JUN 16 2026 · THE NON-GAUSSIAN UNIVERSE · FORTH, HERAKLION

Can we trust higher-order weak lensing?

Baryonic robustness, and learned vs analytical summaries

Andreas Tersenov · FORTH · U. Crete · CEA Paris-Saclay · slides: andreastersenov.github.io/talks/

§0 We are all optimizing statistics; the two-point camp still does not trust the contours

THE 2-POINT CAMP

"I don't believe any of your contours."

WHAT WOULD MAKE HOS FLAGSHIP-GRADE?

- blinding
- systematics
- method limits
- robust covariance
- analytical cross-checks
- null / validation tests
- emulators
- non-Gaussian likelihood
- simplicity

PART 1 OF 2

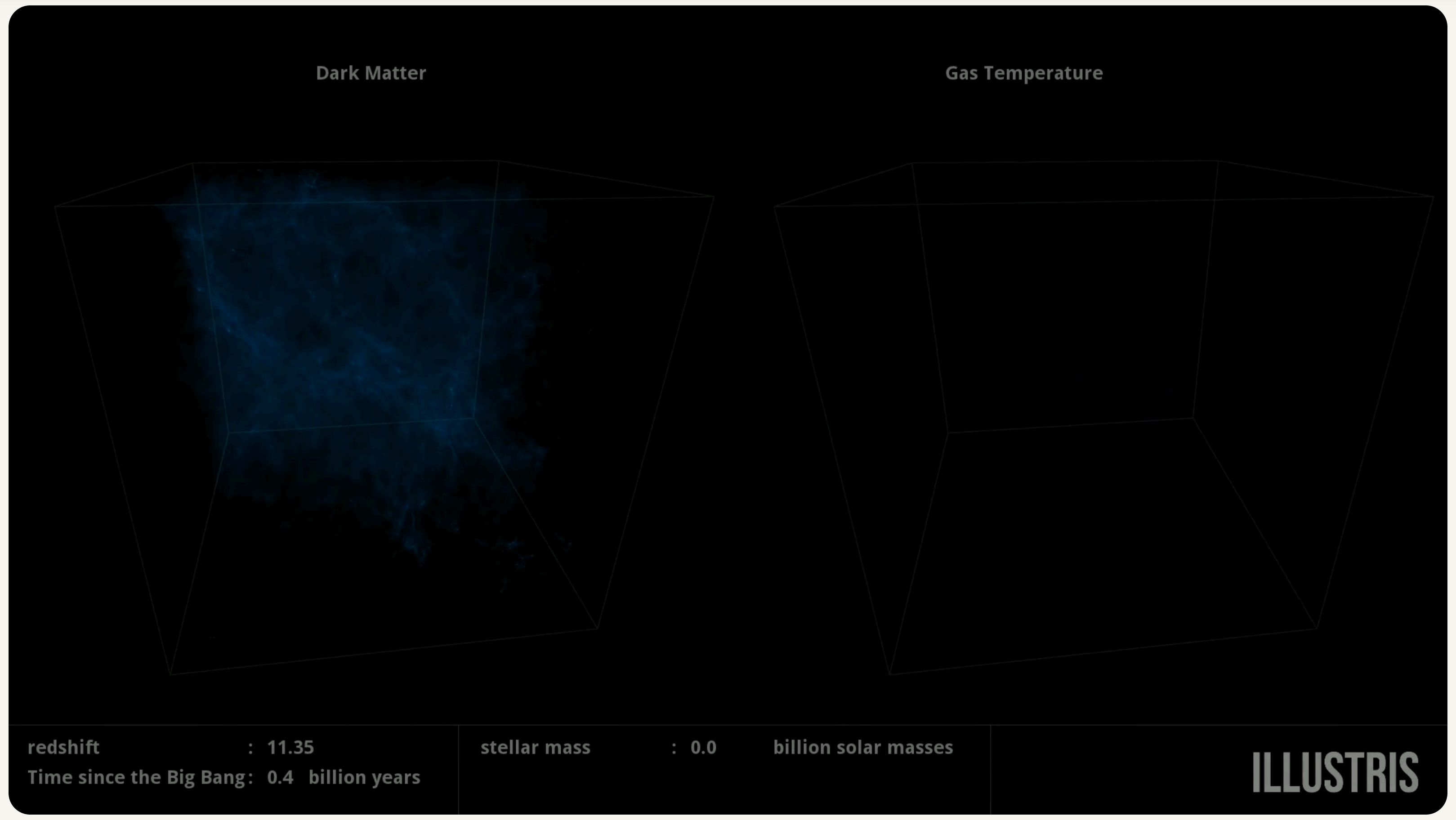
Do baryons break HOS?

Baryonic feedback, the wavelet ℓ1-norm, and the BNT transform.

§1 Stage IV is no longer statistics-limited, it is systematics-limited

TO TRUST A STATISTIC

Before we trust any summary statistic, we have to quantify how each systematic affects it, and at the **contour level** (the inferred parameters).



Illustris: baryonic feedback reshaping the cosmic web

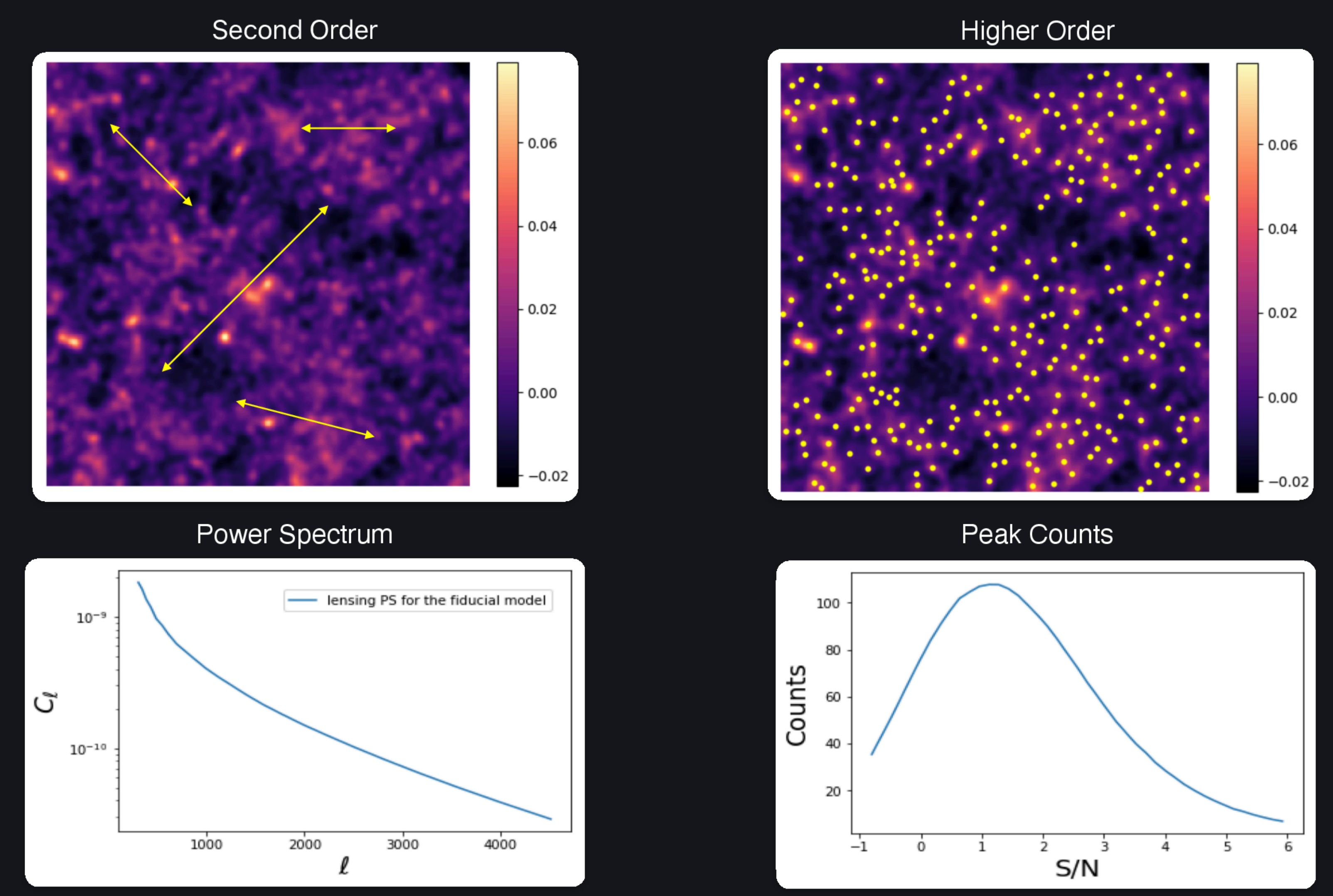
THE SYSTEMATIC AT HAND

Baryonic feedback (AGN, supernovae) suppresses matter on small scales, mimicking cosmological signal and biasing inference, exactly where the constraining power lives and where the feedback models disagree most.

CORE QUESTIONS

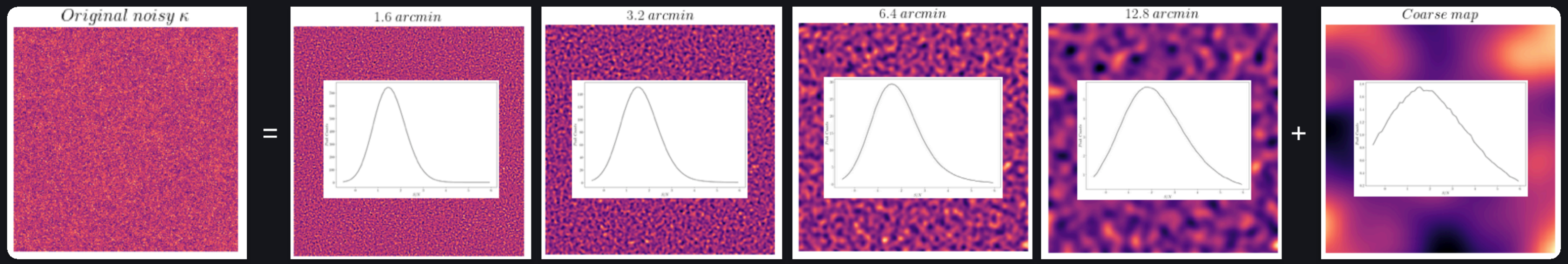
1. How does unmodeled baryonic feedback **bias our non-Gaussian statistics**?
2. After safe scale cuts, do HOS still **outperform the power spectrum**?

§1 Higher-order statistic I: peak counts



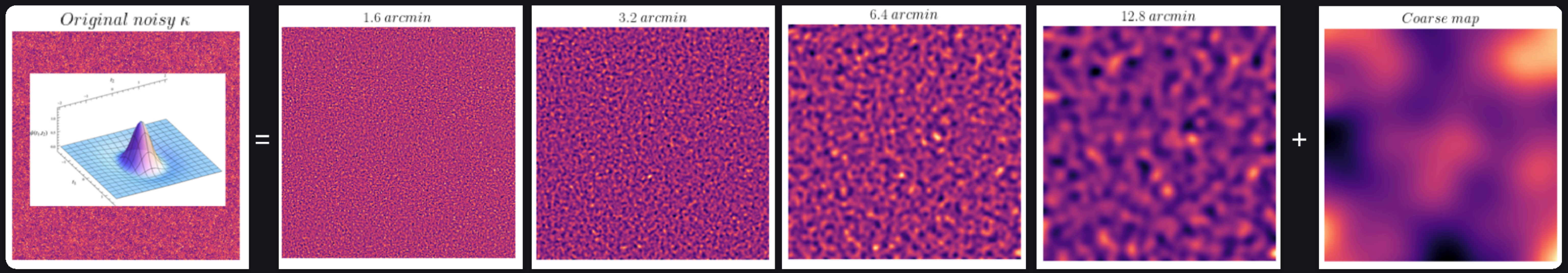
- **Peaks**: local maxima of the SNR field $\nu = (\mathcal{W} * \kappa)(\theta_{\mathrm{ker}}) \,/\, \sigma_n^{\mathrm{filt}}$
- Counted per SNR bin; they trace massive structures
- Simple and well-established, but uses only the high-SNR peaks

§1 Wavelet peaks: a multi-scale peak count via the starlet transform



- Multi-scale, not single-scale
- **Starlet transform**: a map as a sum of wavelet-coefficient images plus a coarse map
- Processes all scales **simultaneously**, for efficiency
- Each band is a different frequency range, so the peak-count covariance is nearly diagonal

§1 Higher-order statistic II: the starlet $\ell 1$-norm



$$\ell_1^{j,i} = \sum_u |\mathcal{S}_{j,i}[u]| = \|\mathcal{S}_{j,i}\|_1$$

- Sum of **absolute starlet coefficients**, per scale and SNR bin
- A **fast, multi-scale** measure of the full void and peak distribution
- Information in **all** pixels: peaks and voids, the full convergence PDF
- No discrete-feature definition

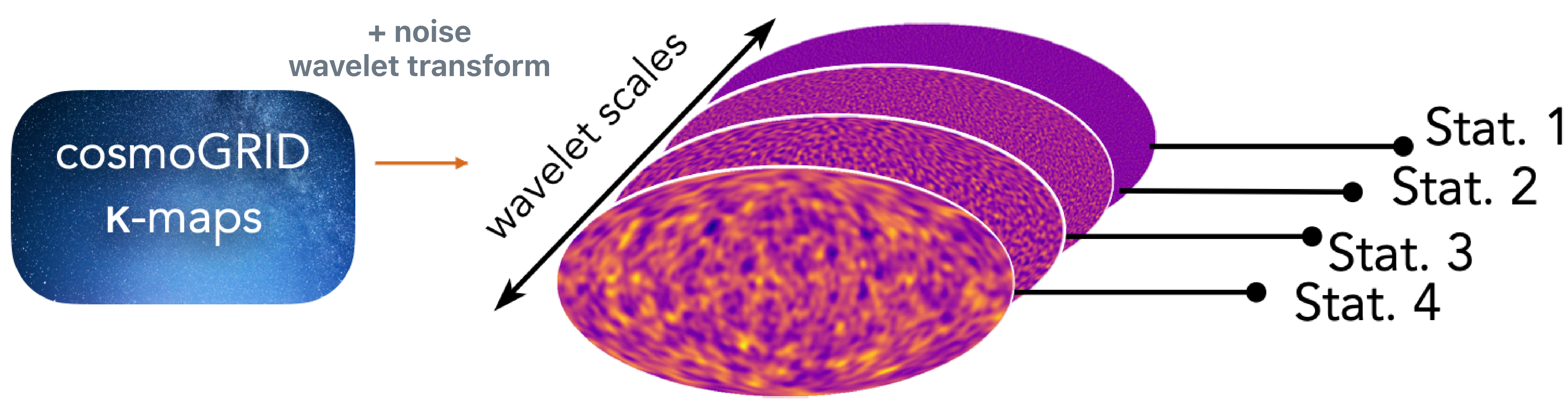
The inference pipeline: neural posterior estimation

replay

1/5 We take **cosmoGRID V1** forward simulations: spherical convergence maps κ at known cosmologies θ.

The inference pipeline: neural posterior estimation

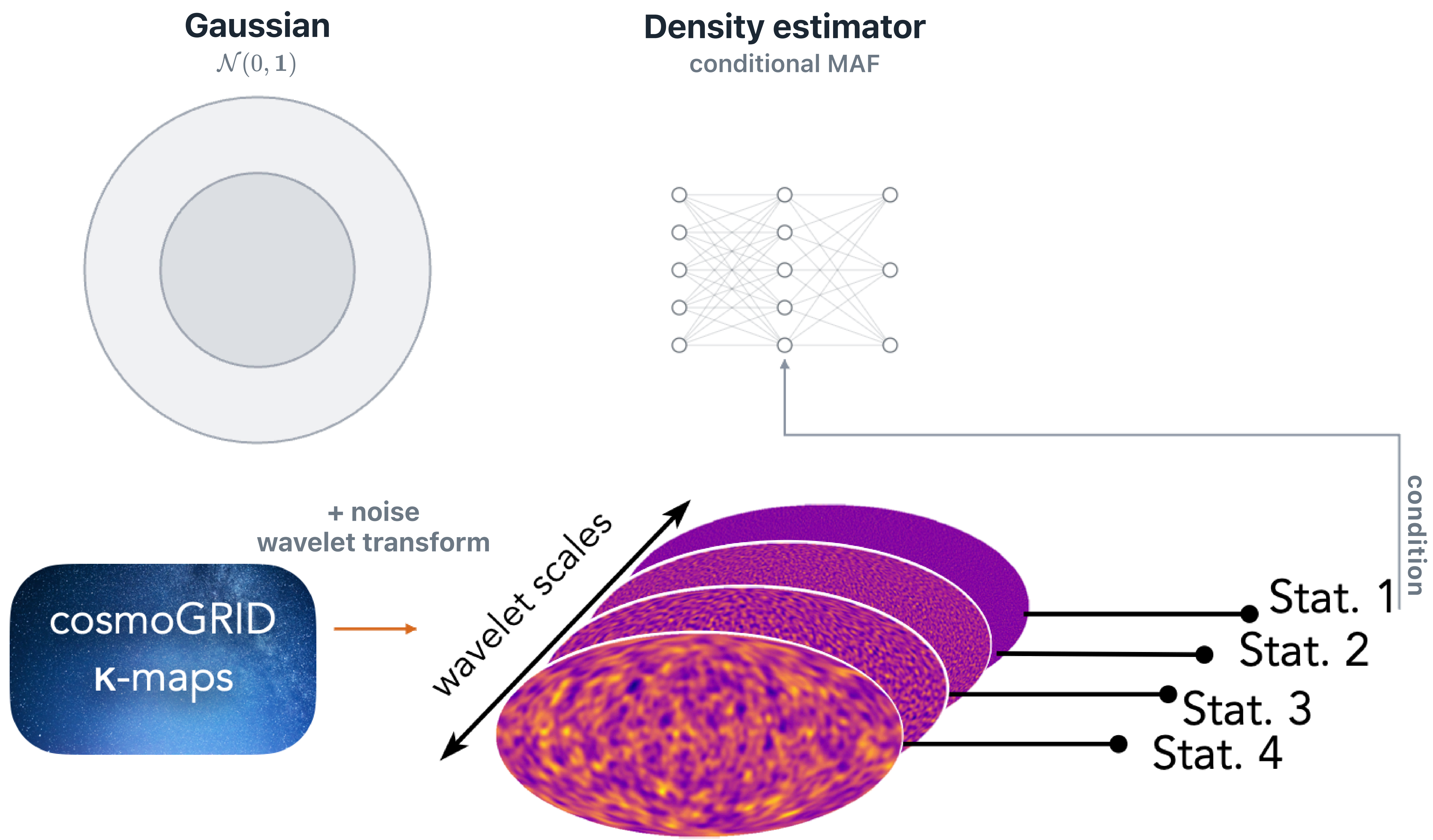
replay



2/5 Add **shape noise** and a **wavelet transform**, then measure **summary statistics** on each scale: the data vector x.

The inference pipeline: neural posterior estimation

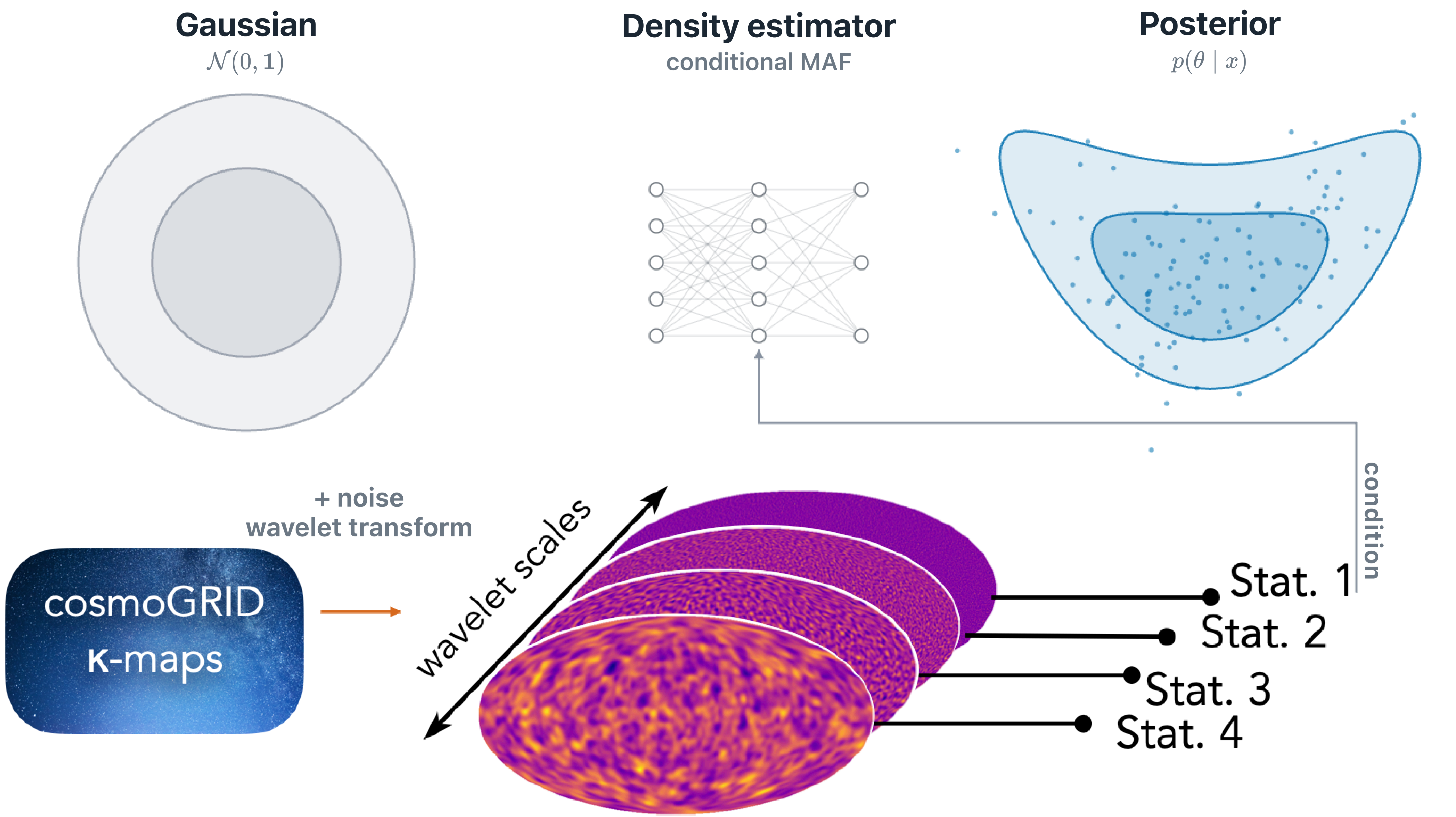
replay



3/5 Those summaries **condition** a neural density estimator, a **conditional MAF**.

The inference pipeline: neural posterior estimation

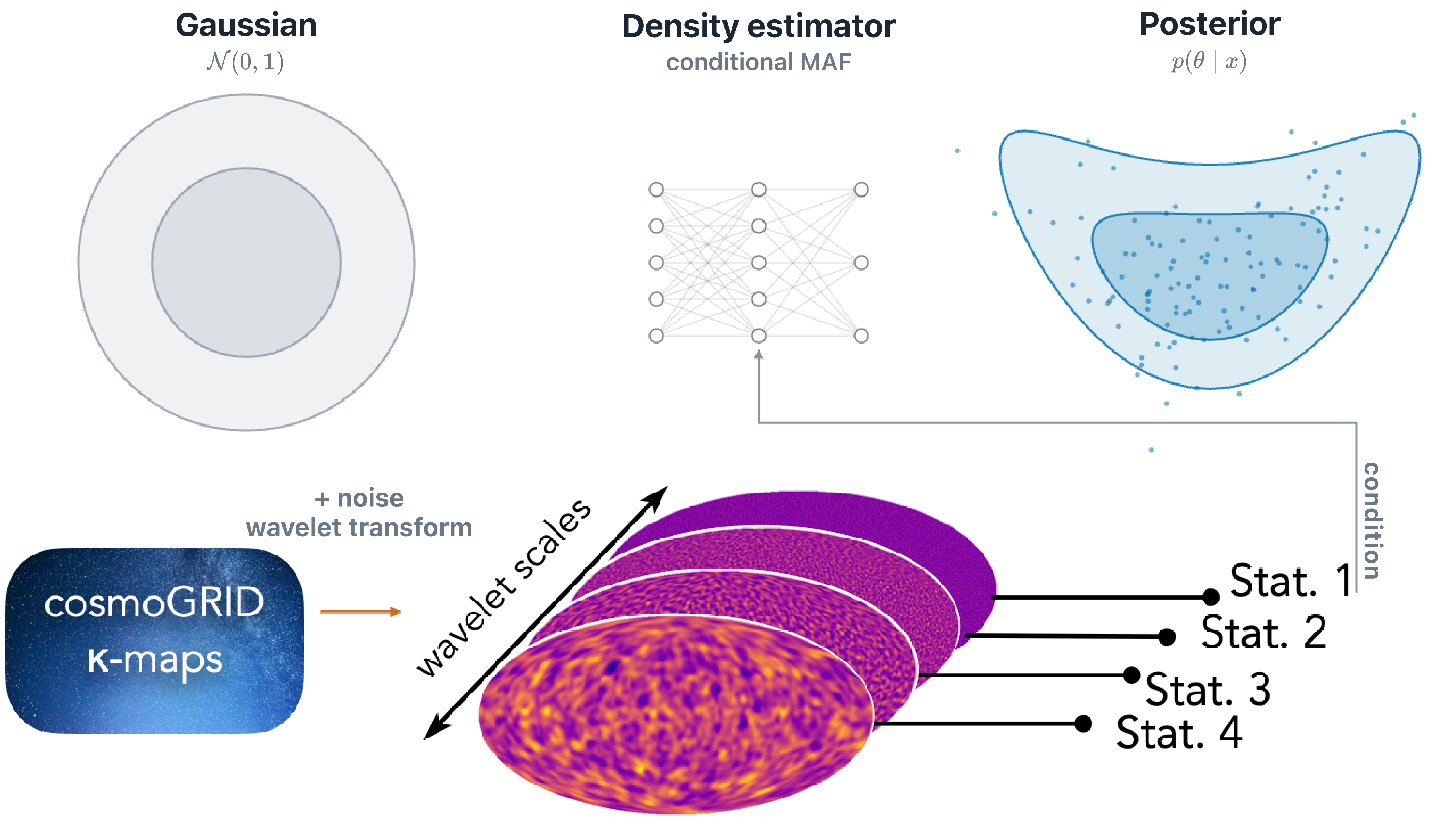
replay



4/5 The flow turns a simple Gaussian $\mathcal{N}(0, 1)$ into the **posterior** $p(\theta \mid x)$.

The inference pipeline: neural posterior estimation

replay



training objective
$\mathcal{L} = -\log p_\phi(\theta \mid x)$

5/5 Trained by maximizing the log-probability of the true parameters: $\mathcal{L} = -\log p_\phi(\theta \mid x)$.

