 BAO and DR1 full shape. *Phys. Rev. D*, 112(8):083513, 2025. doi: 10.1103/w9pk-xsk7.

Hand, N., Feng, Y., Beutler, F., Li, Y., Modi, C., Seljak, U., and Slepian, Z. nbodykit: an open-source, massively parallel toolkit for large-scale structure. *Astron. J.*, 156(4):160, 2018. doi: 10.3847/1538-3881/aadae0.

Heavens, A. F. and Taylor, A. N. A Spherical Harmonic Analysis of Redshift Space. *Mon. Not. Roy. Astron. Soc.*, 275:483–497, 1995. doi: 10.1093/mnras/275.2.483.

Lemos, P. et al. Field-level simulation-based inference of galaxy clustering with convolutional neural networks. *Phys. Rev. D*, 109(8):083536, 2024. doi: 10.1103/PhysRevD.109.083536.

Lesgourgues, J. and Pastor, S. Massive neutrinos and cosmology. *Phys. Rept.*, 429:307–379, 2006. doi: 10.1016/j.physrep.2006.04.001.

Lucchin, F. and Matarrese, S. The Effect of nonGaussian statistics on the mass multiplicity of cosmic structures. *Astrophys. J.*, 330:535–544, 1988. doi: 10.1086/166492.

Makinen, T. L., Charnock, T., Alsing, J., and Wandelt, B. D. Lossless, scalable implicit likelihood inference for cosmological fields. *JCAP*, 11(11):049, 2021. doi: 10.1088/1475-7516/2021/11/049. [Erratum: JCAP 04, E02 (2023)].

Makinen, T. L., Sui, C., Wandelt, B. D., Porqueres, N., and Heavens, A. Hybrid summary statistics, 2025. URL https://arxiv.org/abs/2410.07548.

Maldacena and Martin, J. Non-gaussian features of primordial fluctuations in single field inflationary models. *JHEP*, 05:013, 2003. doi: 10.1088/1126-6708/2003/05/013.

Marinucci, M. et al. The constraining power of the marked power spectrum: an analytical study. *JCAP*, 09:036, 2025. doi: 10.1088/1475-7516/2025/09/036.

Massara, E., Villaescusa-Navarro, F., Ho, S., Dalal, N., and Spergel, D. N. Using the Marked Power Spectrum to Detect the Signature of Neutrinos in Large-Scale Structure. *Phys. Rev. Lett.*, 126(1):011301, 2021. doi: 10.1103/PhysRevLett.126.011301.

Matarrese, S. and Verde, L. The effect of primordial non-Gaussianity on halo bias. *Astrophys. J. Lett.*, 677:L77–L80, 2008. doi: 10.1086/587840.

Mellier, Y. et al. Euclid. I. Overview of the Euclid mission. 5 2024.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *arXiv e-prints*, art. arXiv:2003.08934, March 2020. doi: 10.48550/arXiv.2003.08934.

Percival, W. J., Verde, L., and Peacock, J. A. Fourier analysis of luminosity-dependent galaxy clustering. *Mon. Not. Roy. Astron. Soc.*, 347:645, 2004. doi: 10.1111/j.1365-2966.2004.07245.x.

Philcox, O. H. E., Massara, E., and Spergel, D. N. What does the marked power spectrum measure? Insights from perturbation theory. *Phys. Rev. D*, 102(4):043516, August 2020. doi: 10.1103/PhysRevD.102.043516.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. Learning Transferable Visual Models From Natural Language Supervision. *arXiv e-prints*, art. arXiv:2103.00020, February 2021. doi: 10.48550/arXiv.2103.00020.

Scoccimarro, R., Couchman, H. M. P., and Frieman, J. A. The Bispectrum as a Signature of Gravitational Instability in Redshift-Space. *Astrophys. J.*, 517:531–540, 1999. doi: 10.1086/307220.

Sefusatti, E. and Komatsu, E. The bispectrum of galaxies from high-redshift galaxy surveys: primordial non-Gaussianity and non-linear galaxy bias. *Phys. Rev. D*, 76: 083004, 2007. doi: 10.1103/PhysRevD.76.083004.

Tegmark, M. Measuring cosmological parameters with galaxy surveys. *Phys. Rev. Lett.*, 79:3806–3809, 1997. doi: 10.1103/PhysRevLett.79.3806.

van den Oord, A., Li, Y., and Vinyals, O. Representation Learning with Contrastive Predictive Coding. *arXiv e-prints*, art. arXiv:1807.03748, July 2018. doi: 10.48550/arXiv.1807.03748.

Verde, L., Wang, L.-M., Heavens, A., and Kamionkowski, M. Large scale structure, the cosmic microwave background, and primordial non-gaussianity. *Mon.Not.Roy.Astron.Soc.*, 313:L141, 2000. doi: 10.1046/j.1365-8711.2000.03191.x.