§1 Mitigation: Scale Cuts

§1 Mitigation: Scale Cuts

- Goal: bring tension **below 0.3σ**

§1 Mitigation: Scale Cuts

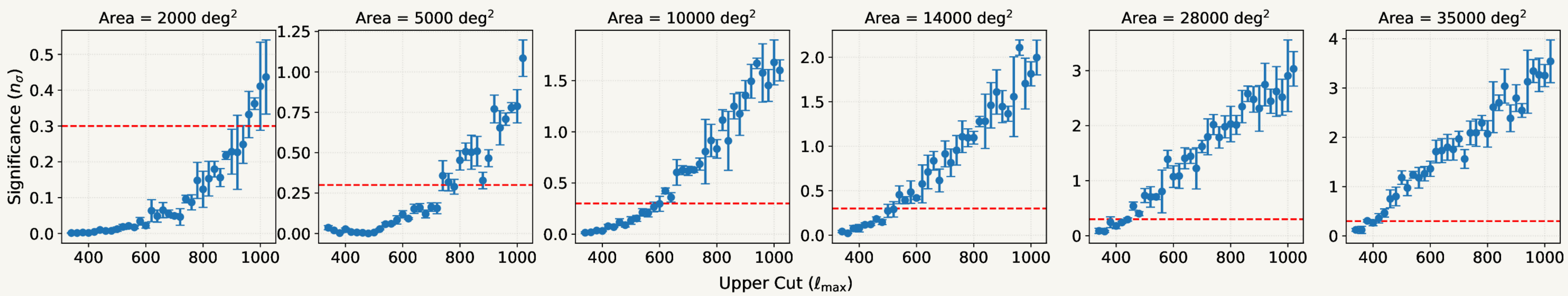
- Goal: bring tension **below 0.3σ**
- C_ℓ: requires an aggressive cut at $\ell \lesssim 400$ → **discards most signal**

§1 Mitigation: Scale Cuts

- Goal: bring tension **below 0.3σ**
- C_ℓ: requires an aggressive cut at $\ell \lesssim 400$ → **discards most signal**
- Starlet: contamination **isolated in one scale** → removing just the finest wavelet band is sufficient across *all* survey areas

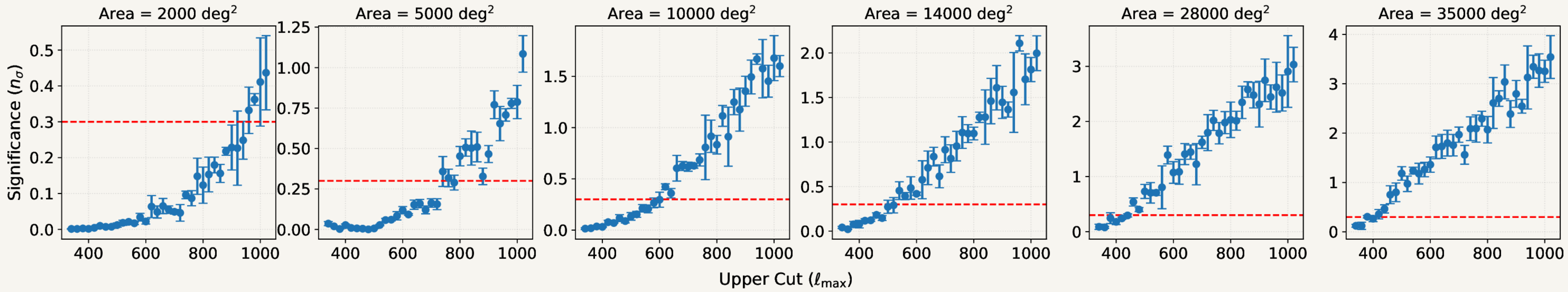
§1 Mitigation: Scale Cuts

- Goal: bring tension **below 0.3σ**
- C_ℓ: requires an aggressive cut at $\ell \lesssim 400$ → **discards most signal**
- Starlet: contamination **isolated in one scale** → removing just the finest wavelet band is sufficient across *all* survey areas



§1 Mitigation: Scale Cuts

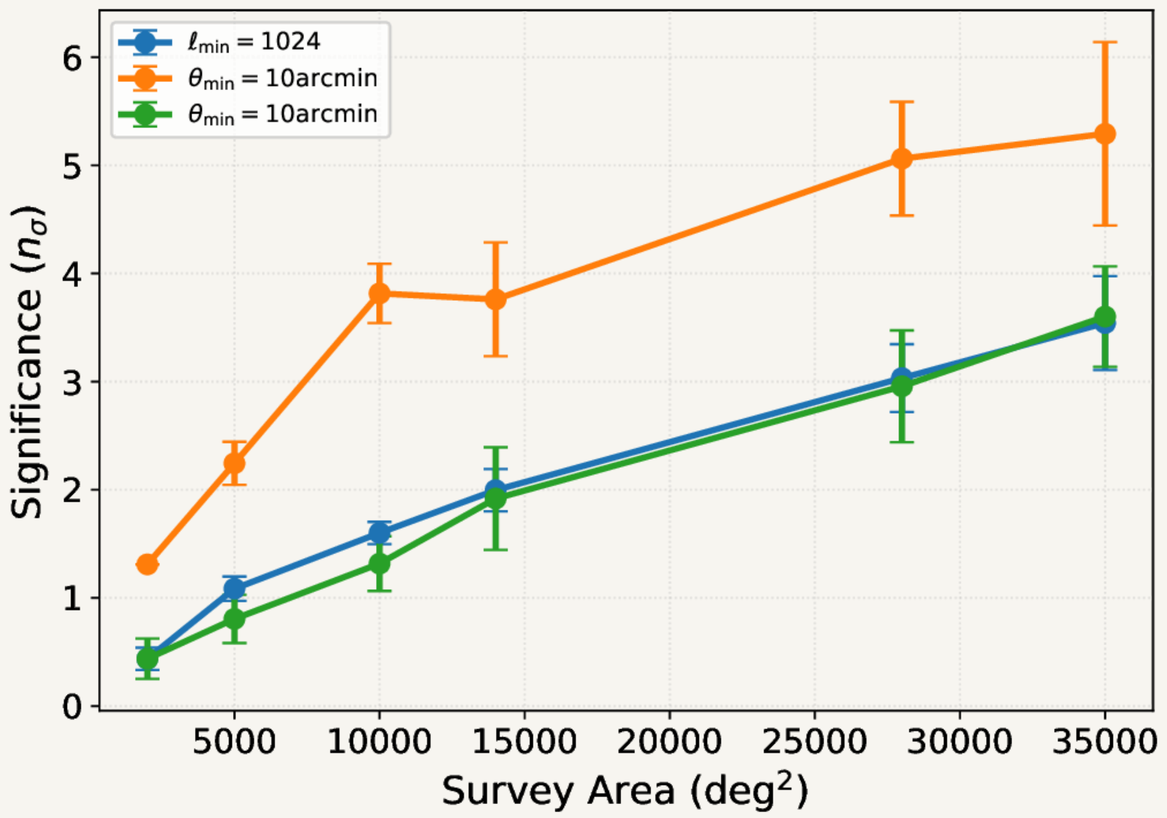
- Goal: bring tension **below 0.3σ**
- C_ℓ: requires an aggressive cut at $\ell \lesssim 400$ → **discards most signal**
- Starlet: contamination **isolated in one scale** → removing just the finest wavelet band is sufficient across *all* survey areas



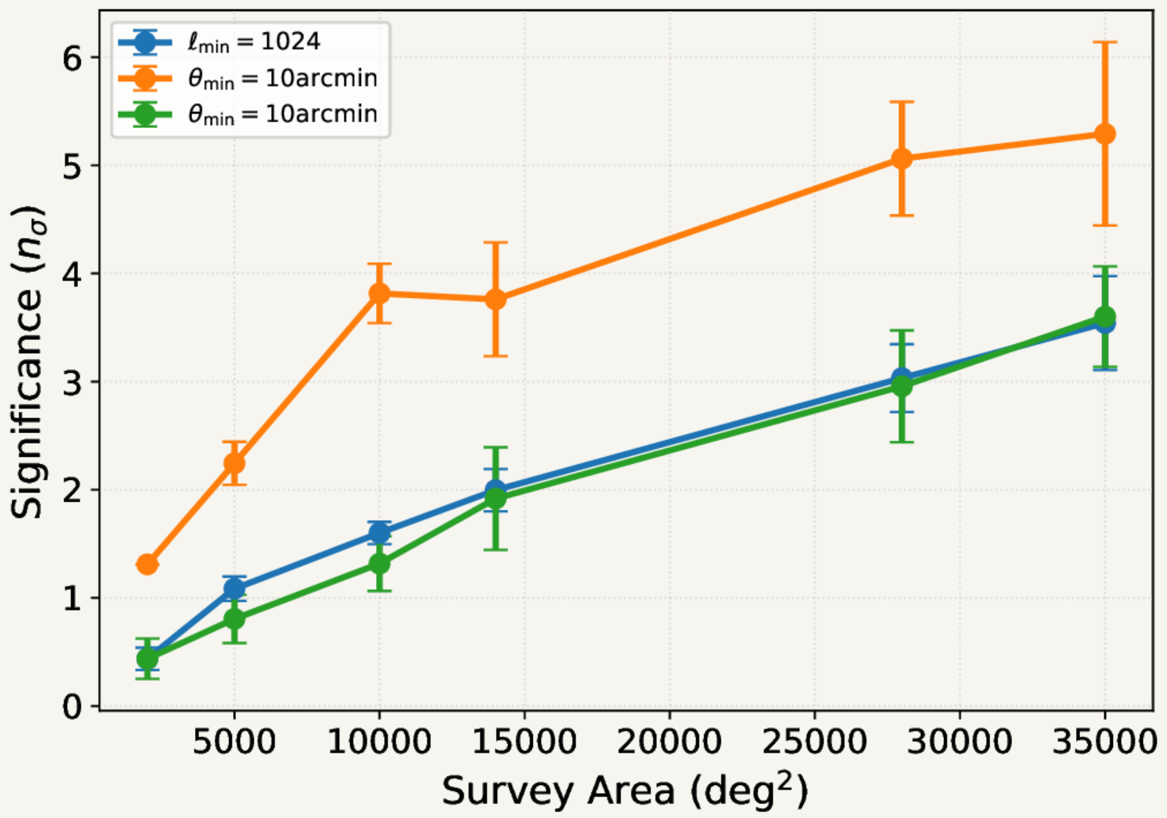
We run this analysis as a function of survey area: more area means smaller statistical errors, and thus higher sensitivity to bias.

§1 Baryonic Bias Scales with Survey Area

§1 Baryonic Bias Scales with Survey Area

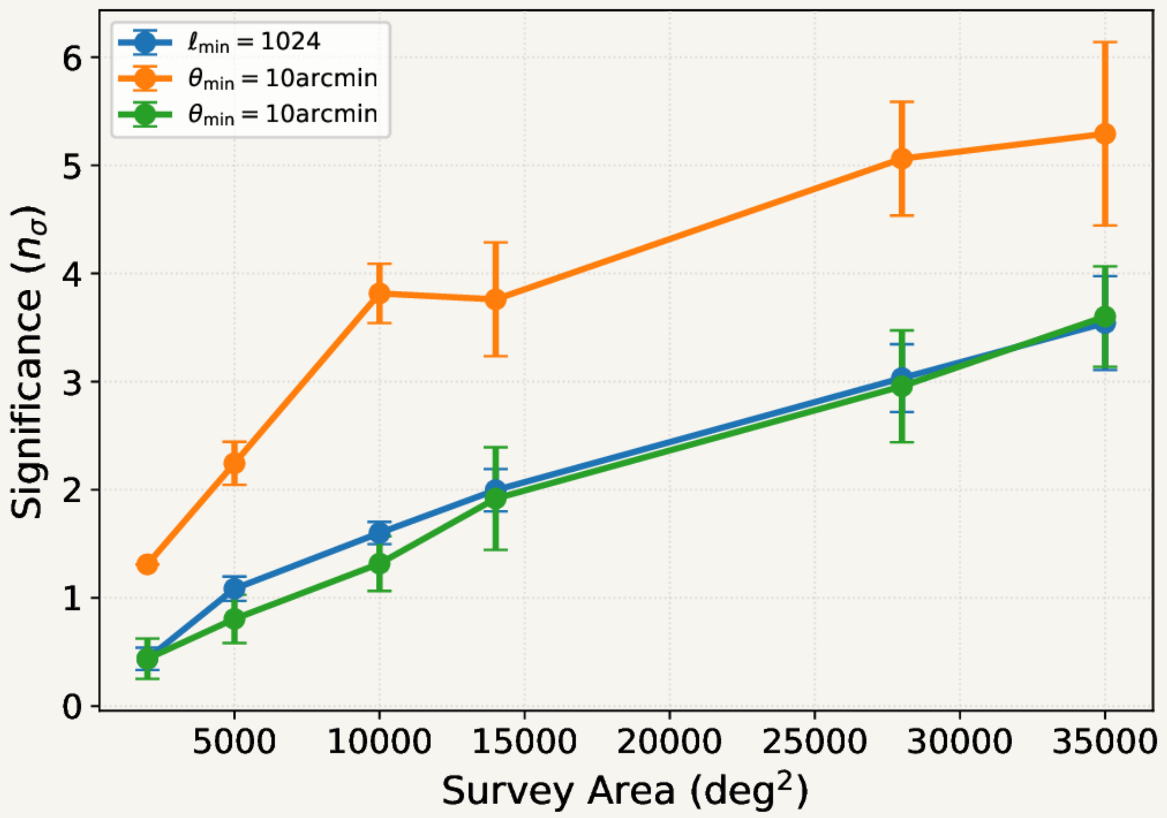


§1 Baryonic Bias Scales with Survey Area



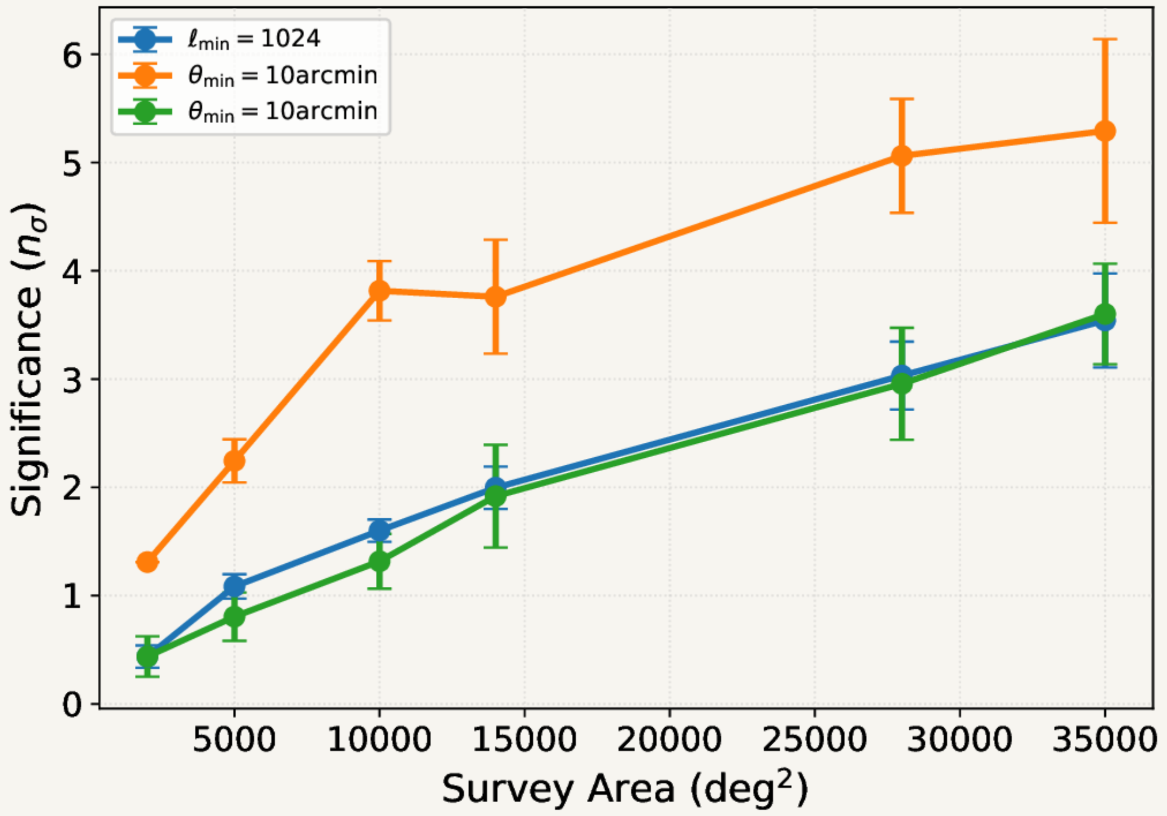
- At **14,000 deg²** (Stage IV): C_ℓ already shows $\sim 2\sigma$ tension

§1 Baryonic Bias Scales with Survey Area



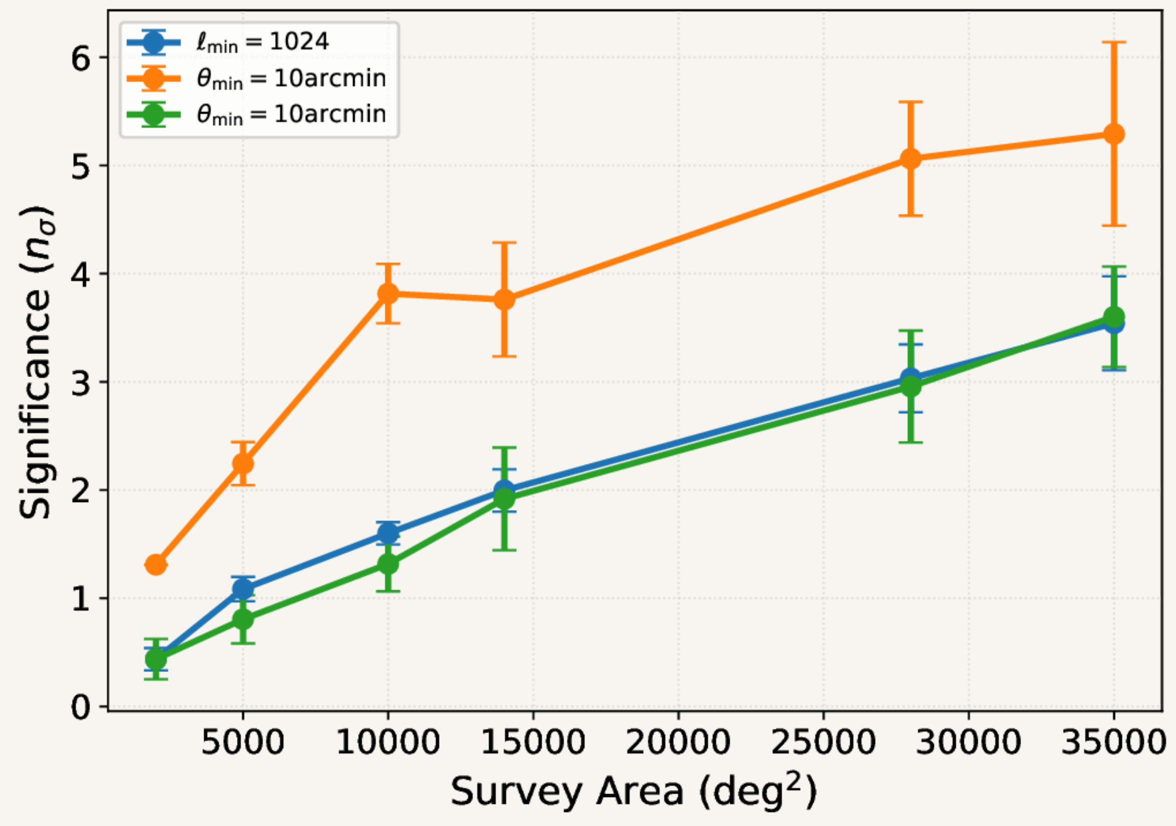
- At **14,000 deg²** (Stage IV): C_ℓ already shows $\sim 2\sigma$ tension
- At **full sky**: tension exceeds 3σ for *all* statistics

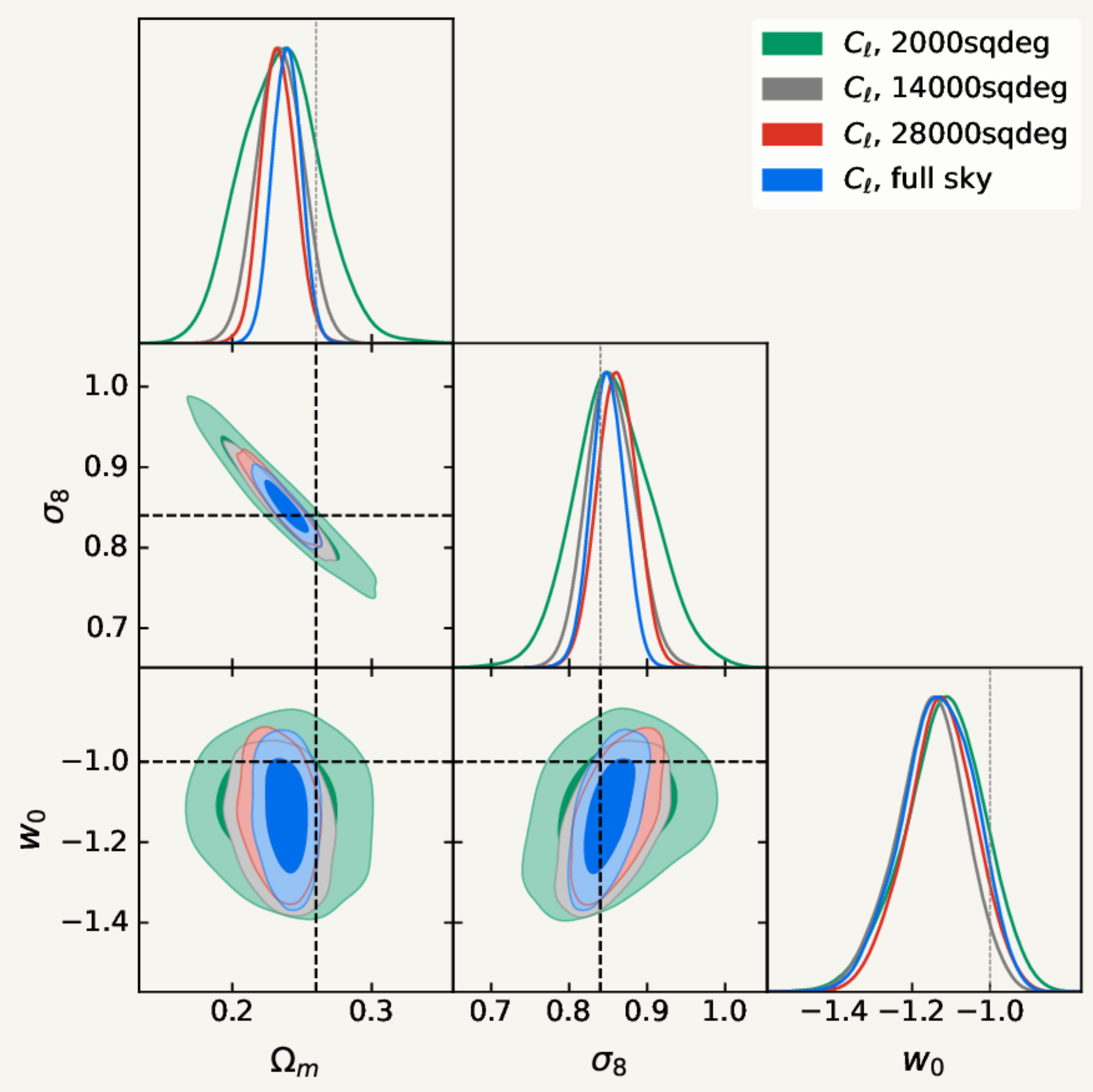
§1 Baryonic Bias Scales with Survey Area



- At **14,000 deg²** (Stage IV): C_ℓ already shows $\sim 2\sigma$ tension
- At **full sky**: tension exceeds 3σ for *all* statistics
- HOS show **higher bias** than C_ℓ: more sensitive to the baryonically-contaminated small scales

§1 Baryonic Bias Scales with Survey Area

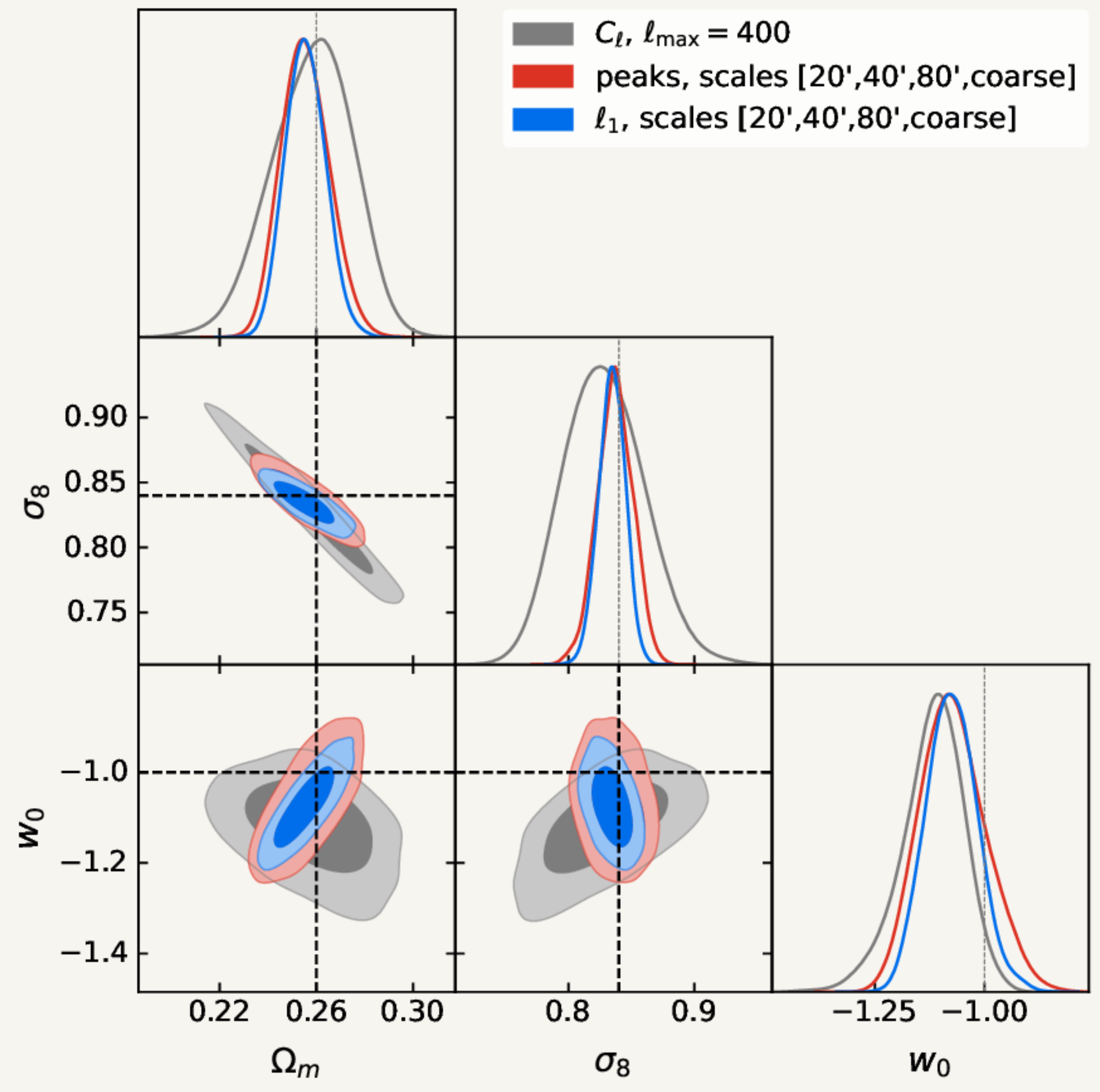




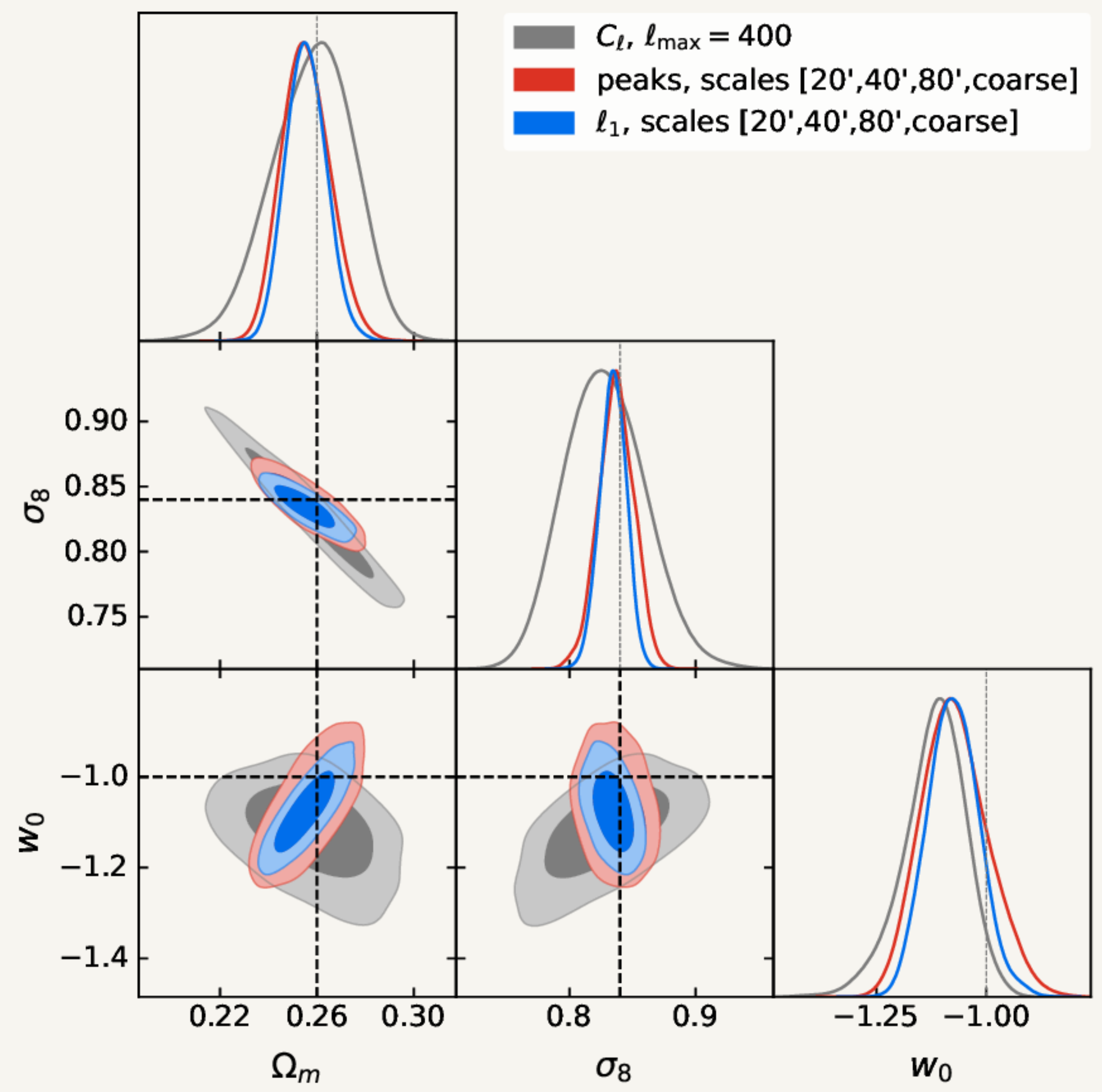
- At **14,000 deg²** (Stage IV): C_ℓ already shows $\sim 2\sigma$ tension
- At **full sky**: tension exceeds 3σ for *all* statistics
- HOS show **higher bias** than C_ℓ: more sensitive to the baryonically-contaminated small scales

§1 Are HOS still useful?

§1 Are HOS still useful?



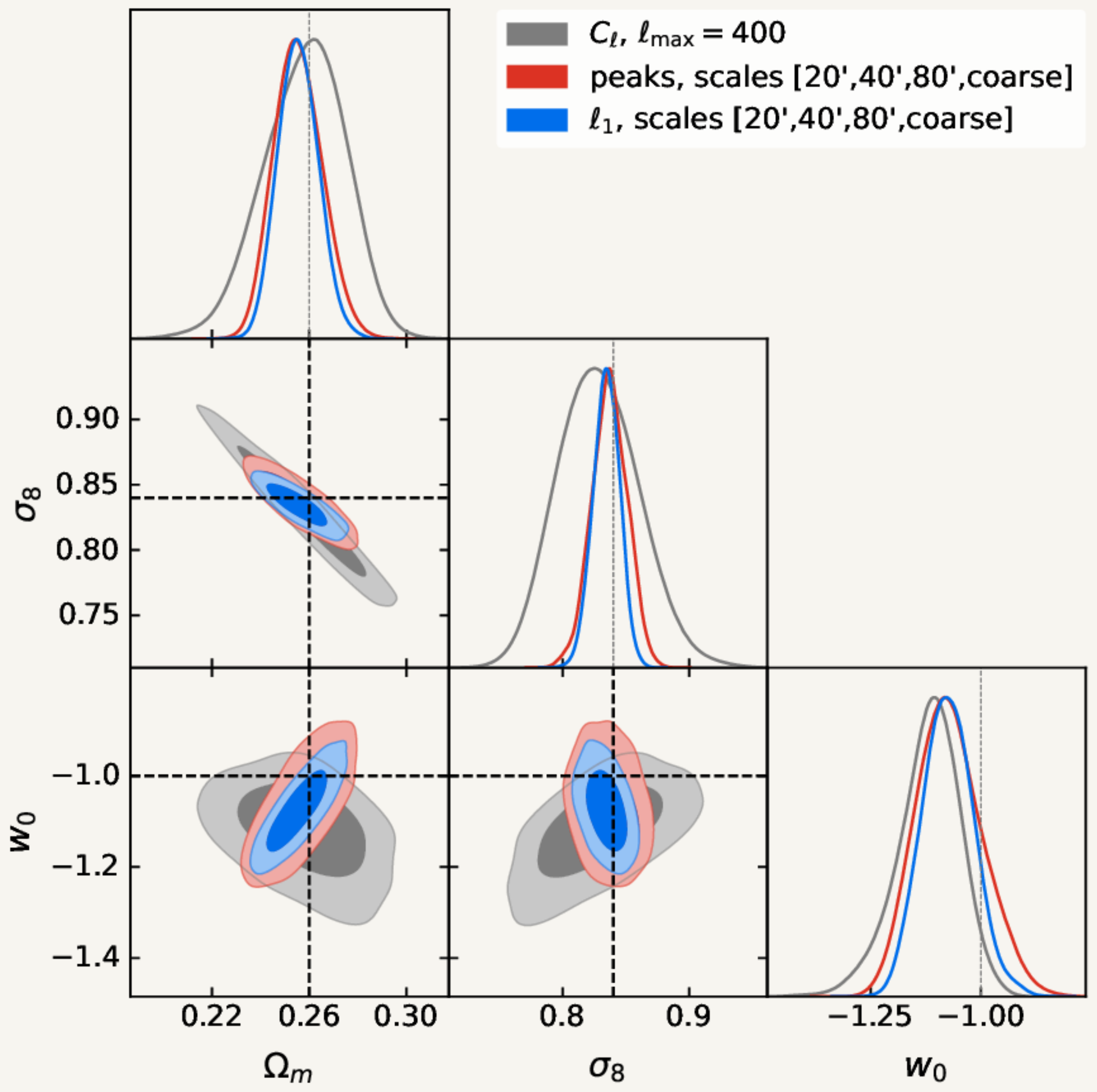
§1 Are HOS still useful?



ON BARYON-SAFE SCALES

- Starlet ℓ_1-norm yields constraints $3\times$ **tighter** than C_ℓ in the full-sky limit
- HOS **still useful**: non-Gaussian signal persists even after the fine-scale cut

§1 Are HOS still useful?



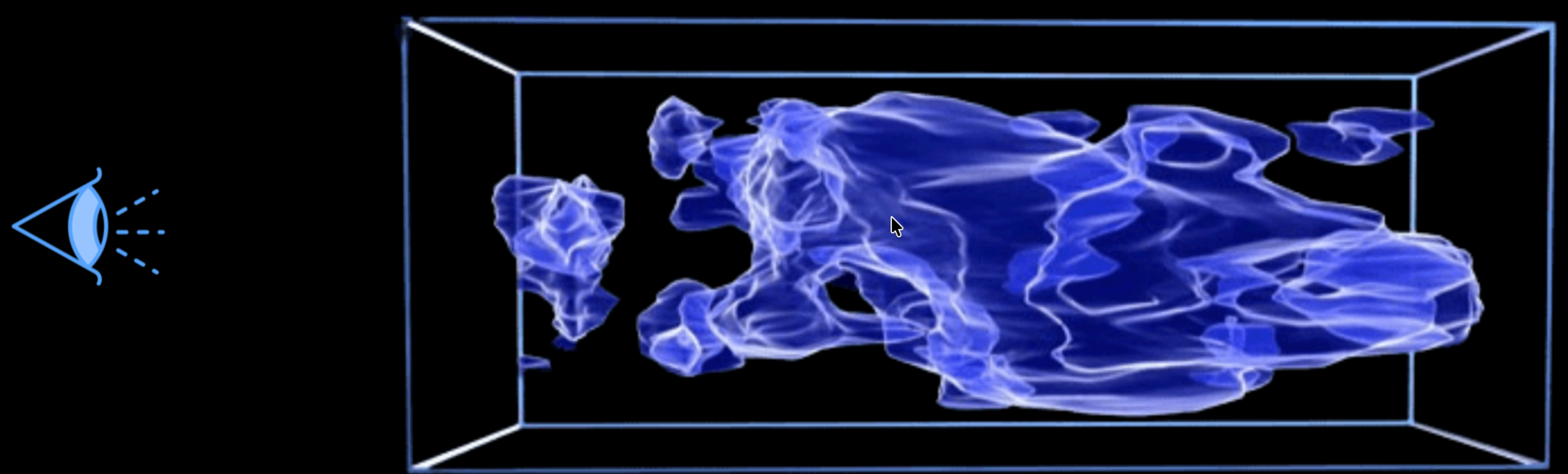
ON BARYON-SAFE SCALES

- Starlet ℓ_1-norm yields constraints $3\times$ **tighter** than C_ℓ in the full-sky limit
- HOS **still useful**: non-Gaussian signal persists even after the fine-scale cut

TAKEAWAY

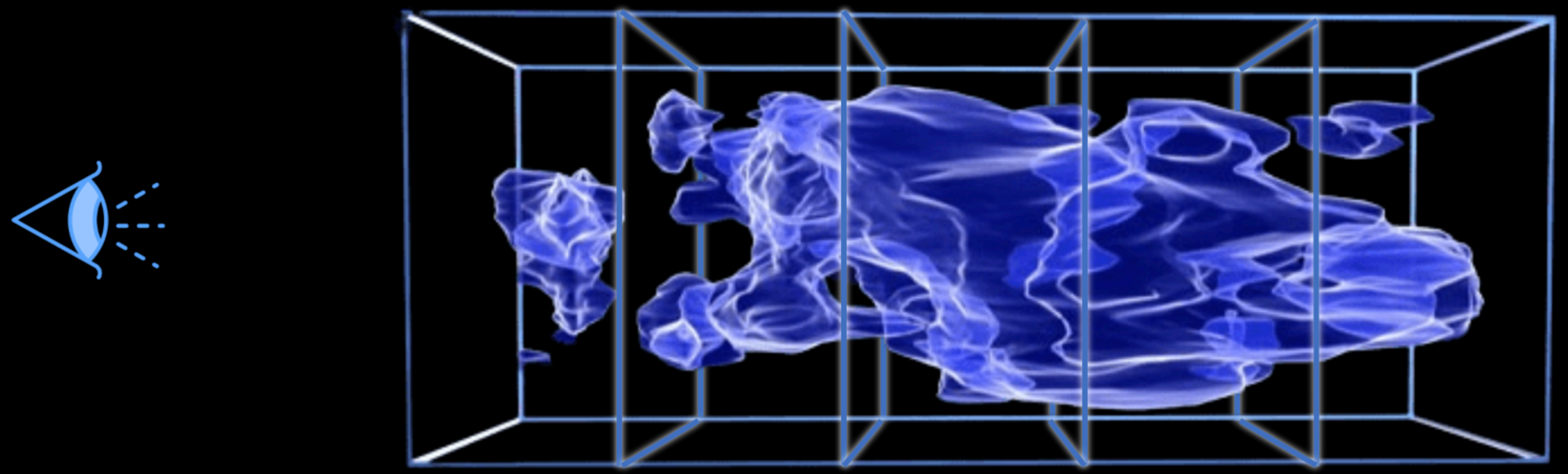
- Baryonic effects are a **dominant systematic**, biasing parameter estimation
- HOS are not merely deep-non-linear probes: they robustly recover information on **quasi-linear scales**

§1 But could we do better than that? Weak lensing tomography



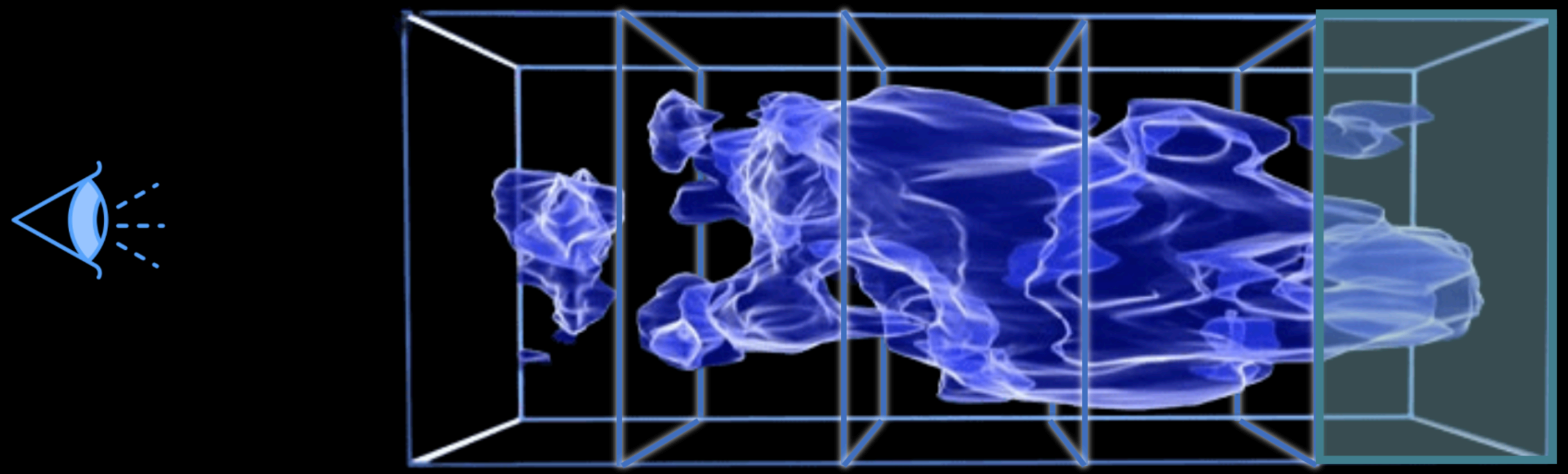
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography



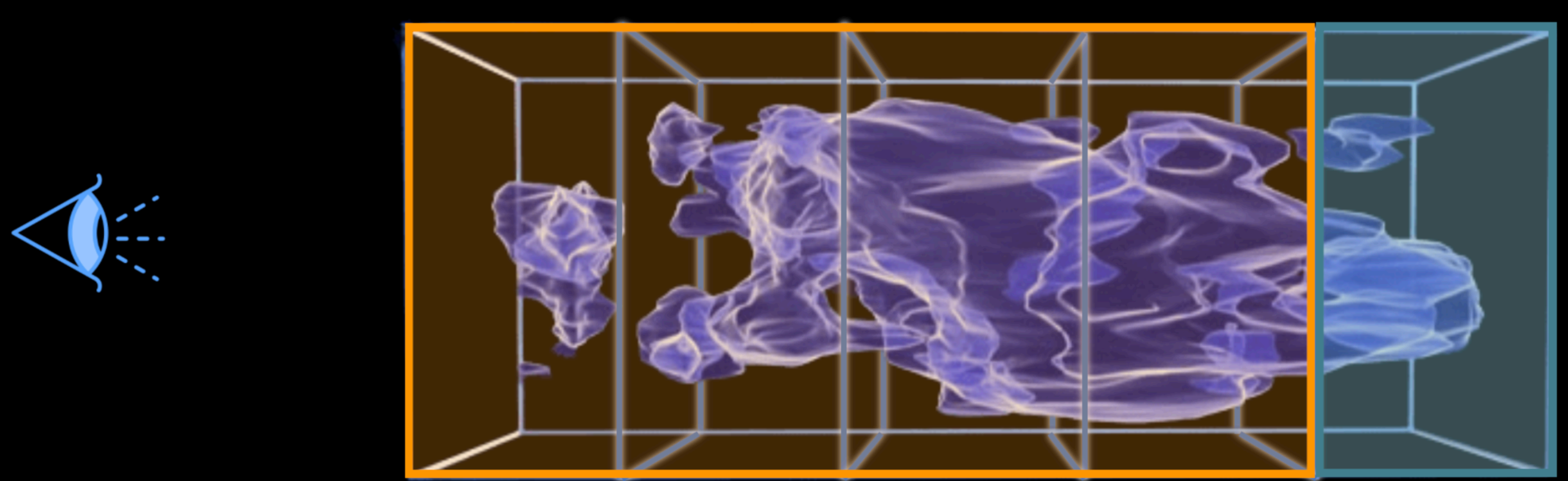
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography



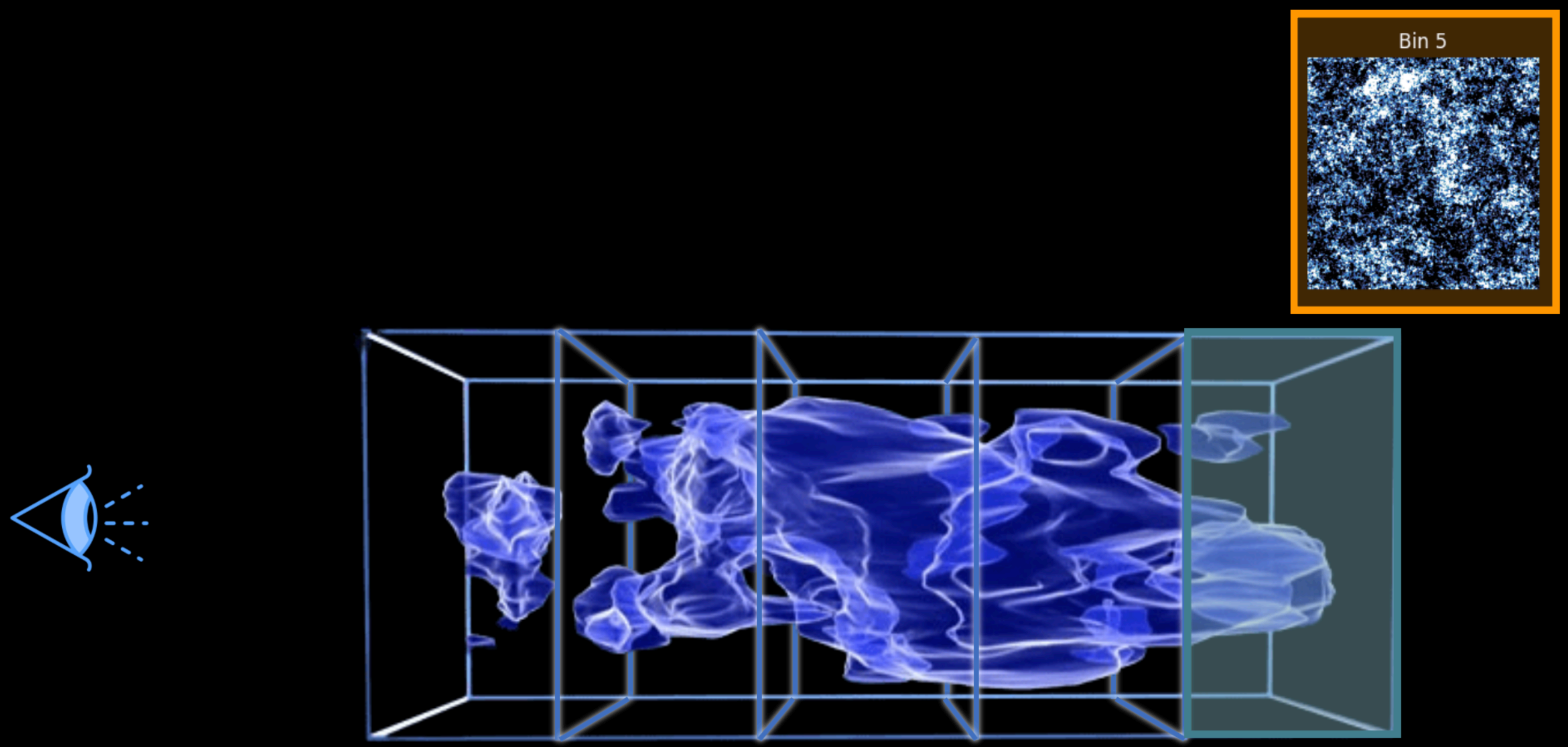
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography



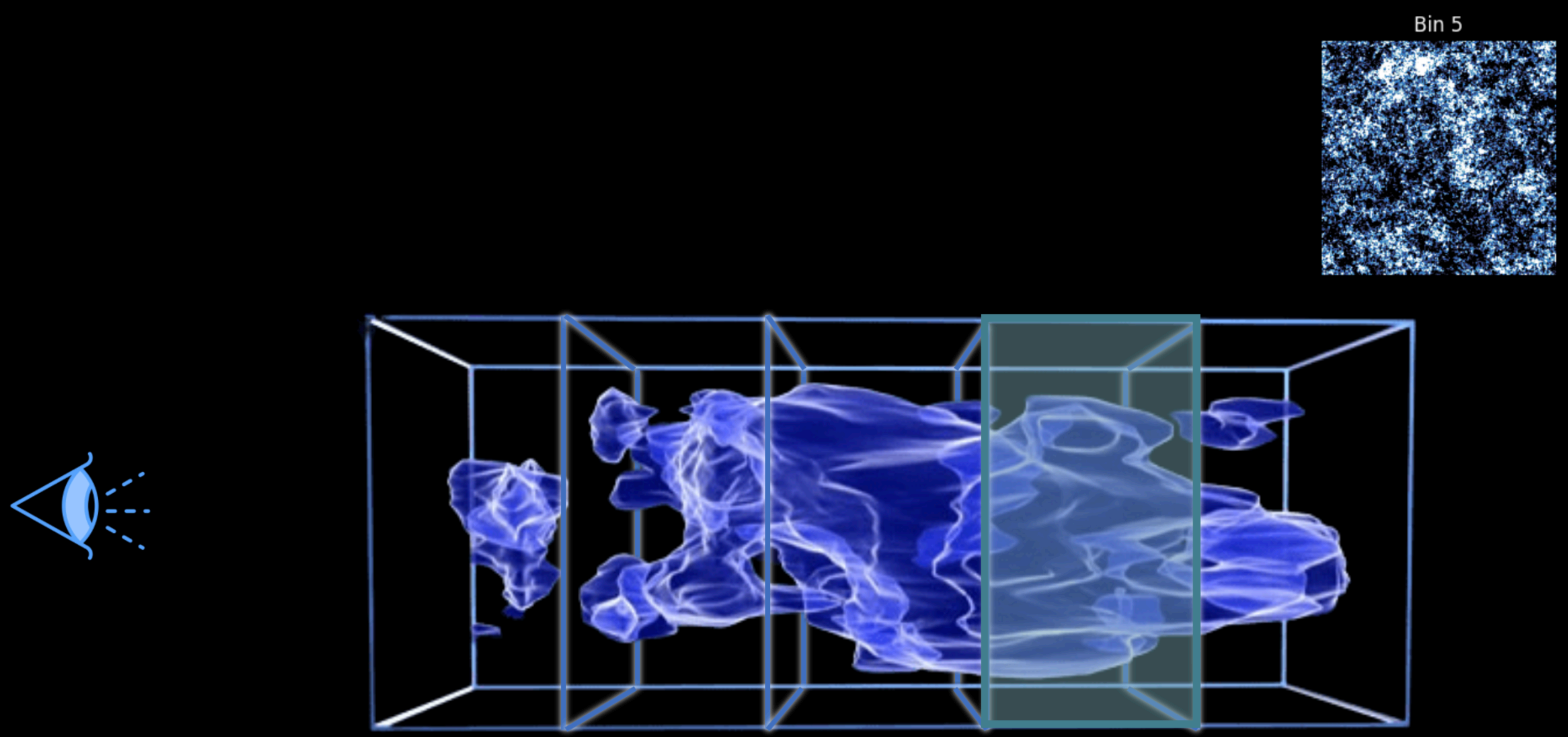
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography



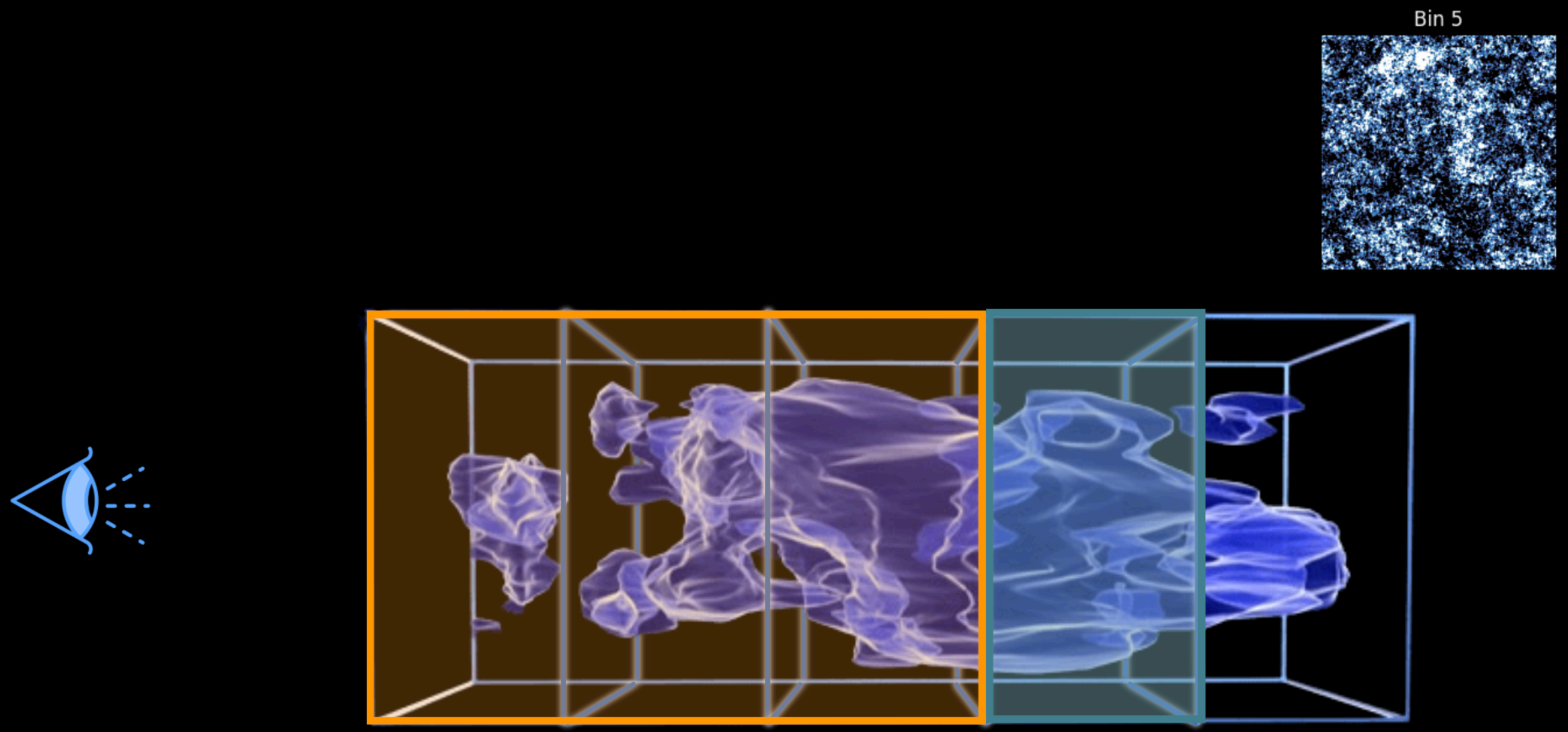
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography



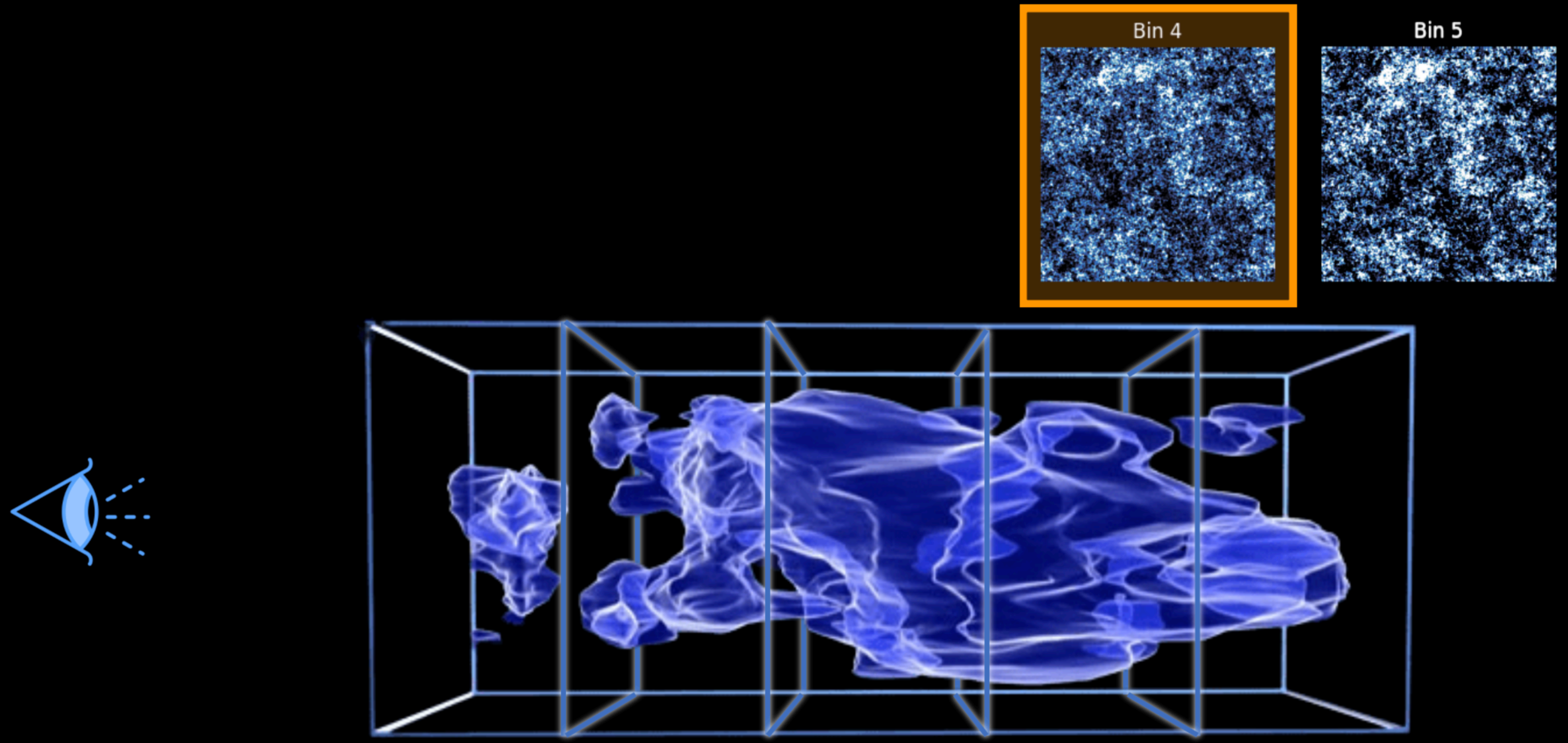
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography



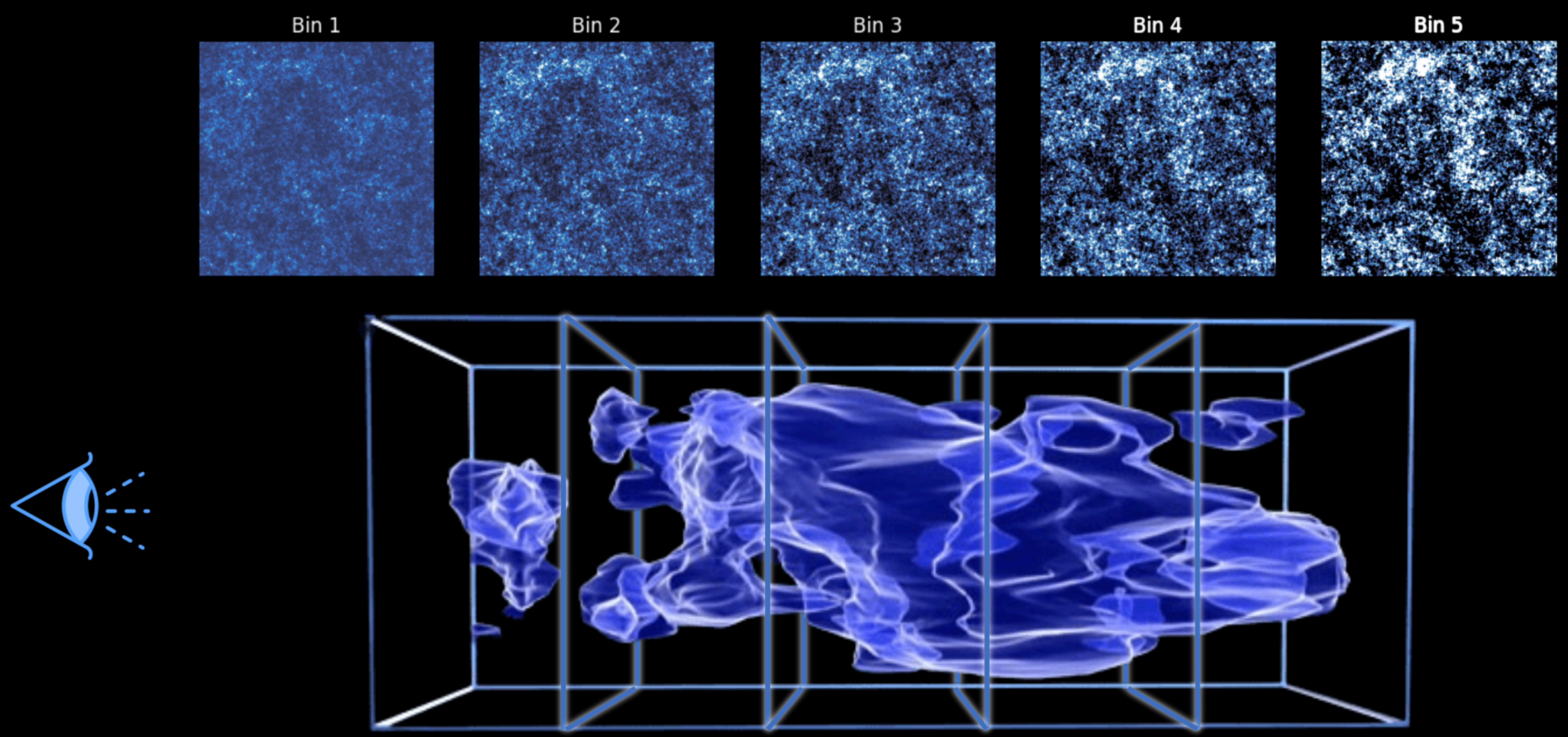
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography



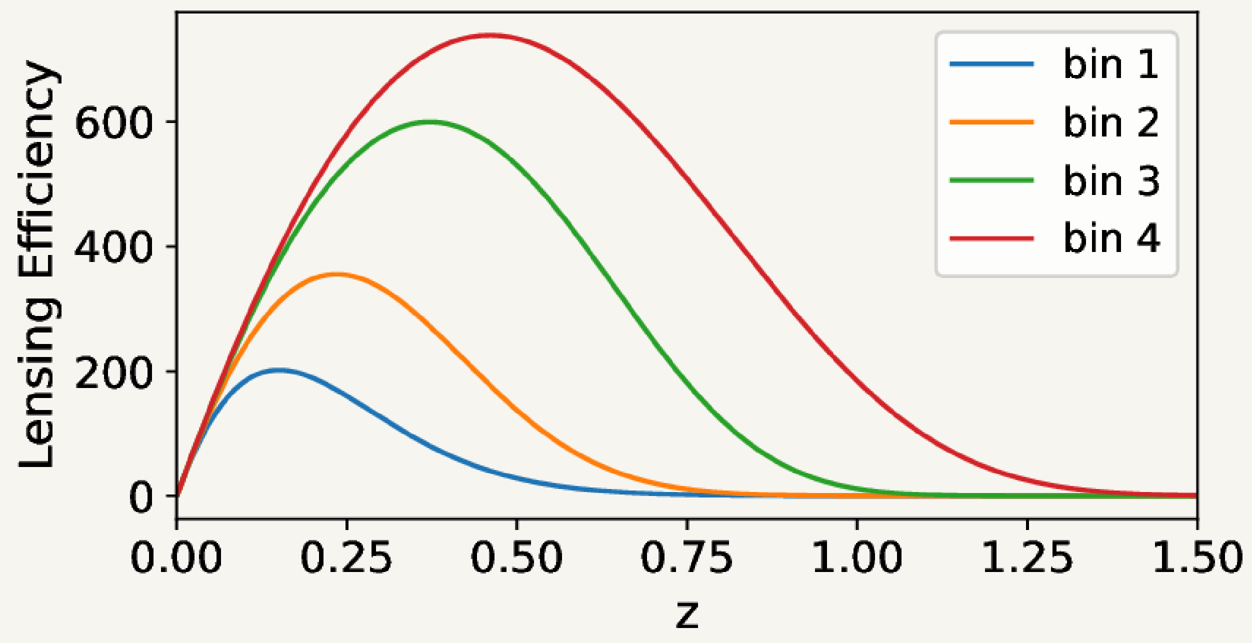
Credit: Justine Zeghal

§1 But could we do better than that? Weak lensing tomography

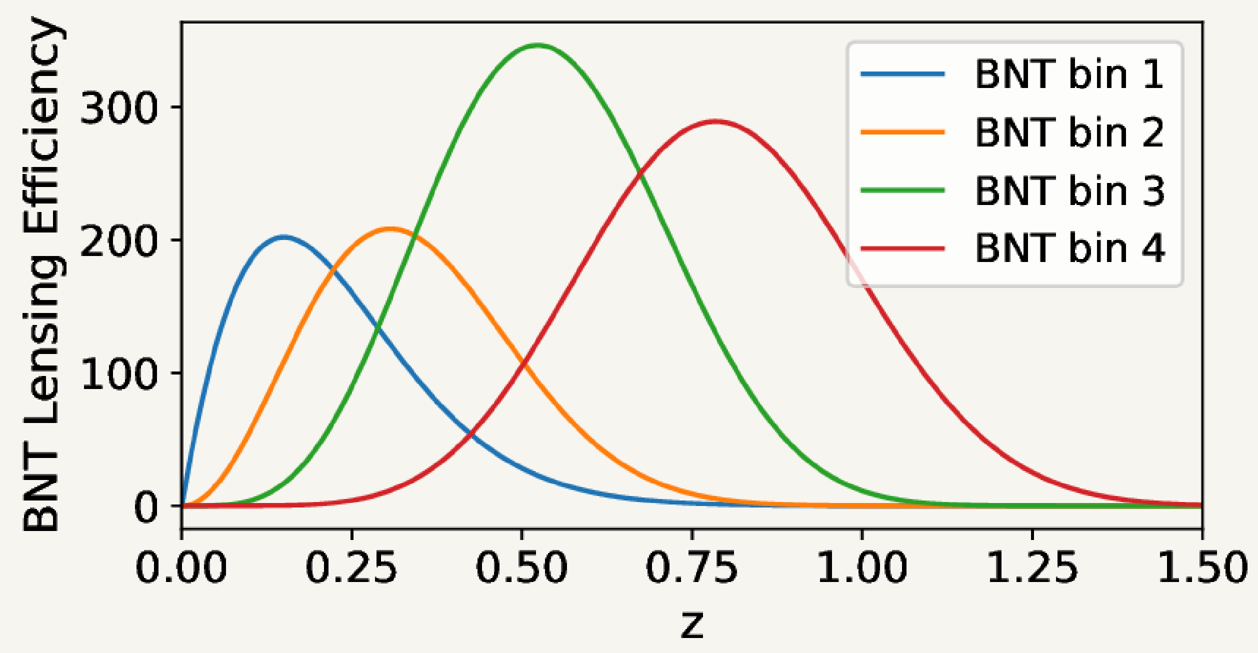


Credit: Justine Zeghal

§1 The BNT transform localizes each tomographic bin in redshift (a linear nulling)



standard: broad, overlapping



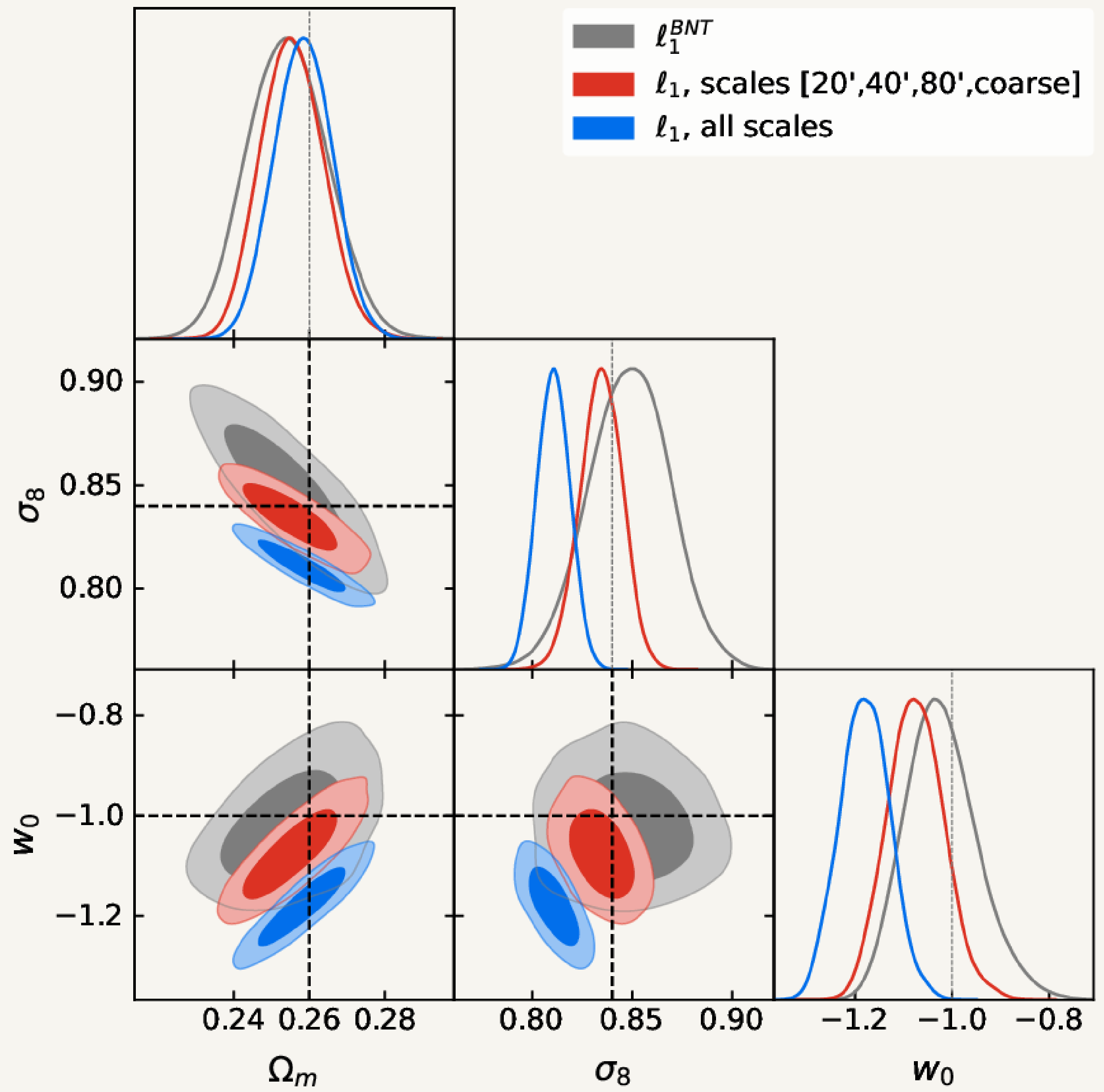
BNT: nulled, localized

WHAT BNT DOES · BERNARDEAU, NISHIMICHI & TARUYA 2014

- A linear, **invertible** nulling of the tomographic bins
- Standard kernels are broad and overlapping, so a fixed angular scale ℓ mixes many physical scales and redshifts
- BNT nulls the low-z lensing efficiency, **localizing each field in redshift** (thin lens-z slices), which sharpens the angular-scale ↔ physical-scale mapping for clean scale cuts

> Our use here: isolate the low-z, small-scale systematics (baryonic feedback) to specific bins, and cut scales only where needed, instead of discarding data everywhere.

§1 But applied to map-based HOS, BNT inflates the per-bin contours



The per-bin HOS contours **inflate dramatically** (gray = BNT; the noise mixing raises the floor).

THE HINGE

- BNT is *invertible*: no information can truly be lost
- Yet a Euclid forecast (**Vinciguerra et al. 2026**) still saw inflated BNT contours, even with explicit cross-bin HOS; recovering the SNR is "highly non-trivial"
- Is the information really lost, or are we just analyzing it wrong?

PART 2 OF 2

Learned vs analytical, and can we trust it?

The analytical ℓ1-norm vs a learned CNN, calibration, and the answer to the BNT puzzle.

§2 Part 2: learned summaries, and the BNT cliffhanger

THE QUESTION

How much better are "optimal", learned summaries than our hand-built summary statistics?

WHAT IS A LEARNED SUMMARY?

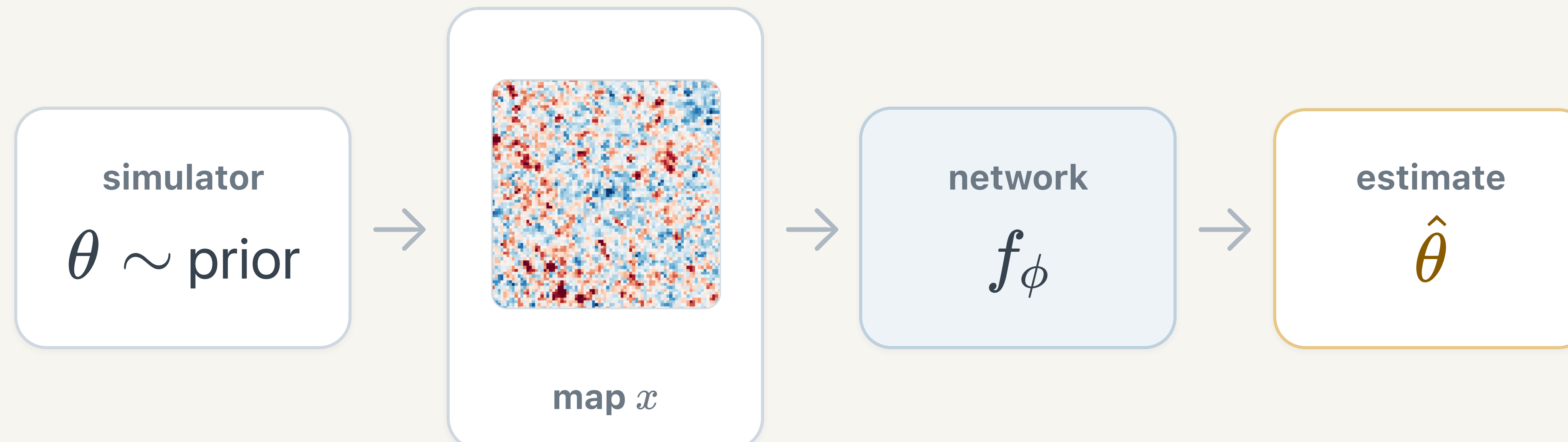
- A neural network that compresses the κ map directly into a few numbers, instead of a hand-designed statistic
- Trained with **VMIM** to keep the cosmological information: the "optimal learned compressor"

AND, LEFT OVER FROM PART 1

...and what the hell is going on with BNT?

Training a neural summary I: regression (MSE)

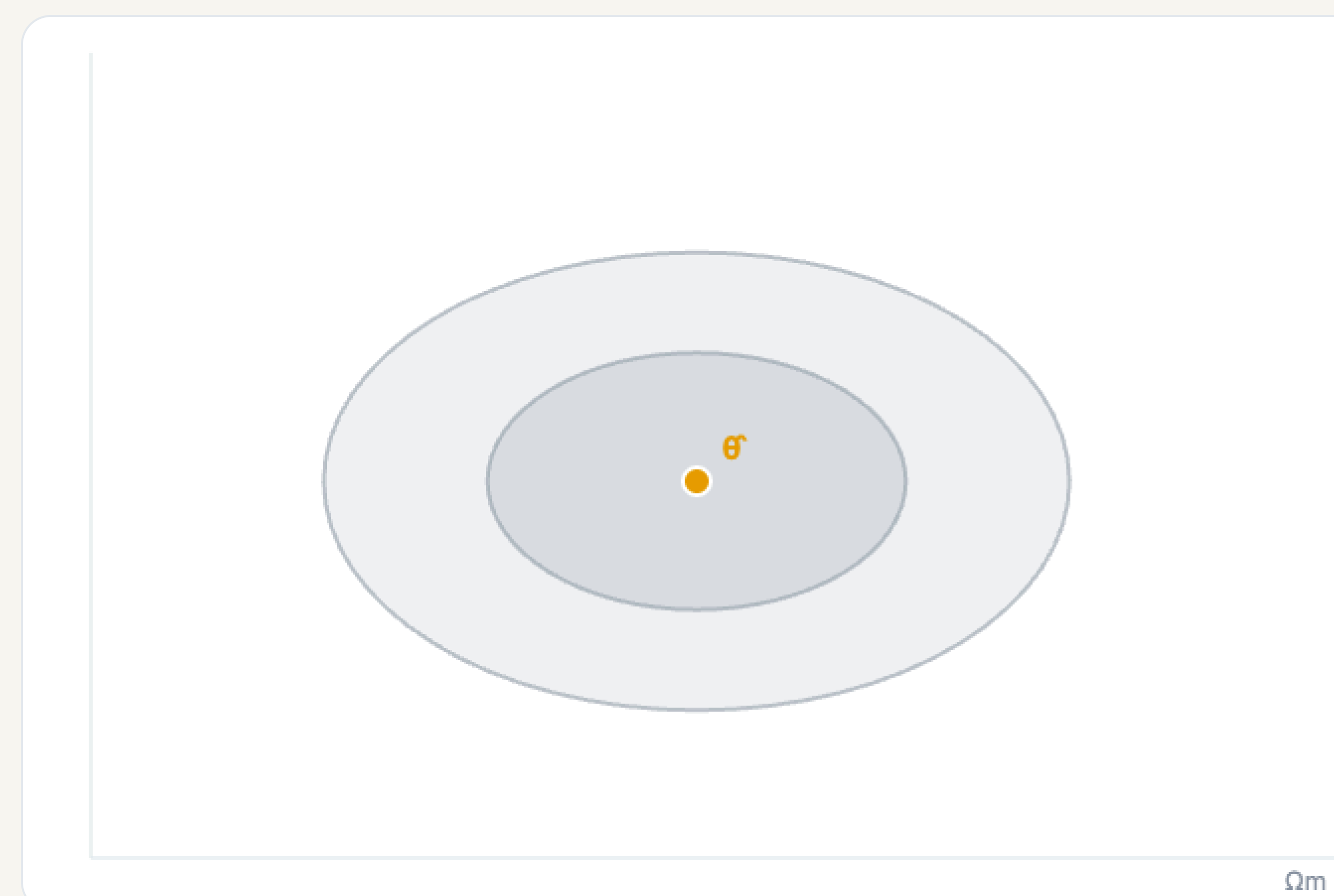
↻ replay



$$\mathcal{L} = \mathbb{E}\,\|\,\theta - f_\phi(x)\,\|^2$$

- A lensing map has $\sim 10^4$ pixels: too many to infer from directly
- Compress it into a low-dimensional **summary** $t = f_\phi(x)$
- **Regression:** train the network to predict the parameters (loss above)

parameter space θ

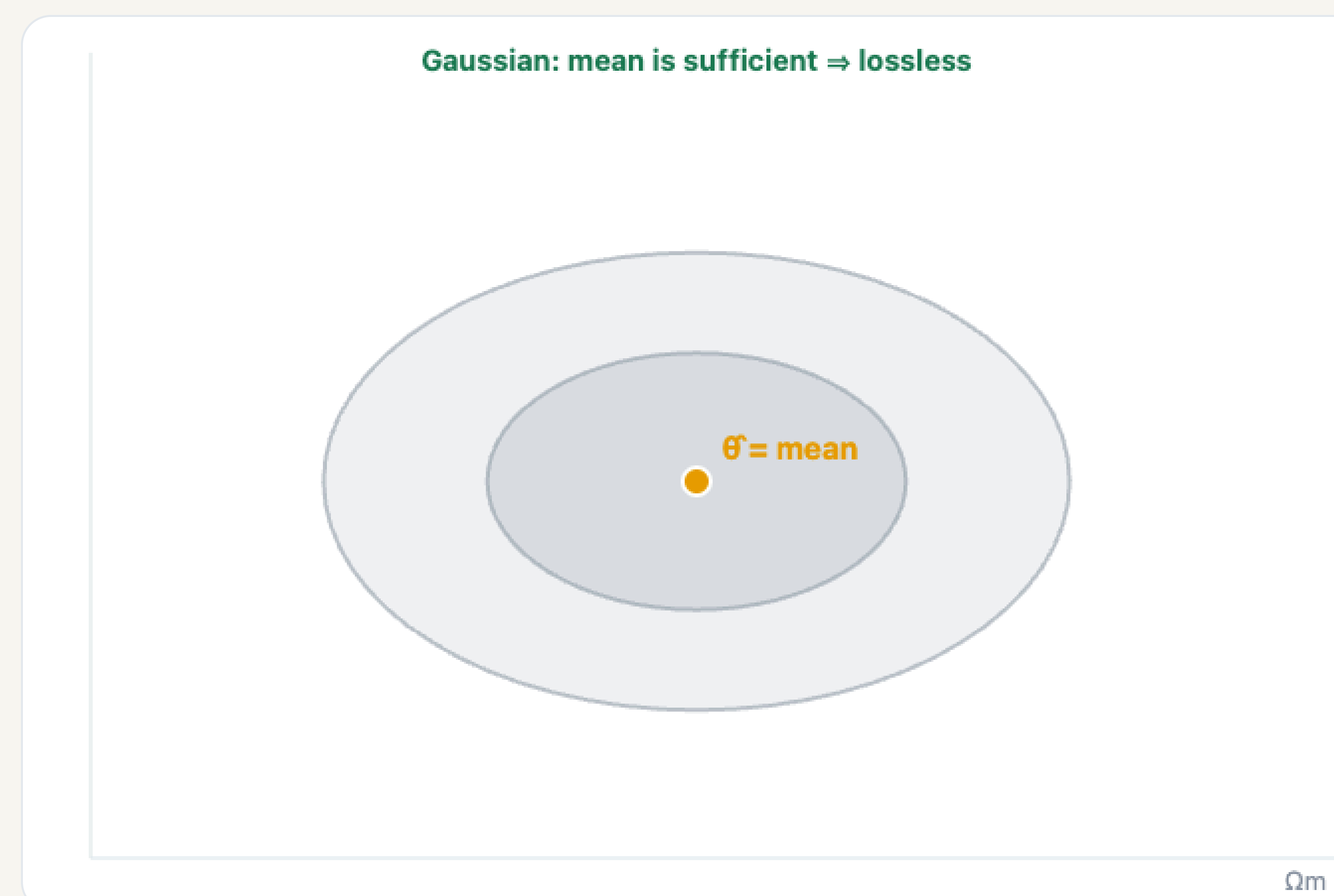


Training a neural summary I: regression (MSE)