Villaescusa-Navarro et al. Ingredients for 21 cm intensity mapping. *The Astrophysical Journal*, 866(2):135, oct 2018. ISSN 1538-4357. doi: 10.3847/1538-4357/aadba0.

White, M. A marked correlation function for constraining modified gravity models. *JCAP*, 11:057, 2016. doi: 10.1088/1475-7516/2016/11/057.

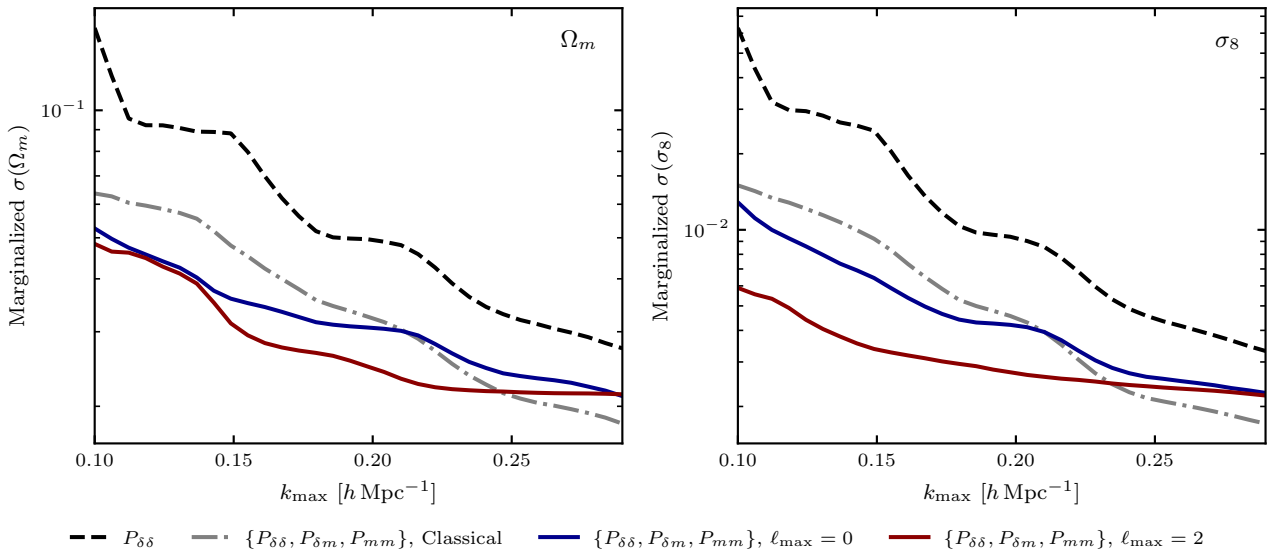


Figure 6. Marginalized Ω_m (left) and σ_8 (right) constraints as a function of $k_{\max}$, for $P_{\delta\delta}$ (black), a reference classical mark (gray), the isotropic ablation $\ell_{\max} = 0$ (blue), and the full learned mark $\ell_{\max} = 2$ (red). The Fisher matrix is evaluated for the full set of summaries $\{P_{\delta\delta}, P_{\delta\Delta}, P_{\Delta\Delta}\}$. The learned marks converge toward the classical baseline as $k_{\max}$ grows.

A. Dependence on $k_{\max}$ and Channel Ablation

In this appendix, we evaluate the robustness of the learned mark to the choice of $k_{\max}$ and to explore the contribution of the different ℓ-channels to the final summary. In Figure 6, we show the marginalized Fisher constraints on Ω_m and σ_8 as a function of $k_{\max}$. As discussed in the main text, the gain of the learned mark is strongest at $k_{\max} \simeq 0.2\,h\,\mathrm{Mpc}^{-1}$, where it significantly outperforms both the unmarked power spectrum and standard marking schemes. At $k_{\max} \sim 0.3\,h\,\mathrm{Mpc}^{-1}$ the classical marks retain the improvement over $P_{\delta\delta}$ while the gain of the learned mark is reduced by the filter taper. On the same configuration grid of Figure 4, at $k_{\max} = 0.3\,h\,\mathrm{Mpc}^{-1}$, the marginalized ratio $\sigma_{\mathrm{class}}/\sigma_{\mathrm{learned}}$ spans a range of $[0.76, 1.1]$ for Ω_m and $[0.72, 1.18]$ for σ_8.