↻ replay

$$\mathcal{L} = \mathbb{E}\,\big\|\,\theta - f_\phi(x)\,\big\|^2$$

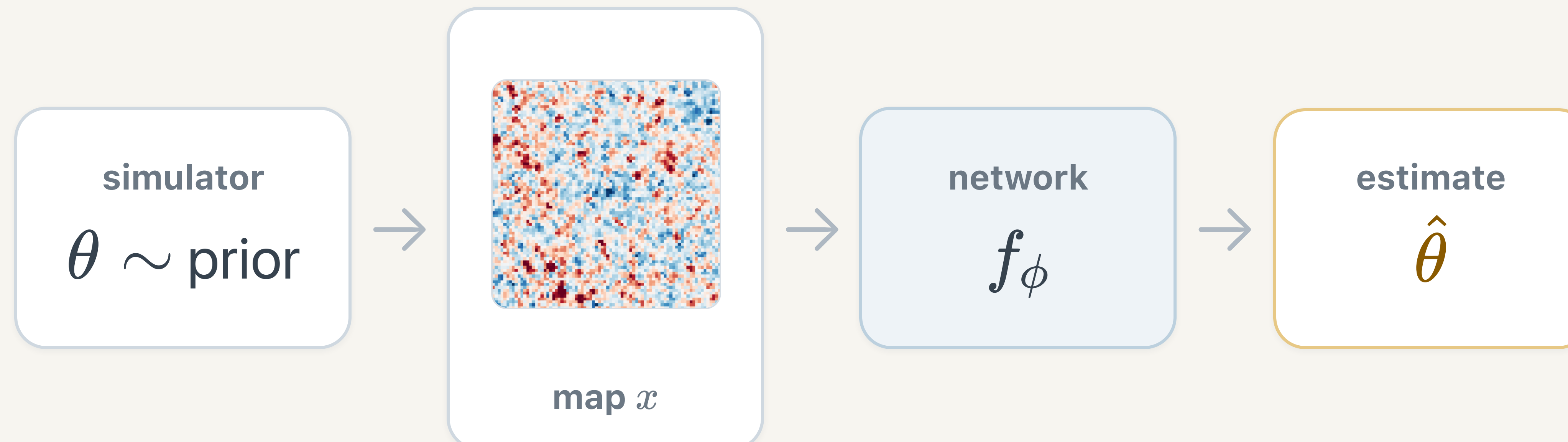
parameter space θ



- Optimum: the posterior **mean**, $\hat{\theta} = \mathbb{E}[\theta \mid x]$
- A compact, information-rich summary
- Gaussian $\Rightarrow$ mean is **sufficient** $\Rightarrow$ lossless

Training a neural summary I: regression (MSE)

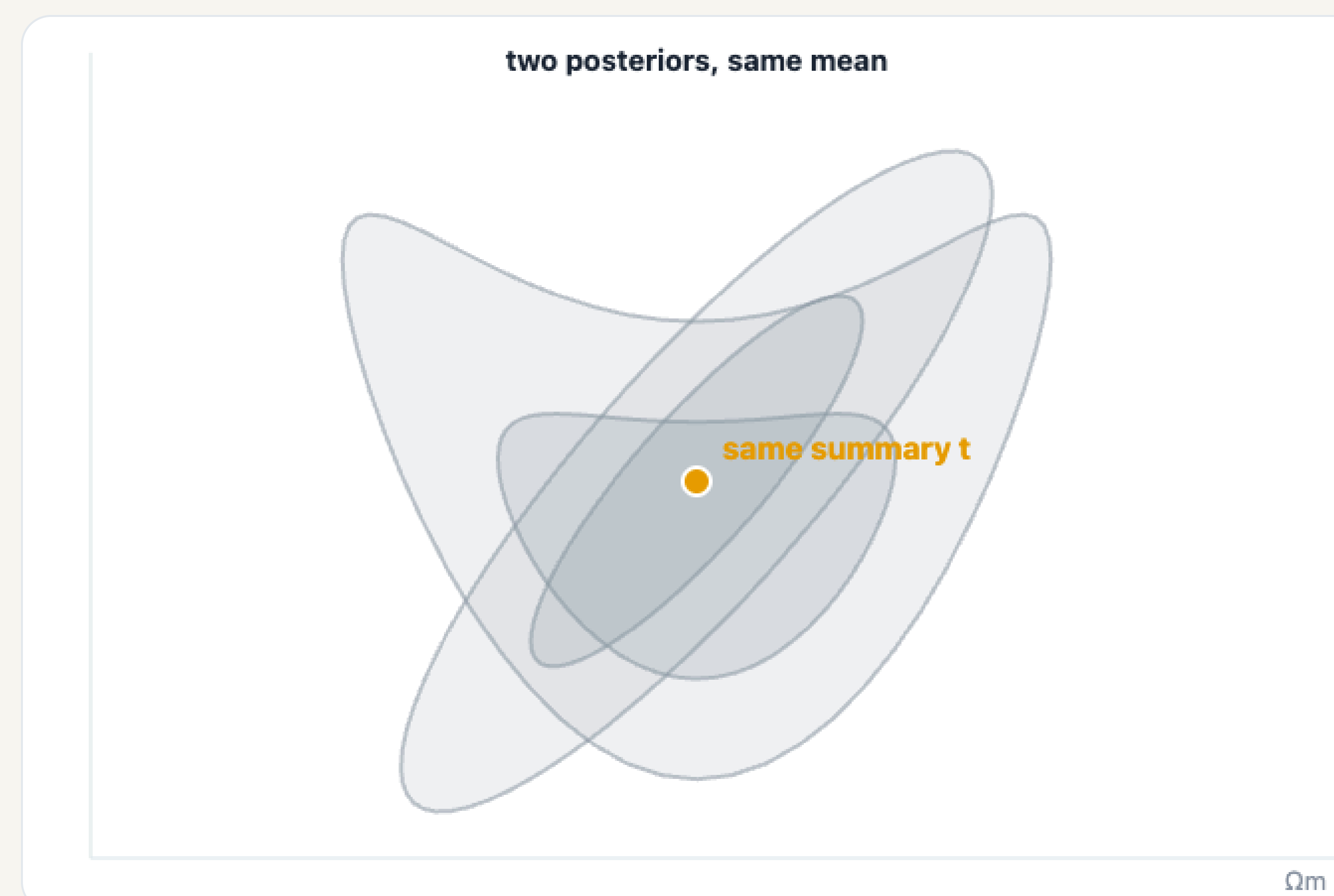
replay



$$\mathcal{L} = \mathbb{E}\,\big\|\,\theta - f_\phi(x)\,\big\|^2$$

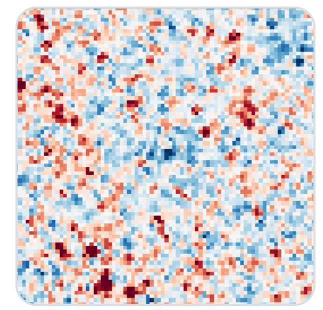
- But the mean is **not always sufficient**
- Same mean, different shape $\Rightarrow$ same summary
- The **non-Gaussian** information is lost

parameter space θ



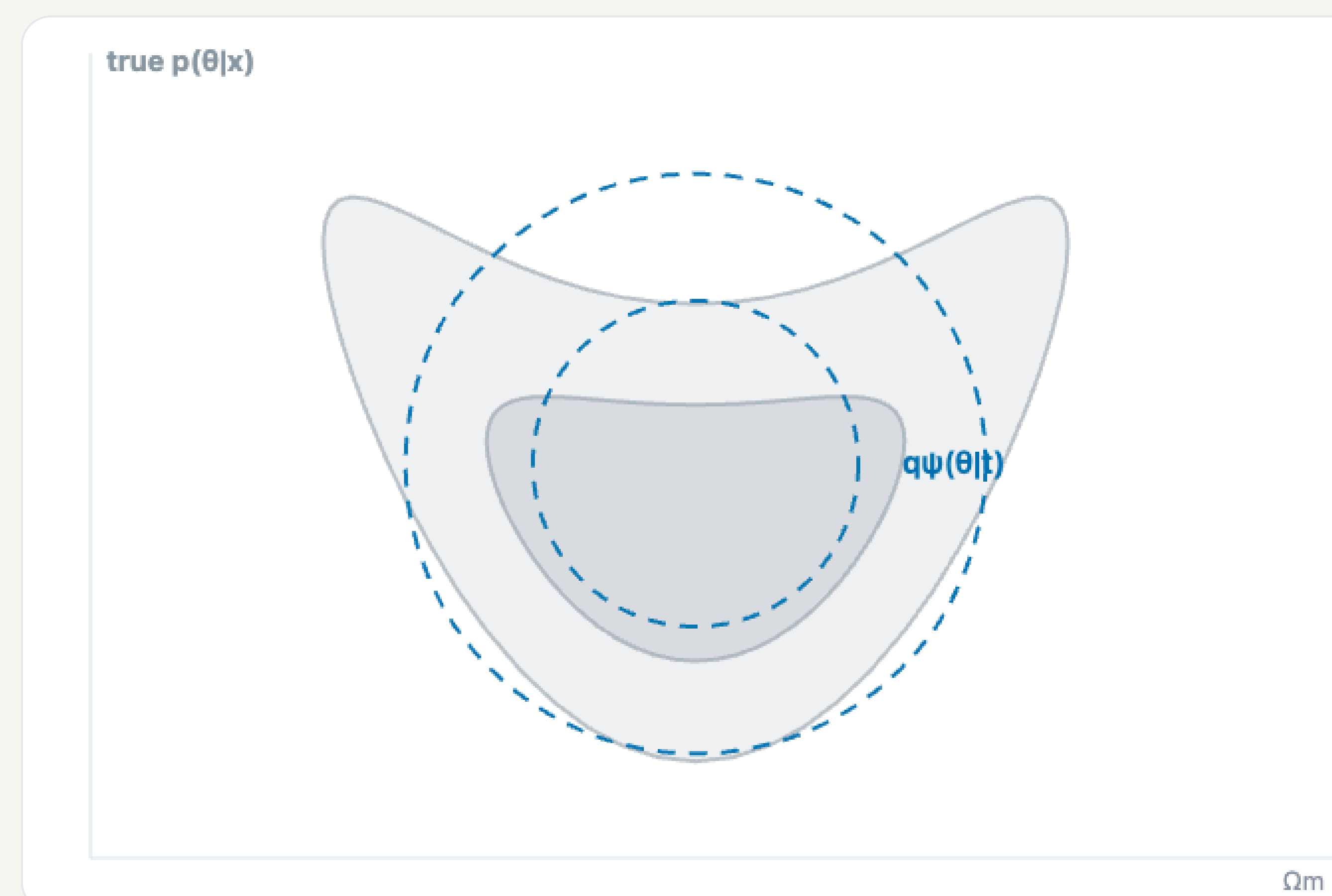
Training a neural summary II: VMIM

replay

simulator $\theta \sim$ prior →  map x → network f_ϕ → summary t → flow $q_\psi(\theta \mid t)$ → posterior $p(\theta \mid t)$

$$\max_{\phi,\psi} I(t;\theta) \;=\; \max_{\phi,\psi} \mathbb{E} \log q_\psi(\theta \mid f_\phi(x))$$

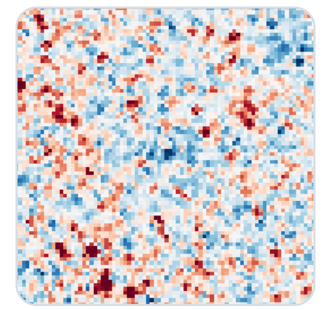
parameter space θ



- Same goal: compress the map to a **summary** $t = f_\phi(x)$
- t is any code; a flow $q_\psi(\theta \mid t)$ rebuilds the full posterior
- Train both to keep **maximal information** about θ (objective above)

Training a neural summary II: VMIM

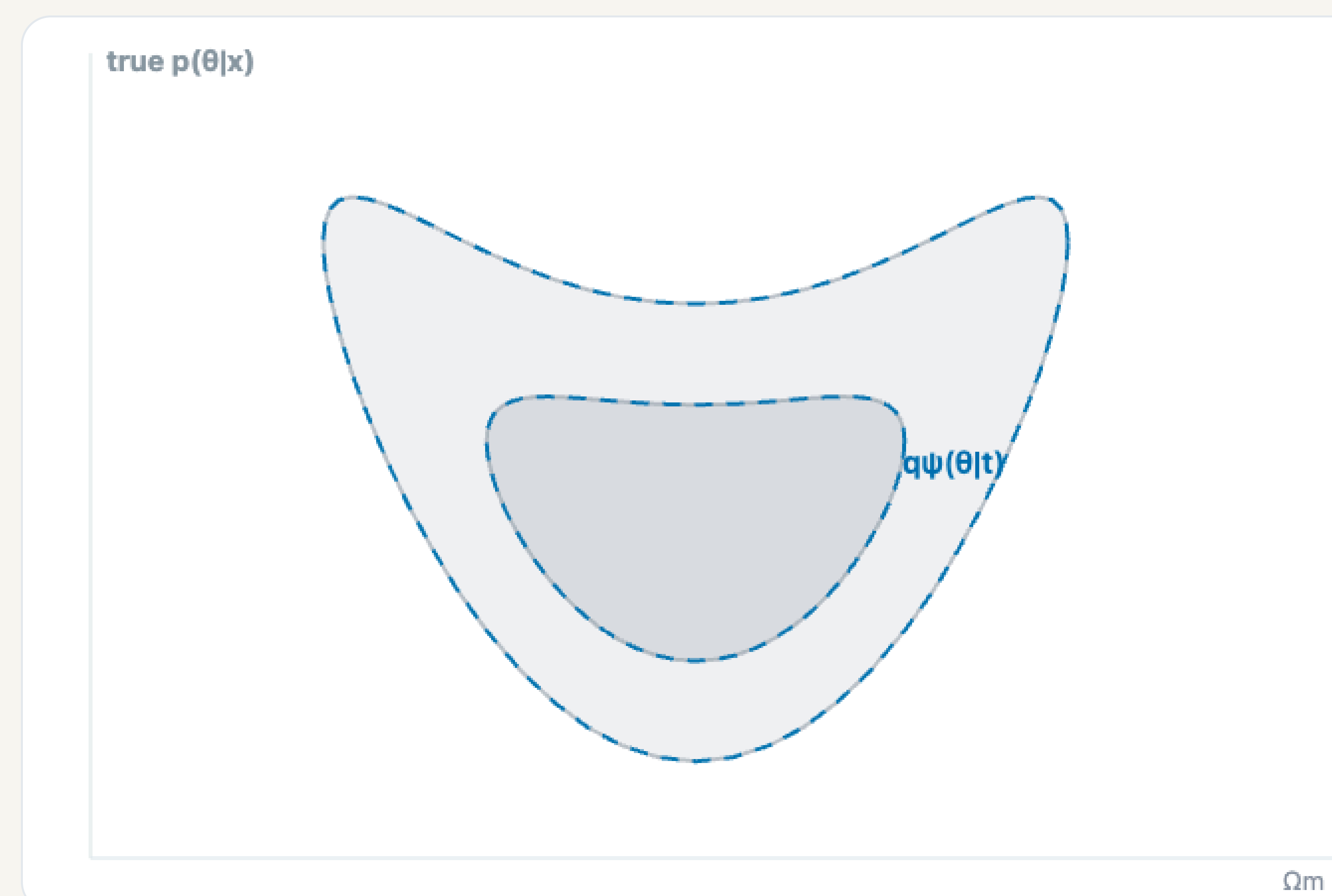
↻ replay

simulator $\theta \sim$ prior →  map x → network f_ϕ → summary t → flow $q_\psi(\theta \mid t)$ → posterior $p(\theta \mid t)$

$$\max_{\phi,\psi} I(t;\theta) = \max_{\phi,\psi} \mathbb{E} \log q_\psi(\theta \mid f_\phi(x))$$

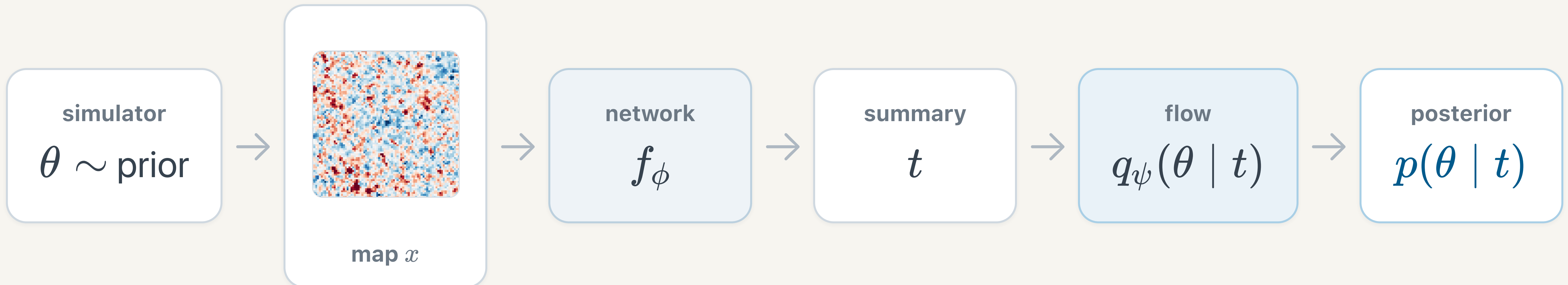
- Maximizing $I(t;\theta)$ reshapes the flow to match $p(\theta \mid x)$
- At the optimum, t is a **sufficient statistic**
- $p(\theta \mid x) = p(\theta \mid t)$

parameter space θ



Training a neural summary II: VMIM

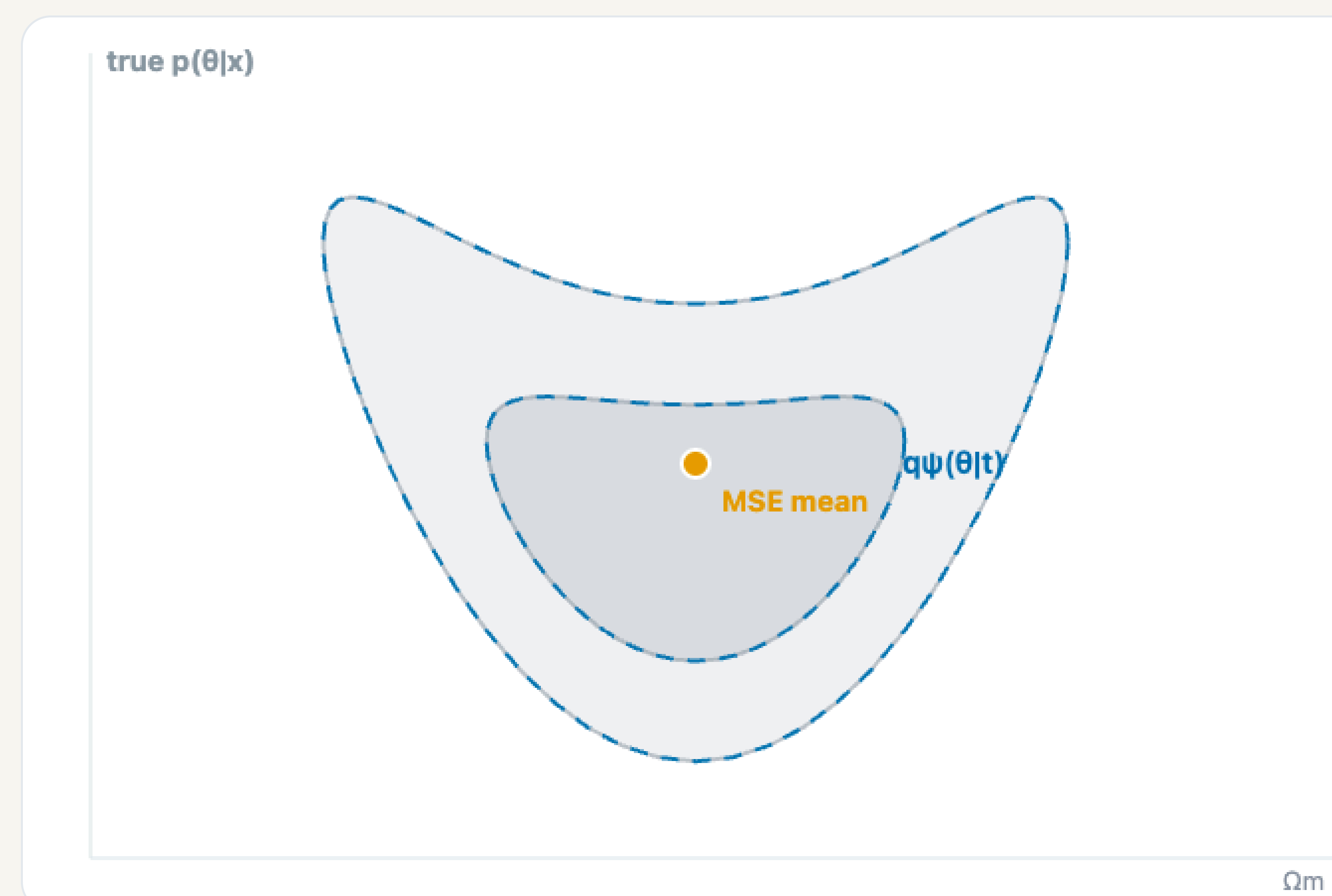
↻ replay



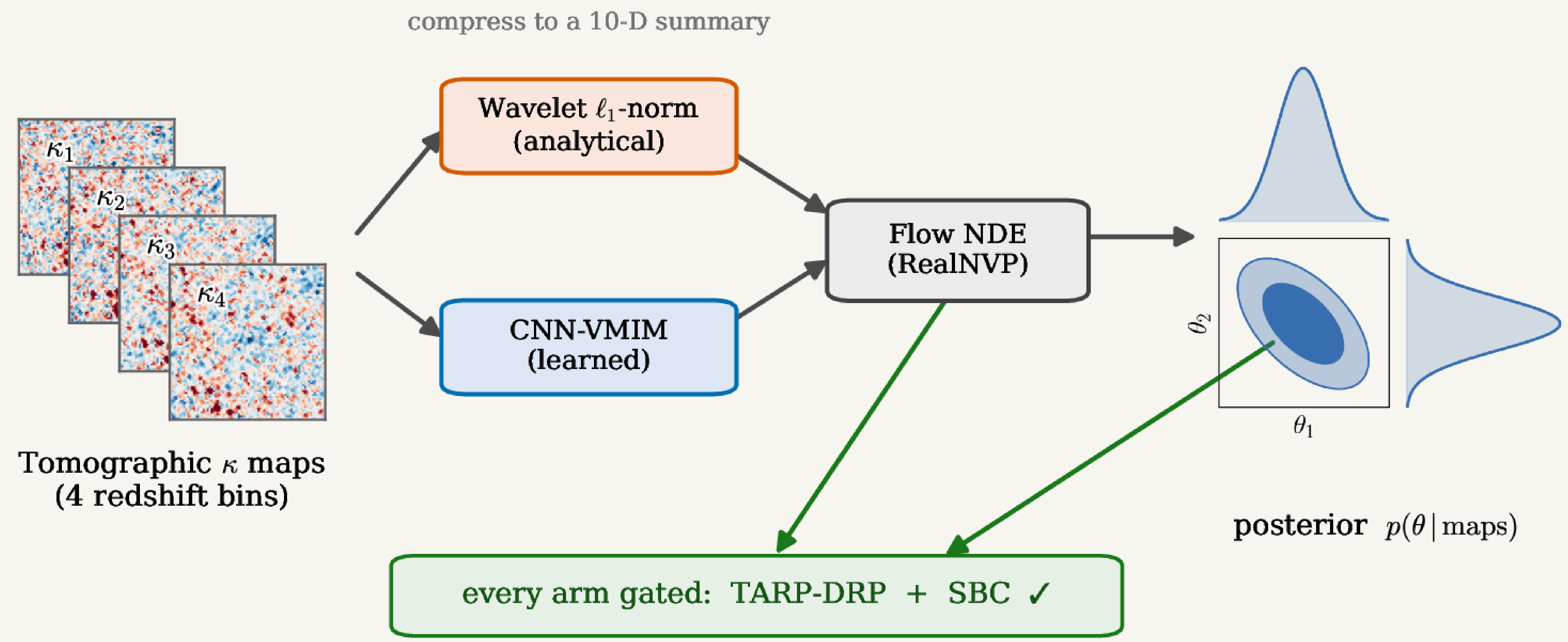
$$\max_{\phi,\psi} I(t;\theta) = \max_{\phi,\psi} \mathbb{E} \log q_\psi(\theta \mid f_\phi(x))$$

- Regression keeps only the **mean** (dot)
- VMIM keeps the **full shape**
- Gaussian: they agree; **non-Gaussian**: VMIM stays sufficient

parameter space θ



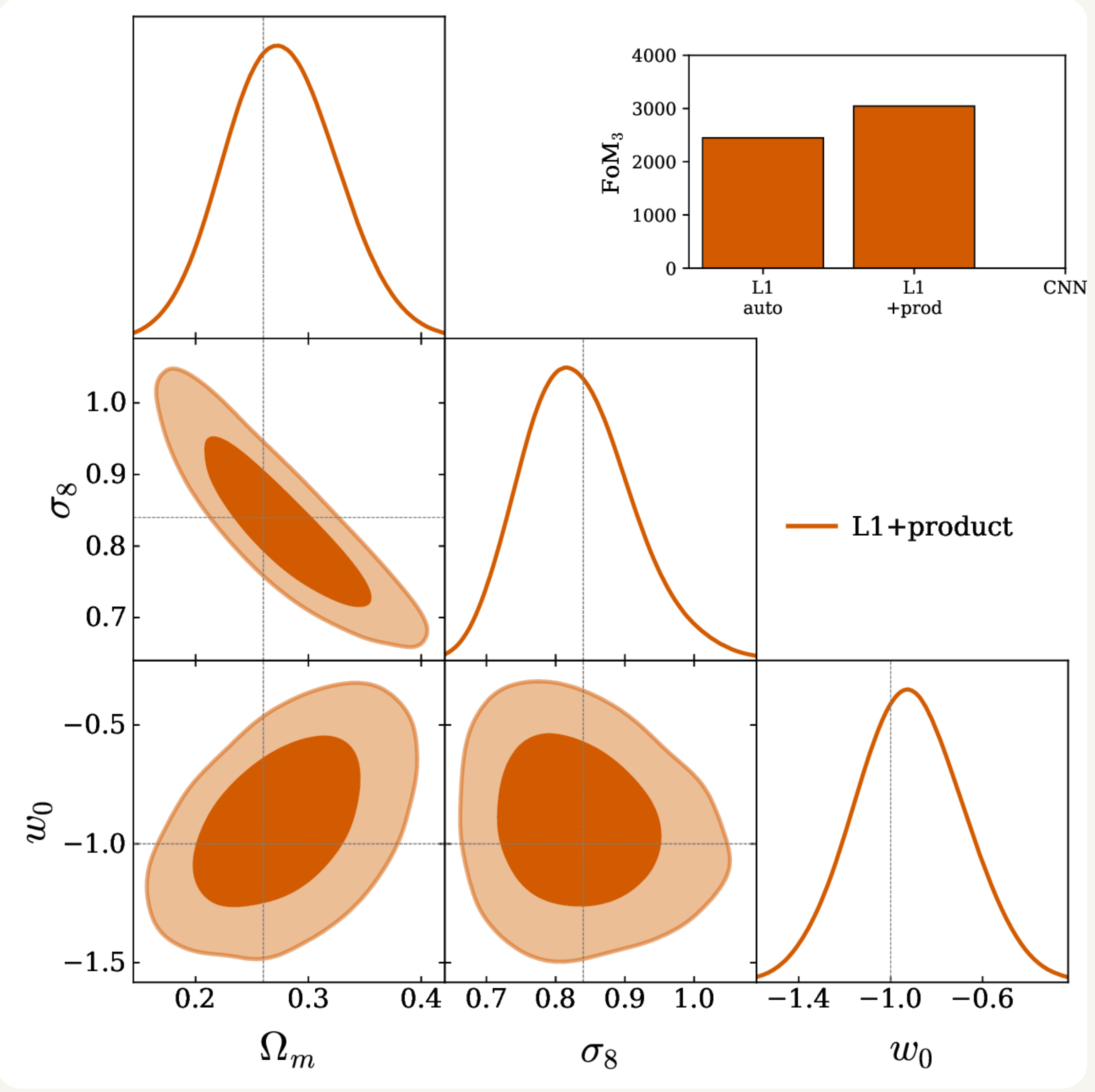
§2 The comparison, done fairly: same maps, same flow, both calibrated



APPLES TO APPLES

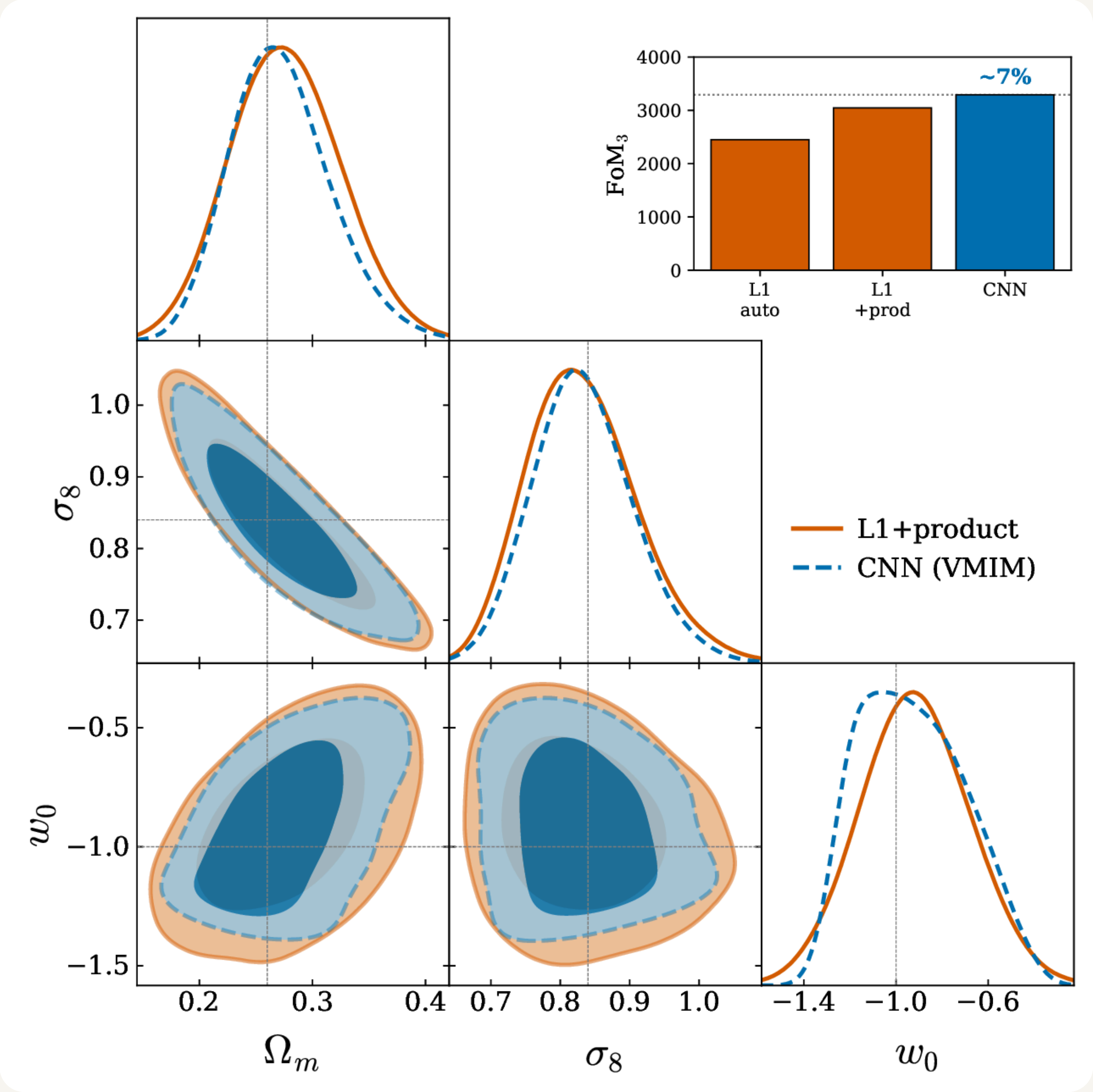
- Same κ maps → **ℓ1-norm** or **CNN-VMIM** → the *same* flow NDE → posterior
- Flat-sky **10° patches** (cross-maps physically buildable), both arms calibrated
- **324k patches, 899 cosmologies**

§2 The analytical ℓ1-norm almost reaches the optimal CNN (~7%): The hand-built ℓ1+product is *near-sufficient*



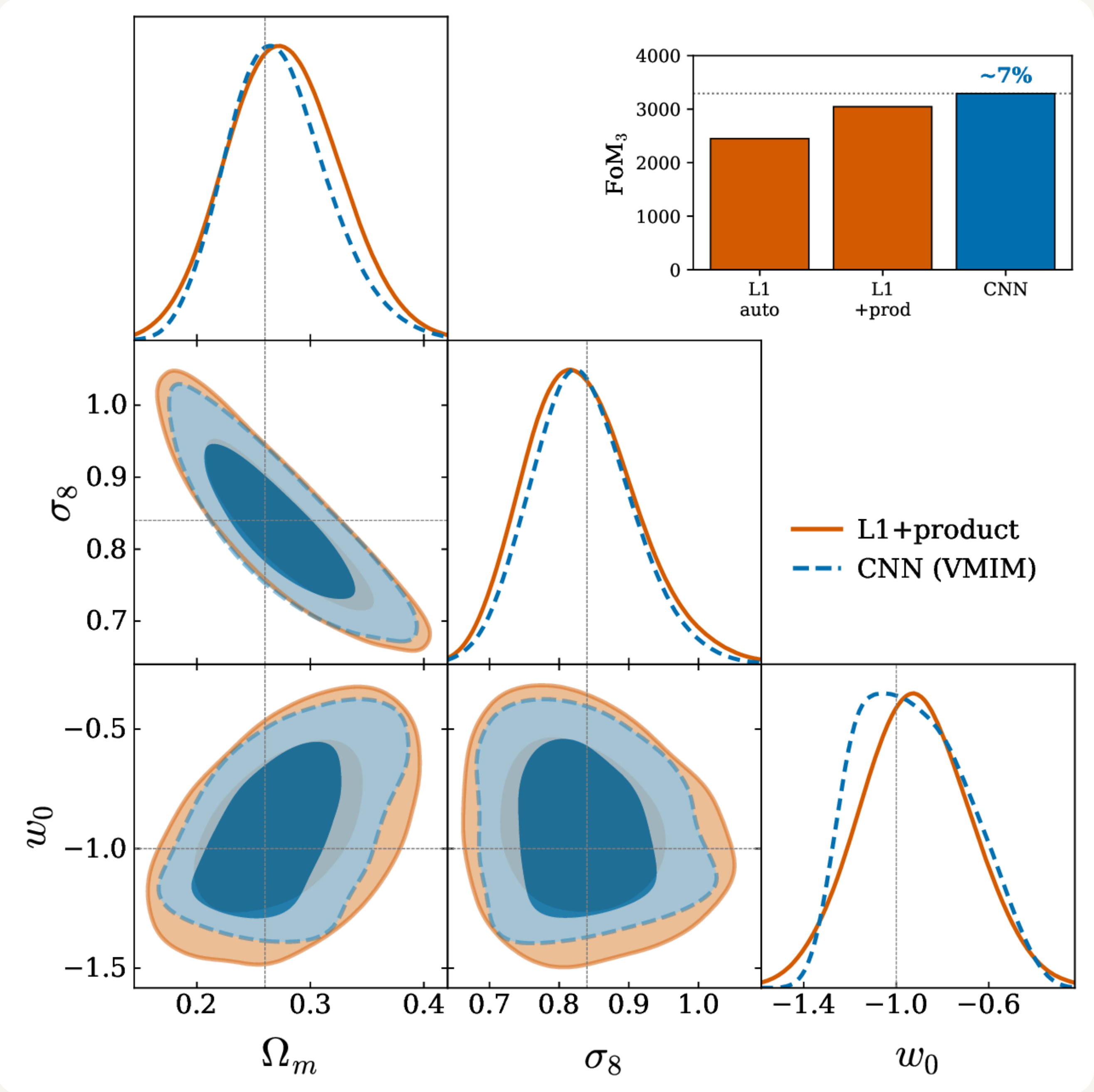
ℓ1 and CNN posteriors nearly coincide

§2 The analytical ℓ1-norm almost reaches the optimal CNN (~7%): The hand-built ℓ1+product is *near-sufficient*

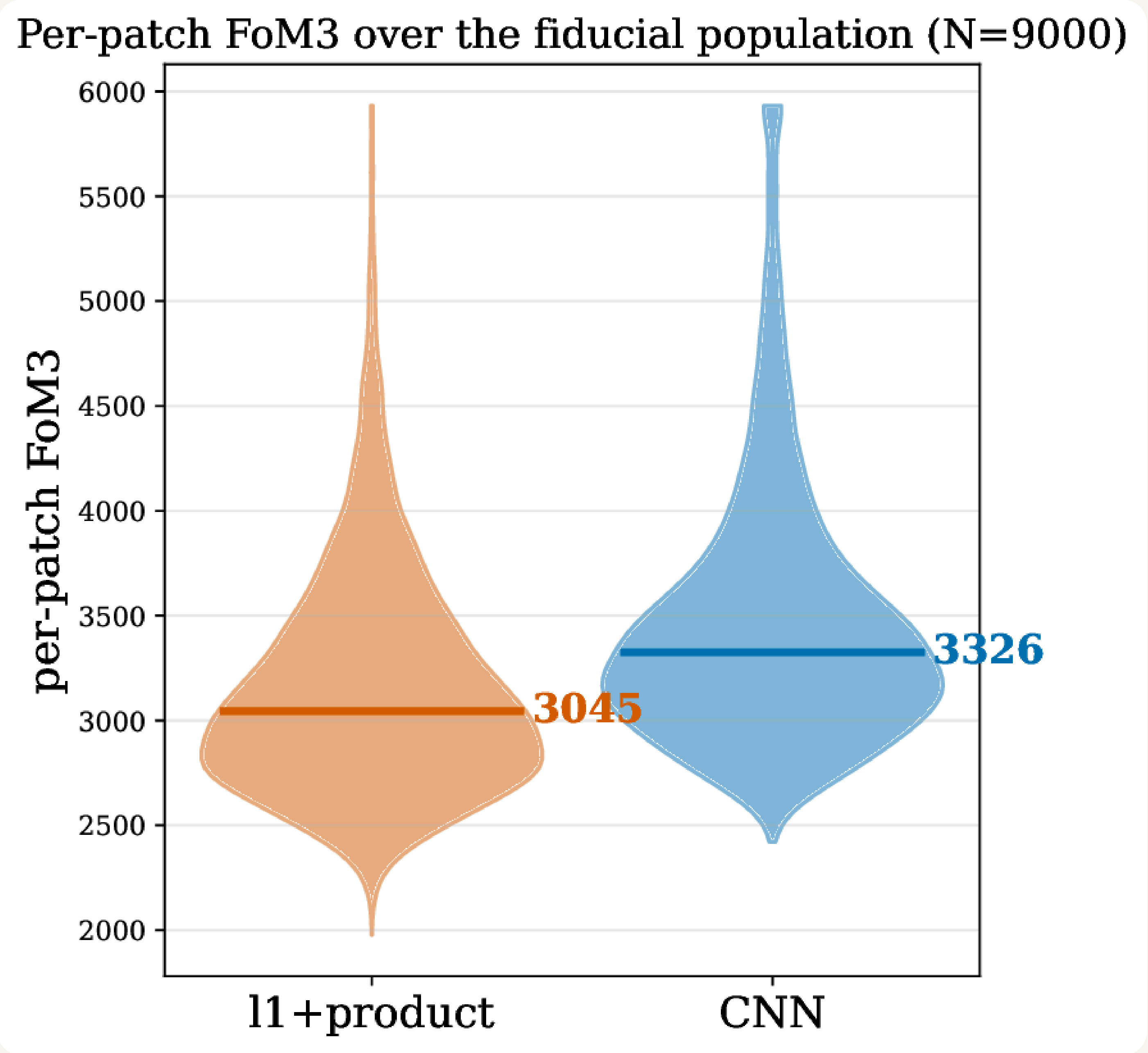


ℓ1 and CNN posteriors nearly coincide

§2 The analytical ℓ1-norm almost reaches the optimal CNN (~7%): The hand-built ℓ1+product is *near-sufficient*

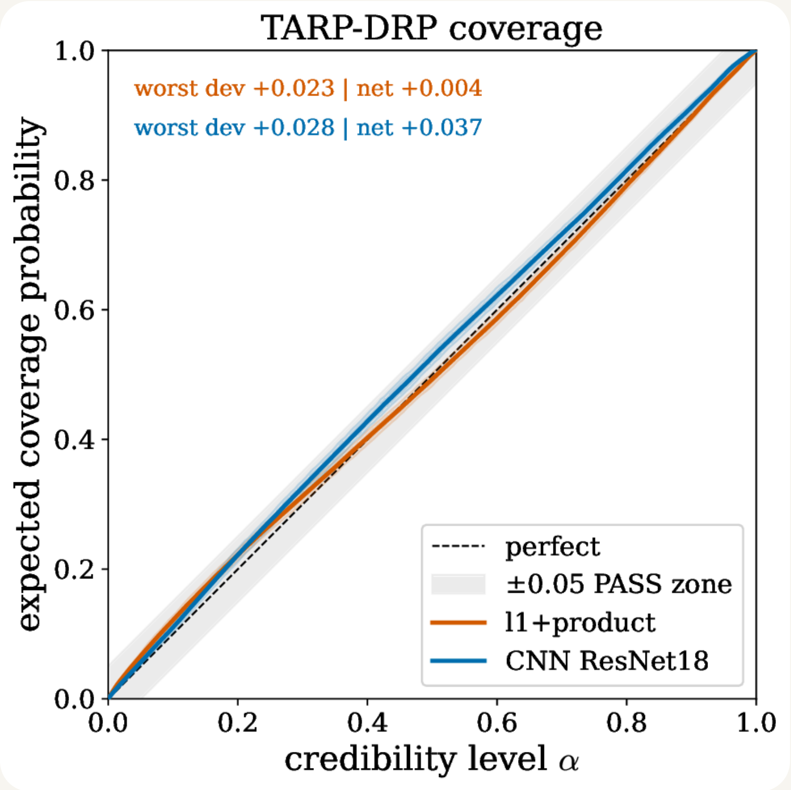


ℓ1 and CNN posteriors nearly coincide

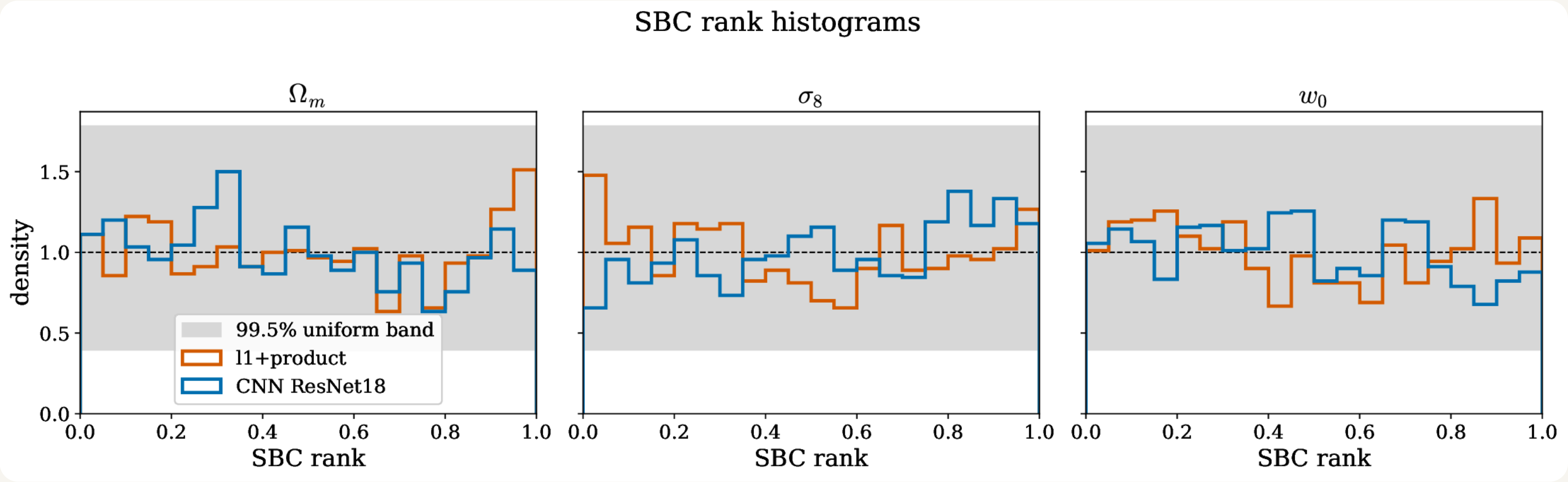


near-identical FoM3 across patches (3045 vs 3326)

§2 Can we trust it? Every arm passes the same TARP + SBC tests reasonably



TARP-DRP coverage: on the diagonal

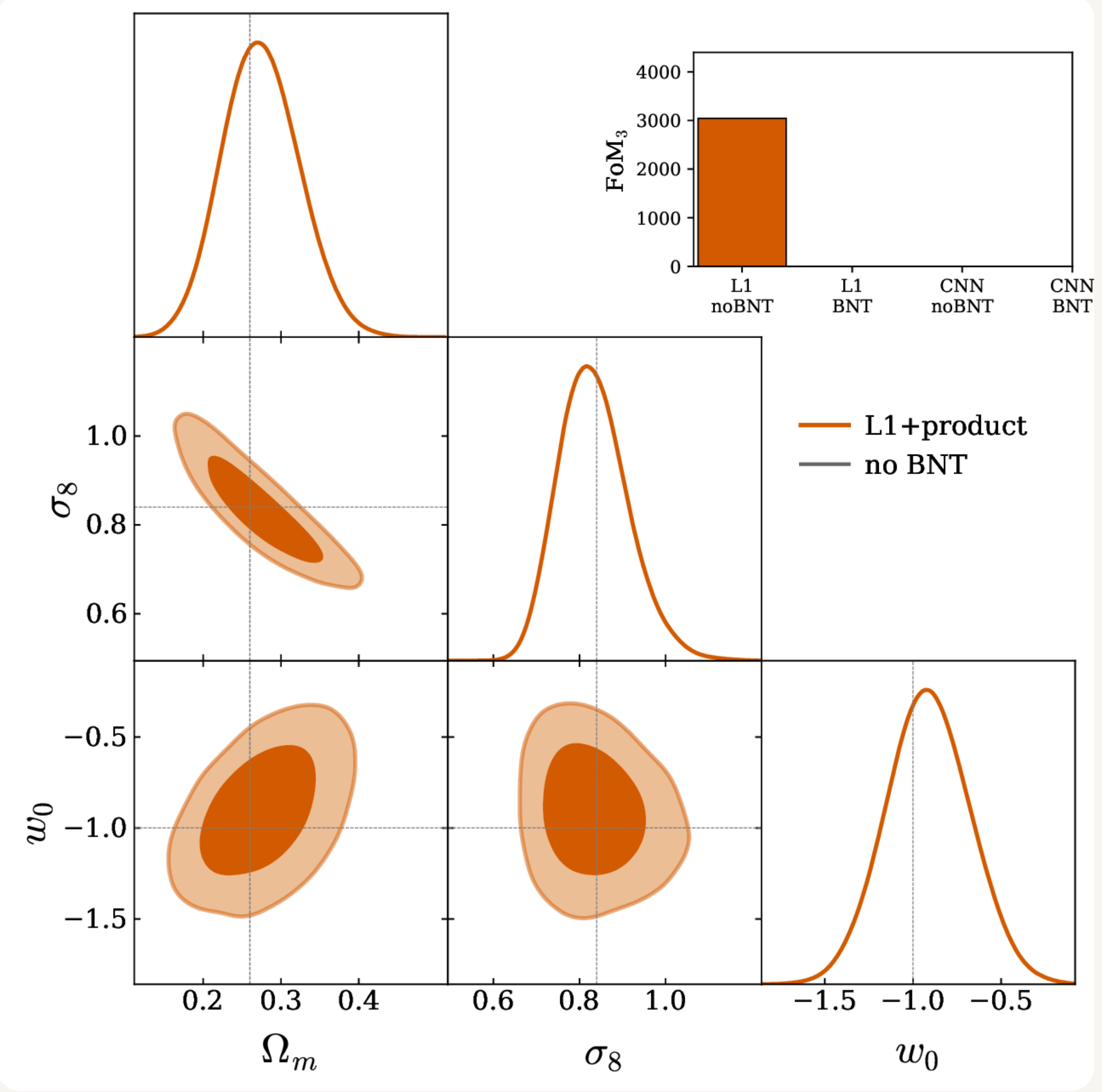


SBC ranks: flat within the 99.5% band

TIGHT IS NOT THE SAME AS CORRECT

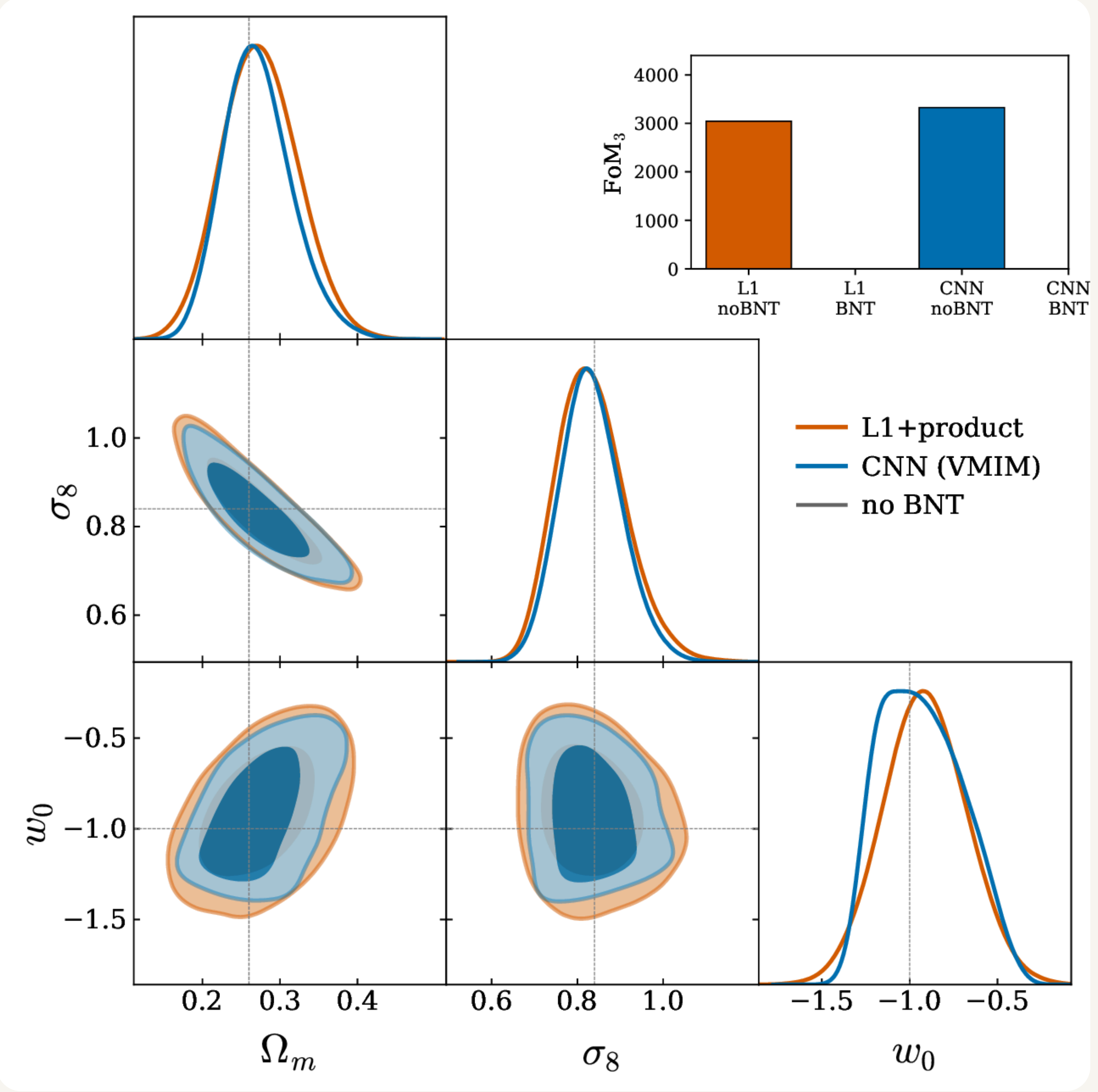
- Both arms pass the *same* battery: varied-θ TARP-DRP coverage + SBC rank uniformity
- The constraining power is real