We further explore the contribution of the ℓ-channels by comparing the learned mark to an isotropic $\ell_{\max} = 0$ setting. Keeping the same evaluation pipeline at $k_{\max} = 0.2\,h\,\mathrm{Mpc}^{-1}$, on the grid of classical marks the isotropic ablation recovers a significant fraction of the MSE ratio in the generalization task, with $\mathrm{MSE}_{\mathrm{class}}/\mathrm{MSE}_{\mathrm{learned}} \in [0.91, 1.33]$ for Ω_m and $\in [0.83, 1.34]$ for σ_8. In the Fisher forecast at the fiducial cosmology, the isotropic ablation recovers $\sigma_{\ell_{\max}=2}/\sigma_{\ell_{\max}=0} \sim 0.81$ on Ω_m and ~ 0.65 on σ_8 at $k_{\max} \simeq 0.2\,h\,\mathrm{Mpc}^{-1}$. A modulation of the local density is therefore responsible for most of the gain, while the anisotropic channels and their cross-interactions provide a refinement that further tightens the final constraining power of the learned mark. At smaller scales, the effect of the anisotropic channels is reduced by resolution and smoothing effects, and the mark behaves as an effective isotropic modulation of density.

B. Hyperparameters and Model Structure

This appendix collects the hyperparameters used to train the results reported in the main text. A full training run of the model takes ≈ 3 hours on a single NVIDIA GH200 GPU.

B.1. Optimization and Contrastive Loss

The marker network and the summary and parameter embedders are trained jointly with AdamW under the schedule of Table 1. A short bootstrap phase first pretrains the parameter and $P_{\delta\delta}$ projectors under the same contrastive loss (Section 2.3); the $P_{\delta\delta}$ projector is then frozen for the main run so that it remains a stable residualization reference. The objective is an InfoNCE variant that projects the mark summaries, the unmarked $P_{\delta\delta}$ summary, and the parameters $\boldsymbol{\theta}$ into a shared embedding space, with explicit residualization of the mark embedding against the $P_{\delta\delta}$ embedding. Its projector geometry and negative-sampling scheme are listed in Table 2.

Hyperparameter	Value
Optimizer	AdamW
Base learning rate	5×10^{-4}
Weight decay	10^{-4}
Gradient-clip norm	10.0
Batch size	16
Mark-parameters LR mult.	2.0
θ-proj. LR mult.	0.25
LR scheduler	plateau (γ=0.5)
patience / floor	8 / 10^{-7}
Early-stop patience	20 ep. (Δ=10^{-5})
Validation fraction	0.15

Table 1. Optimization hyperparameters for the contrastive training run.

Hyperparameter	Value
Embedding dimension	16
Mark / $P_{\delta\delta}$ proj.	$(\texttt{Lin LN GELU})^{\times 2}\texttt{Lin}$
Projector hidden width	64
θ projector	linear
Temperature τ	0.05
k-modes	47 $\left(k_{\max}=0.30\,h\,\mathrm{Mpc}^{-1}\right)$
Real negatives / batch	128
Synthetic negatives / batch	384 / anchor
Synthetic close negatives / batch	32 / anchor
Neighborhood scale	0.15
Global exclusion scale	0.03

Table 2. InfoNCE loss and projector heads. Neighborhood and exclusion scales are in normalized θ.

B.2. Network Architectures

The marker module evaluates the filtered fields of Equation (5) in Fourier space, producing 9 channels that are reduced to the rotation invariants $\{E_0, E_1, E_2, I_3\}$, non-linearly processed through independent MLPs and summed together with an additional cross term. Its structure is summarized in Table 3. The held-out parameter recovery reported in the main text uses a symmetric-pyramid MLP regressor that maps the marked-summary spectra $\{P_{\delta\delta}(k), P_{\Delta\Delta}(k), P_{\delta\Delta}(k)\}$ to the cosmological parameters $\boldsymbol{\theta}$. The raw spectra are first processed into three physically motivated channels per k-bin, concatenated into a length-$3N_k$ vector, and fed into the pyramid body, whose architecture settings are listed in Table 4. To train the MLP for parameter regrssion, we use AdamW with Plateau LR scheduler, with early stopping based on the validation loss.

Hyperparameter	Value
SH channels	9 (ℓ=0, 1, 2)
Radial parameterisation	per-ℓ Fourier-feature MLP
Radial MLP width	16
Fourier-feature frequencies	8
Radial seed $r_0/r_{\max}$	0.3
Taper $k_{\mathrm{pass}}/k_{\mathrm{Nyq}}$	0.85
Taper floor	0
Input transform	$\mathrm{sgn}(x)\log(1+x)$
Per-invariant MLP	`Lin GELU Lin`
MLP width	16
Cross-term width	8

Table 3. Marker module. Per-invariant MLPs use zero-output initialization to set $M \equiv 1$ as first mark state.

Hyperparameter	Value
Input features	$\log_{10} P_{\delta\delta},\ \log_{10}(P_{\Delta\Delta}/P_{\delta\delta}),\ r_k$
Base width B	$3N_k$
Hidden widths	$[B, 2B, 4B, 4B, 2B, B]$
Block	`Lin LN GELU Drop`
Dropout	0.1
Loss	$\sum_d \log(\mathrm{MSE}_d + \epsilon)$

Table 4. Symmetric-pyramid MLP regressor for held-out parameter recovery.