§2 BNT revisited: the per-bin ℓ1 collapses (0.26×), the channel-mixing CNN is lossless (0.96×)



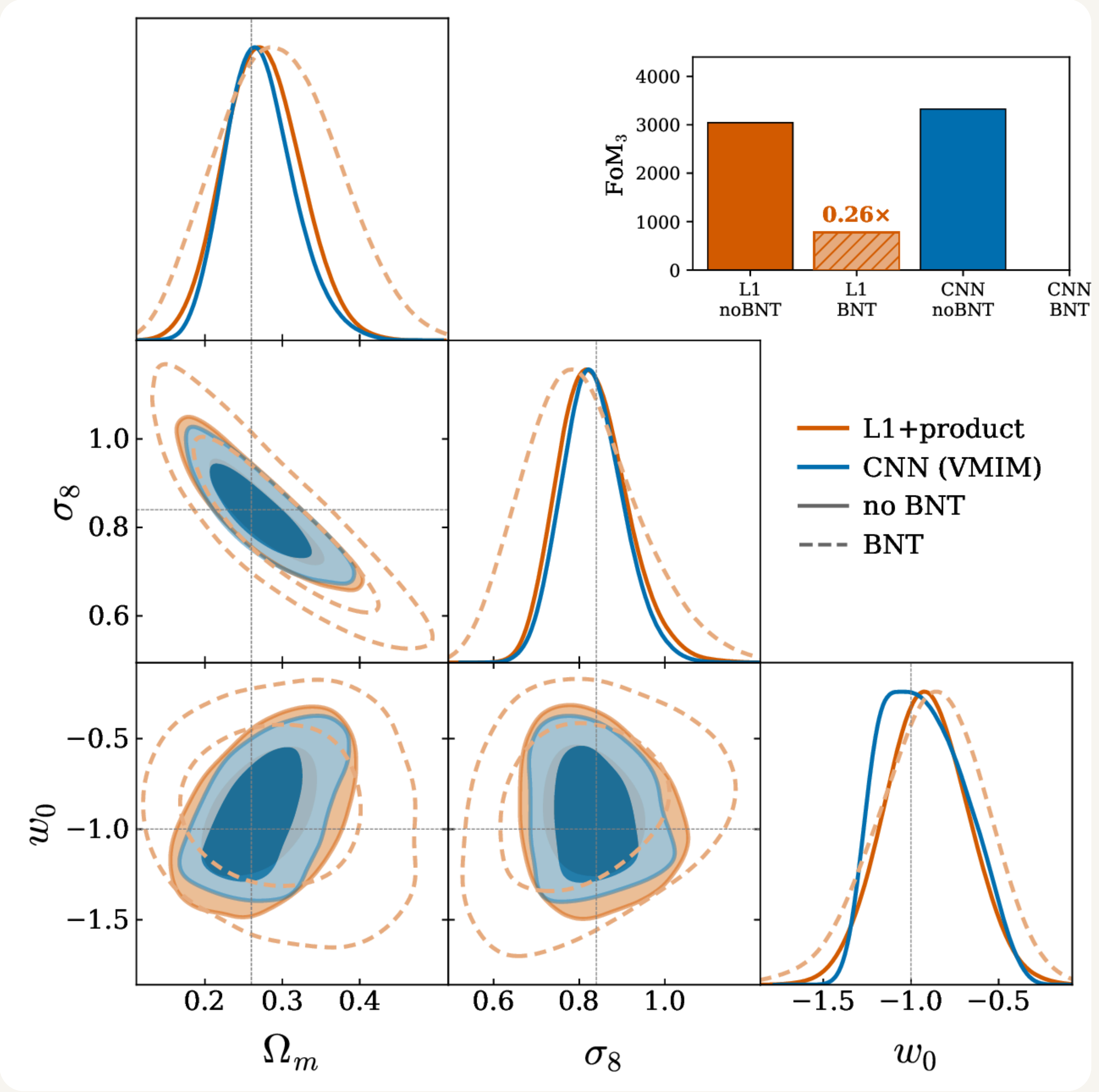
SAME TRANSFORM, OPPOSITE FATES

§2 BNT revisited: the per-bin ℓ1 collapses (0.26×), the channel-mixing CNN is lossless (0.96×)



SAME TRANSFORM, OPPOSITE FATES

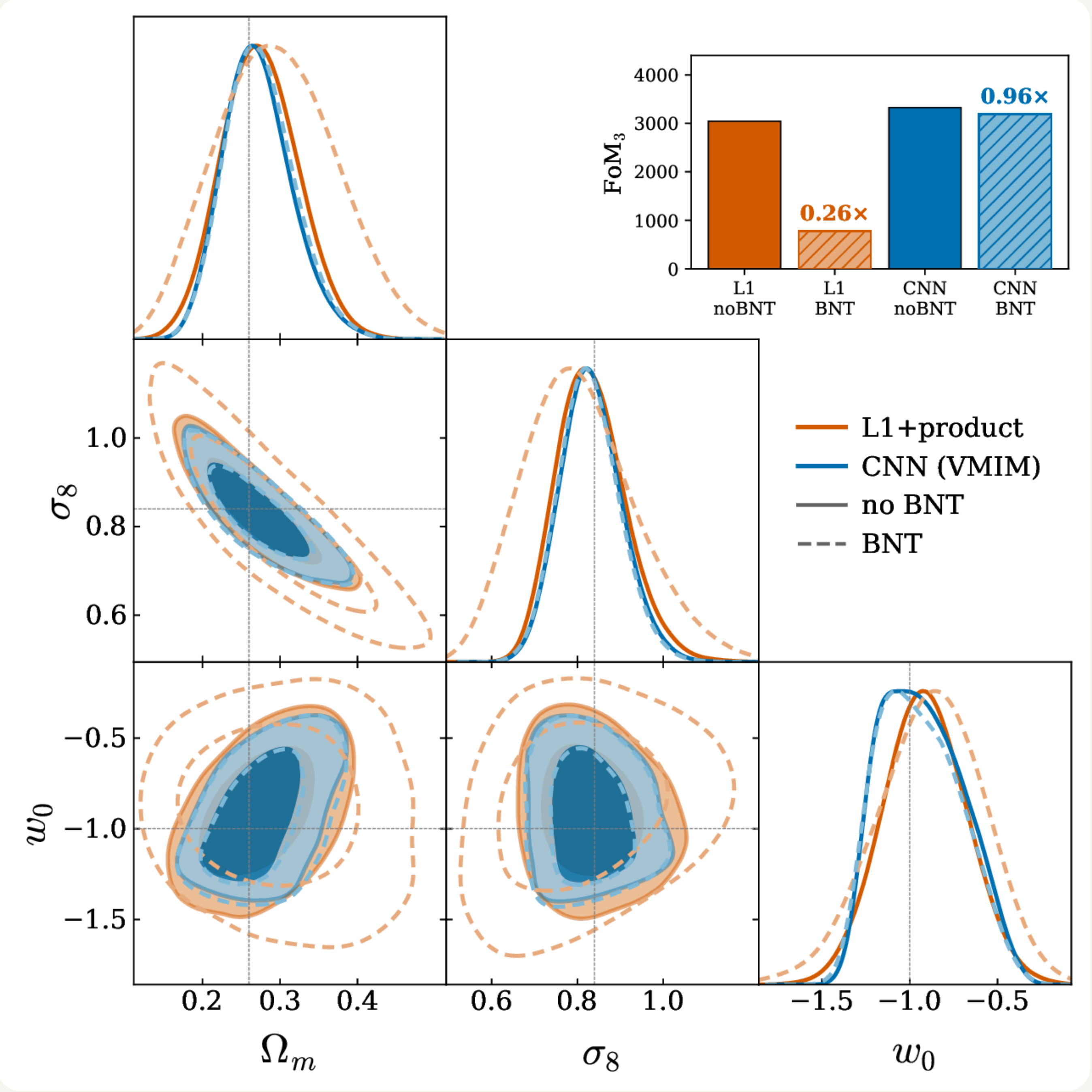
§2 BNT revisited: the per-bin ℓ1 collapses (0.26×), the channel-mixing CNN is lossless (0.96×)



SAME TRANSFORM, OPPOSITE FATES

- Per-channel **ℓ1+product collapses**: 3045 → 779 (**0.26×**; σ_8 +65%)

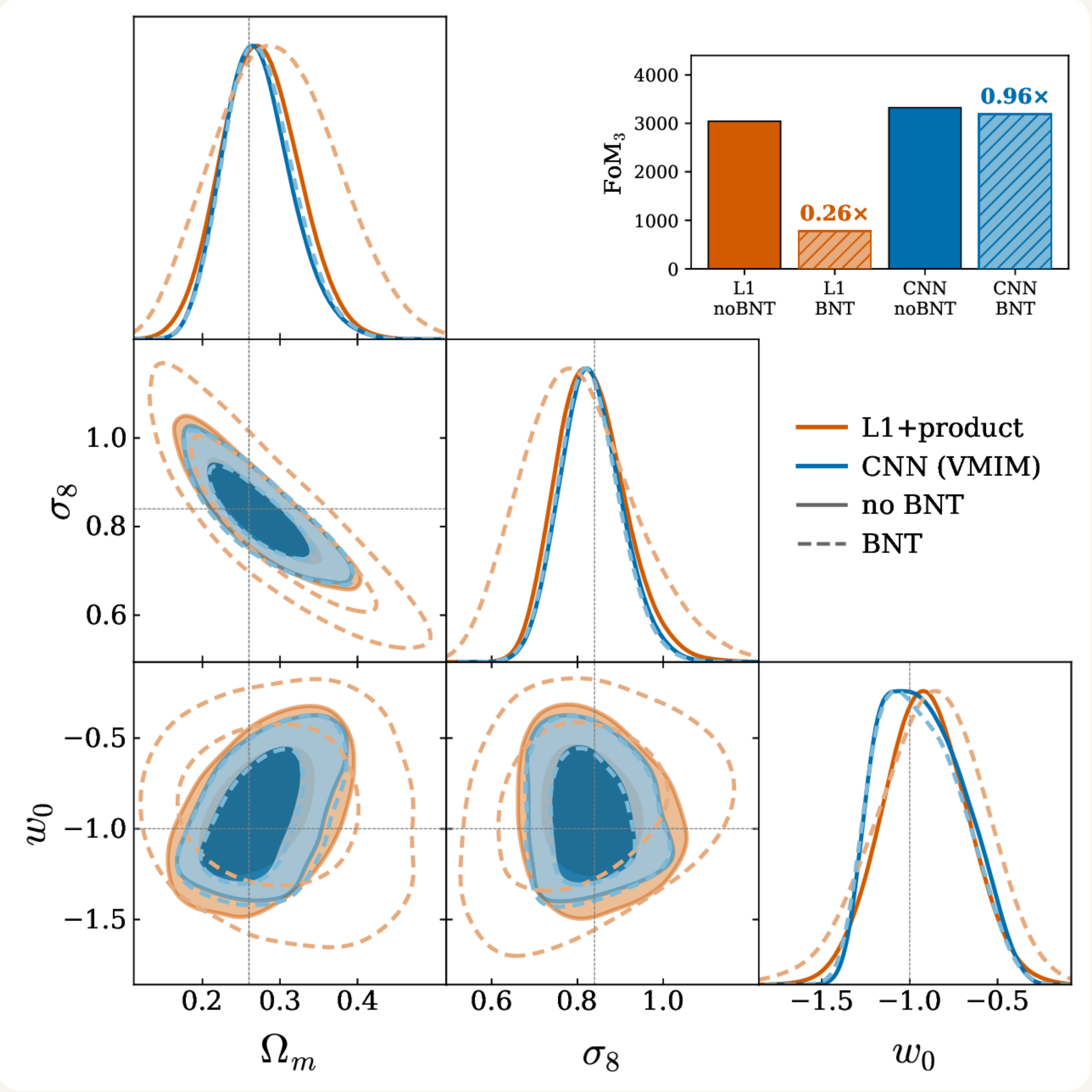
§2 BNT revisited: the per-bin ℓ1 collapses (0.26×), the channel-mixing CNN is lossless (0.96×)



SAME TRANSFORM, OPPOSITE FATES

- Per-channel **ℓ1+product collapses**: 3045 → 779 (**0.26×**; σ_8 +65%)
- Channel-mixing **CNN** is **lossless**: 3326 → 3186 (**0.96×**)

§2 BNT revisited: the per-bin ℓ1 collapses (0.26×), the channel-mixing CNN is lossless (0.96×)

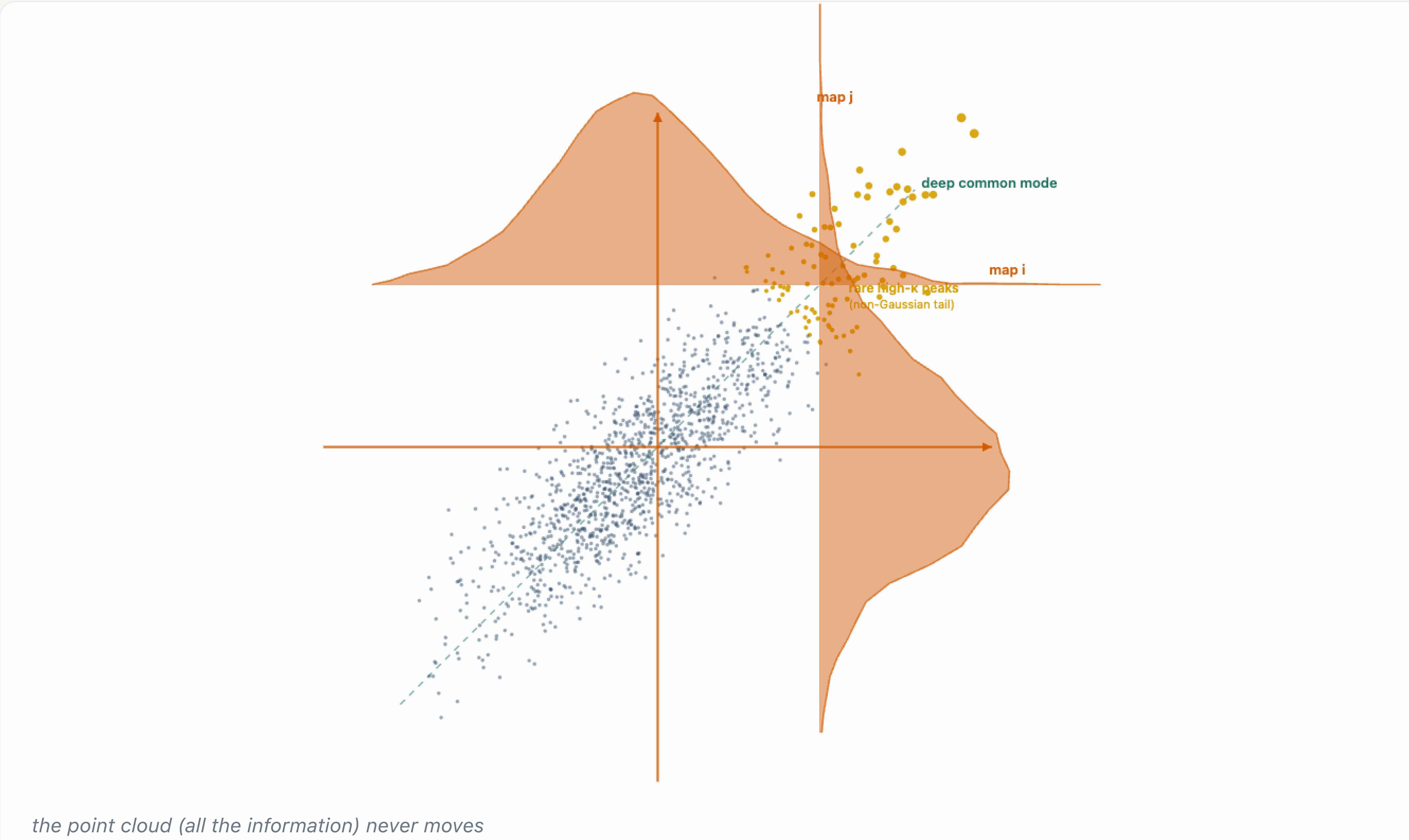


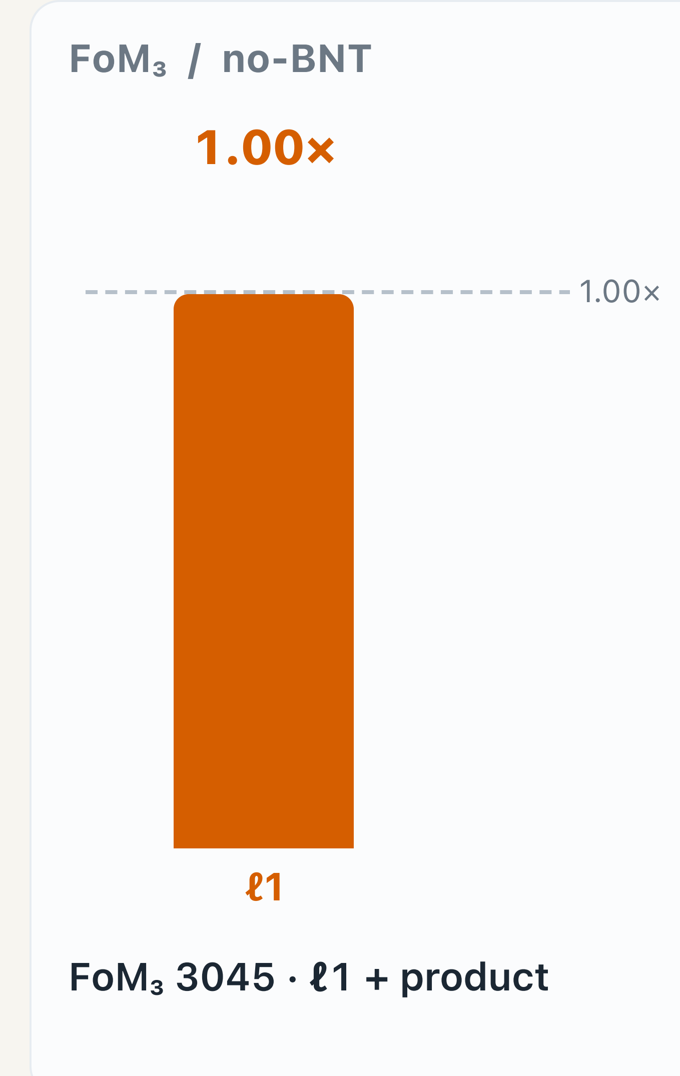
SAME TRANSFORM, OPPOSITE FATES

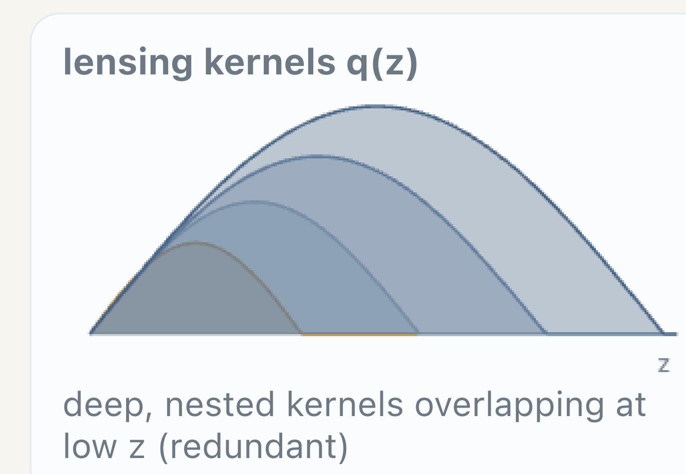
- Per-channel **ℓ1+product** **collapses**: 3045 → 779 (**0.26×**; σ_8 +65%)
- Channel-mixing **CNN** is **lossless**: 3326 → 3186 (**0.96×**)
- The collapse is itself *calibrated*: a real loss

Same information, a different frame: why BNT collapses the ℓ1-norm but not the CNN

↻ replay



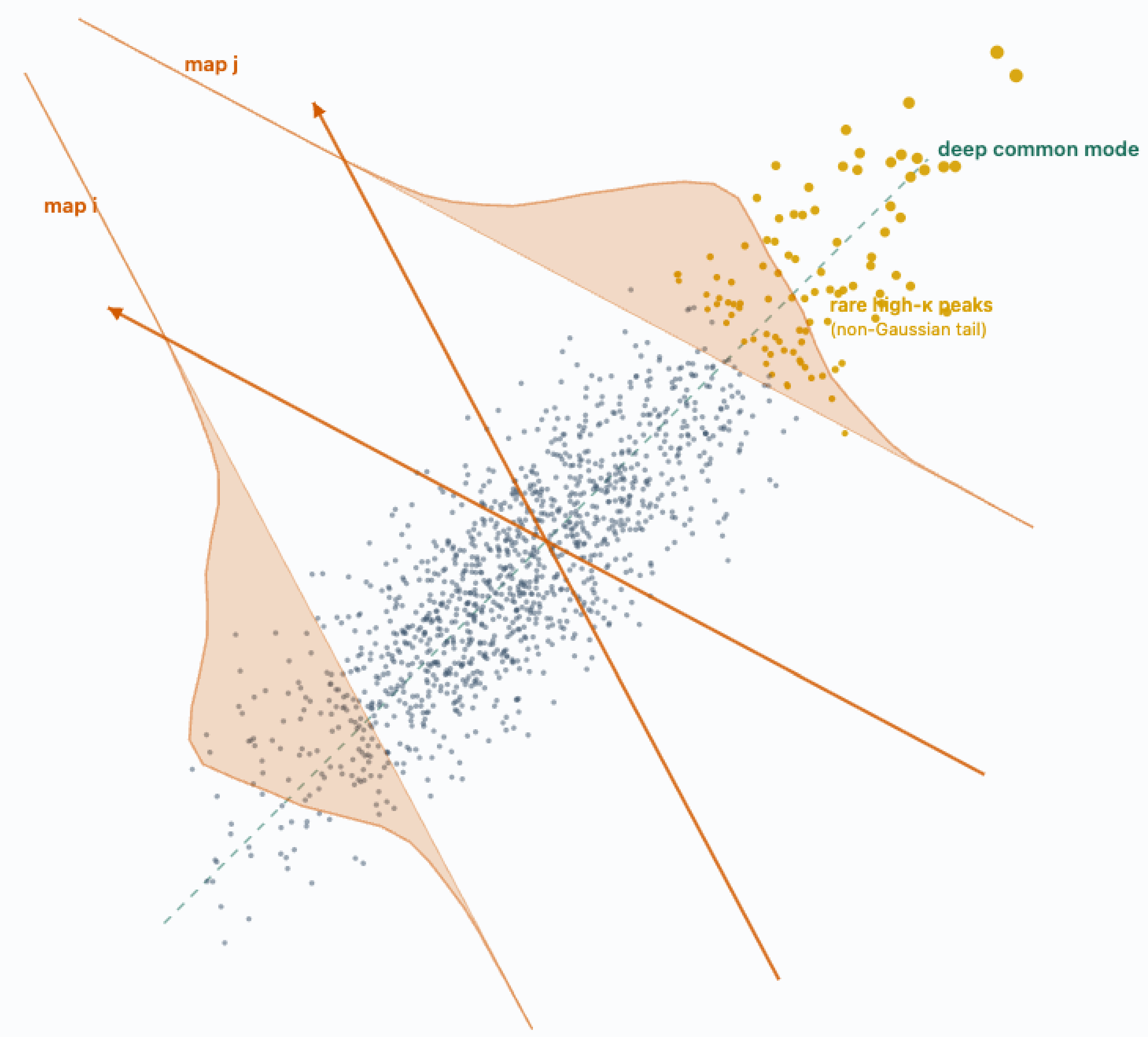




1/5 Each pixel is a point in 4-channel space; together the pixels form a **point cloud**. The **ℓ1-norm** of one map is its **projection** onto that axis (a 1-D marginal). The cloud is elongated along a **deep common mode**, with a non-Gaussian tail of **rare high-κ peaks**, so both projections are **rich**.

Same information, a different frame: why BNT collapses the ℓ1-norm but not the CNN

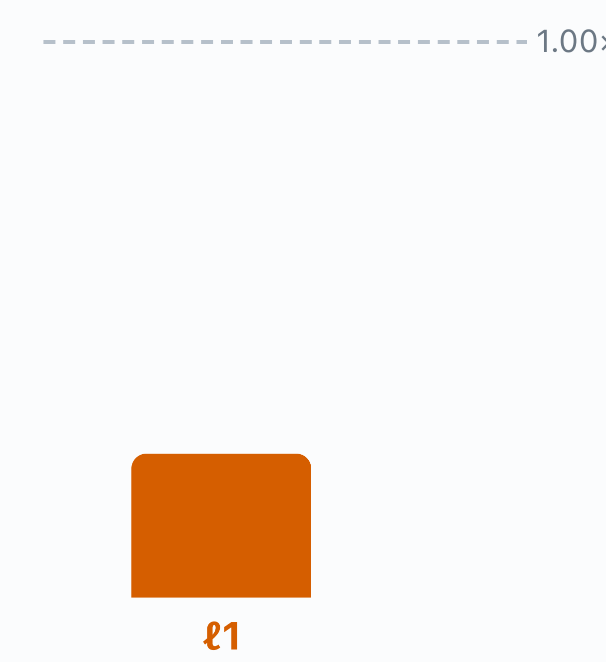
↻ replay



the point cloud (all the information) never moves

FoM_3 / no-BNT

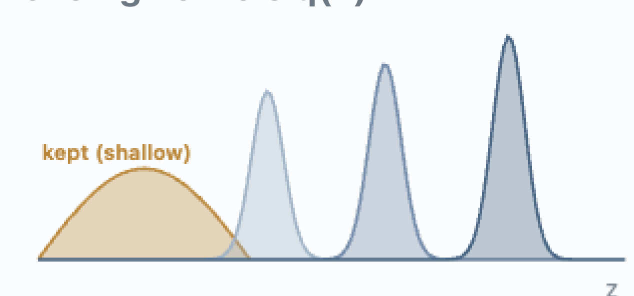
0.26×



3045 → 779

$\sigma(\sigma_8)$ +65% (calibrated loss)

lensing kernels q(z)

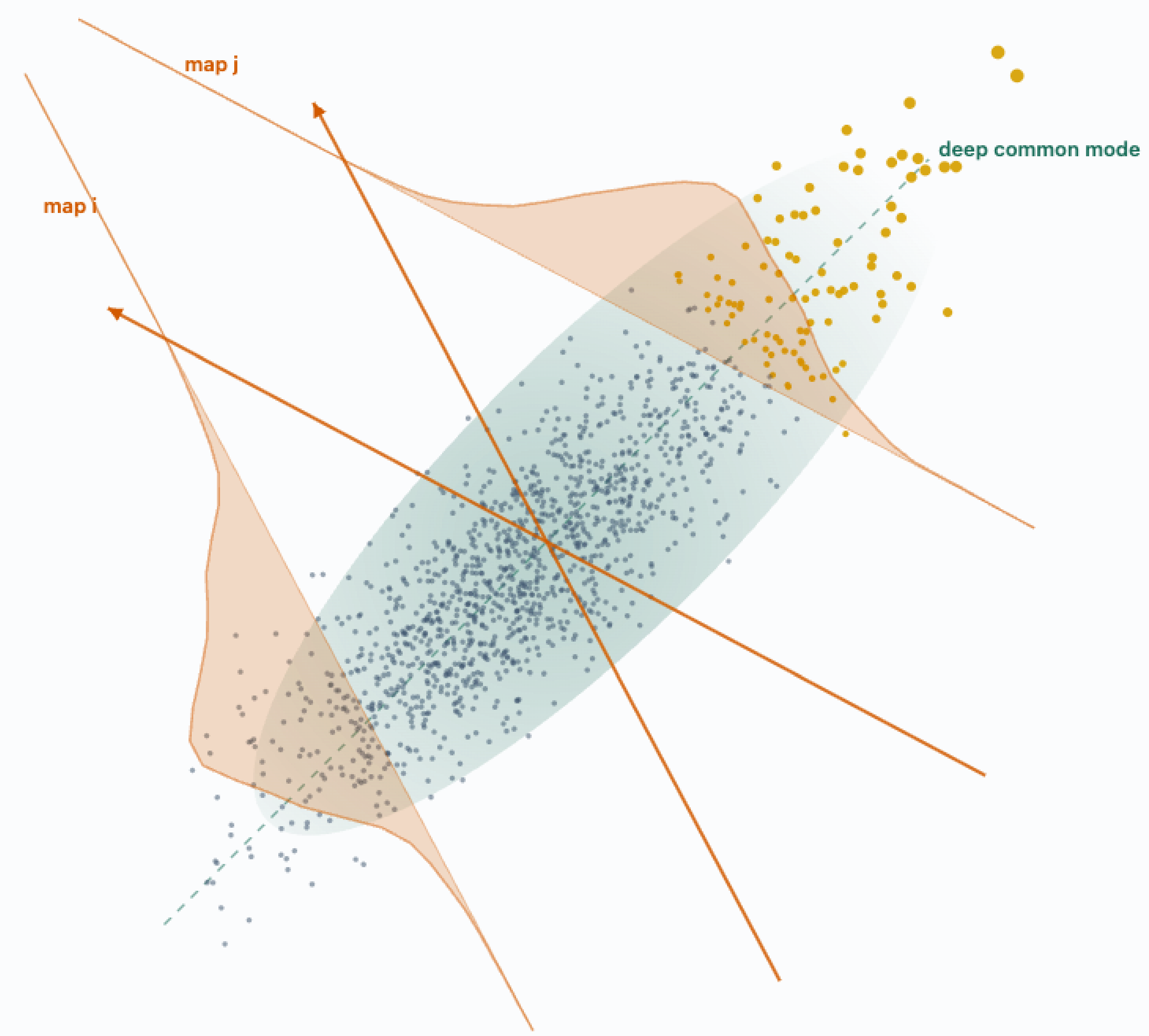


1 shallow map + 3 thin lens-z slices (signal-poor)

2/5 **BNT re-orients the measuring axes** off the deep mode, onto thin, signal-poor slices (with amplified, correlated noise). The projections **flatten toward noise** and FoM_3 collapses to 0.26×. **The cloud has not moved.**

Same information, a different frame: why BNT collapses the ℓ1-norm but not the CNN

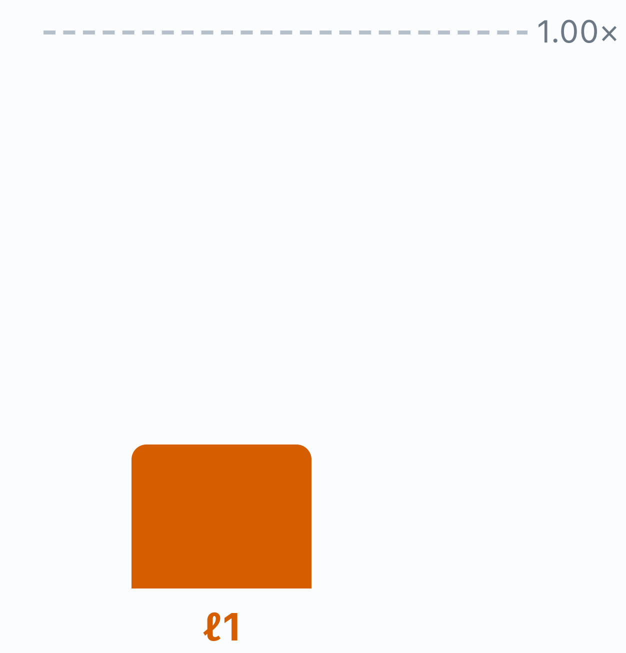
↻ replay



the point cloud (all the information) never moves

FoM_3 / no-BNT

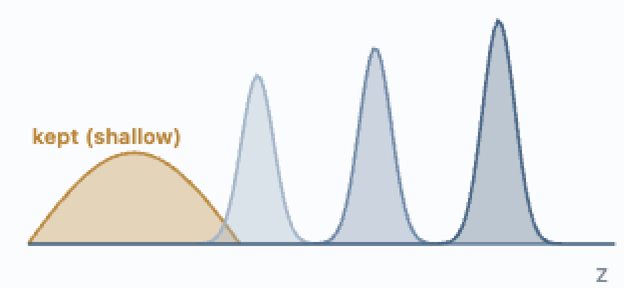
0.26×



3045 → 779

info is intact, just off-axis

lensing kernels q(z)

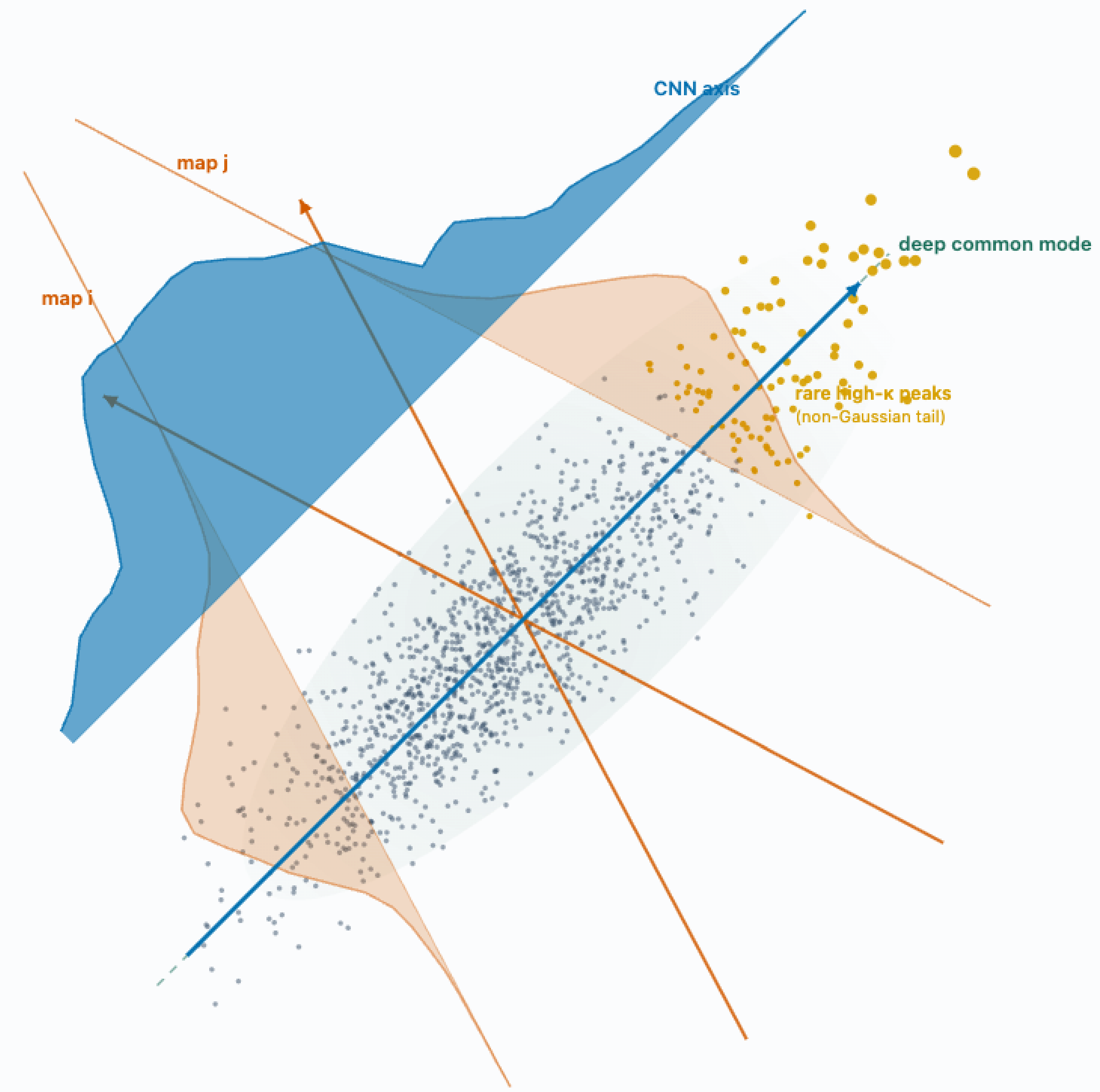


1 shallow map + 3 thin lens-z slices (signal-poor)

3/5 So where did it go? The cosmology lives in the cloud's **shape**, the **relations between maps**. In this frame, no single-map projection can see it.

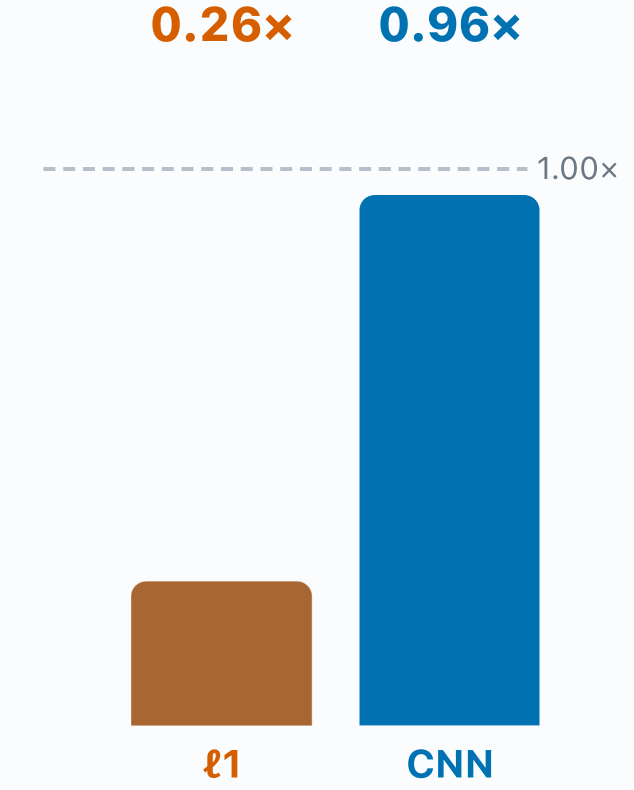
Same information, a different frame: why BNT collapses the ℓ1-norm but not the CNN

↻ replay



the point cloud (all the information) never moves

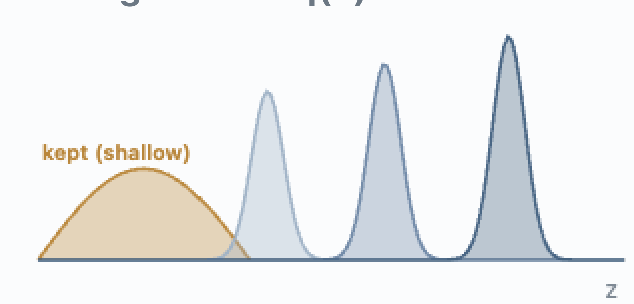
FoM_3 / no-BNT



3326 → 3186 · CNN

channel-mixing ⇒ frame-agnostic

lensing kernels q(z)

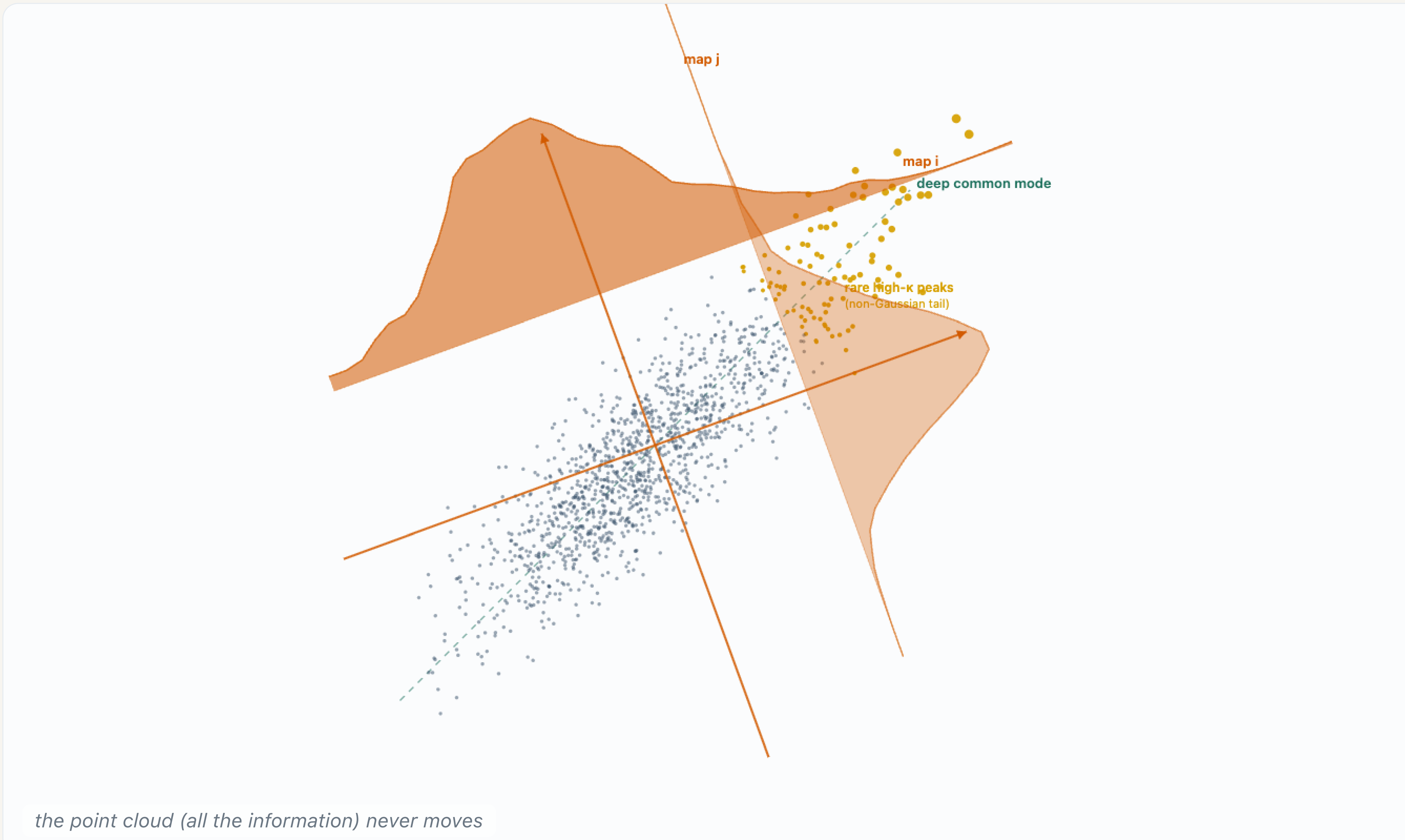


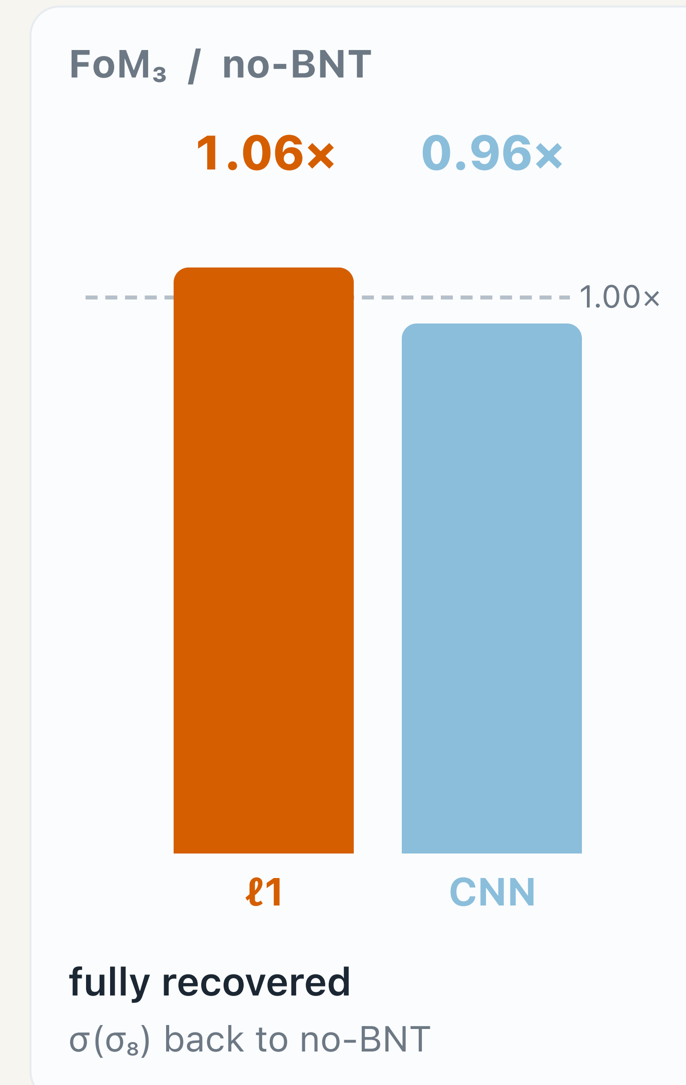
1 shallow map + 3 thin lens-z slices (signal-poor)

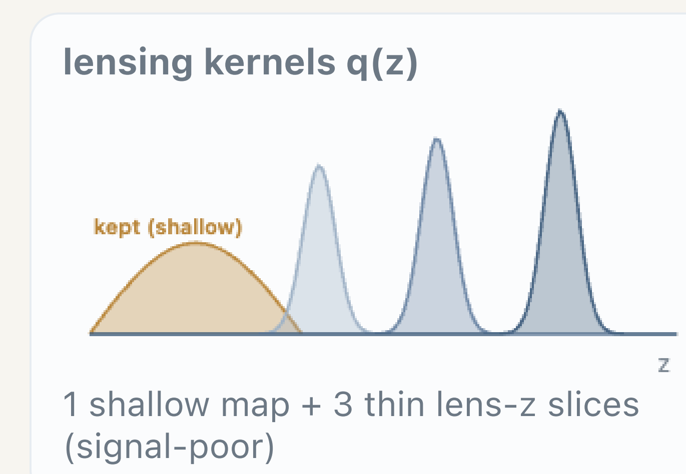
4/5 The **CNN mixes channels first**, so it can draw **its own axis back along the cloud** (undo B for free) and read it richly again, $FoM_3 \approx 0.96\times$. **Basis-robust, not "smarter."**

Same information, a different frame: why BNT collapses the ℓ1-norm but not the CNN

↻ replay



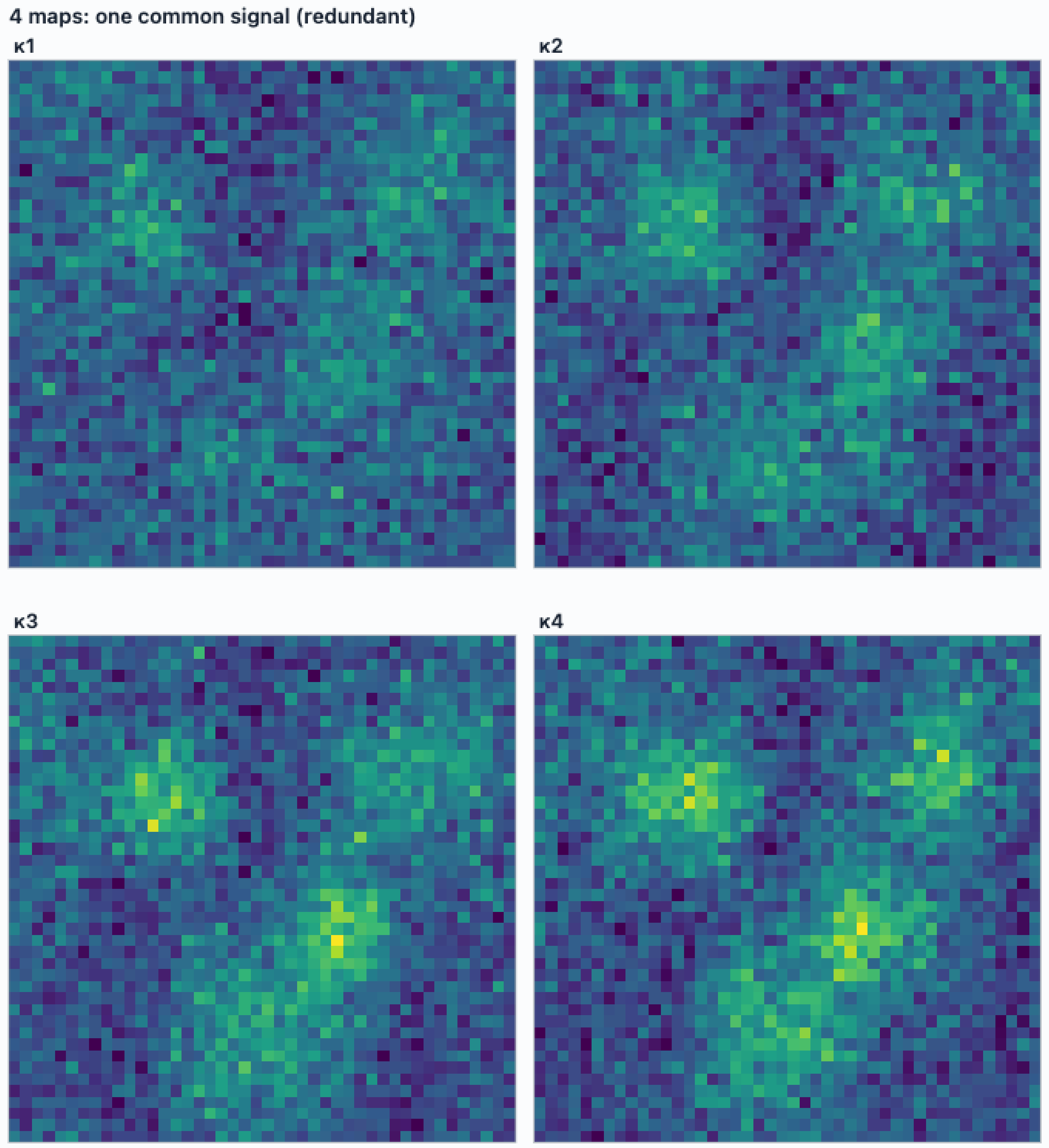


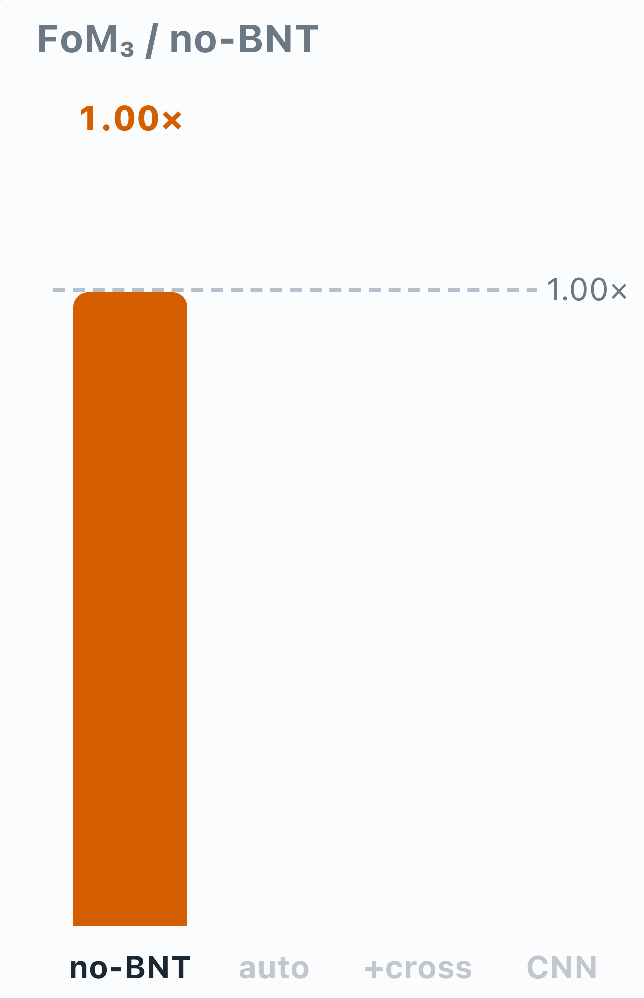


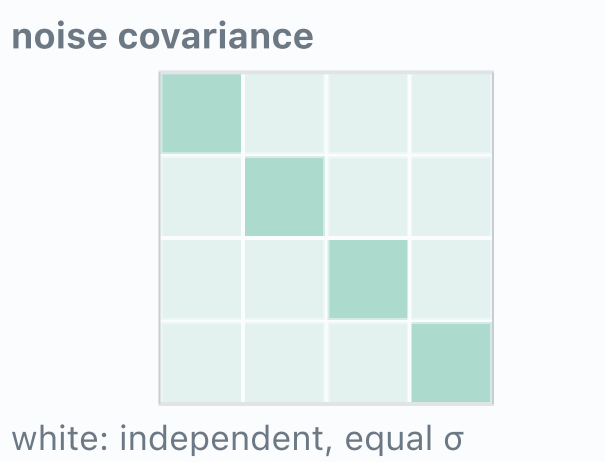
5/5 **Whitening** rotates to a **different clean frame** and the per-map projections come back, FoM_3 recovers to **1.06×**. Nothing was lost; the collapse was the **frame**.

Signal & noise under BNT: who can read it

replay



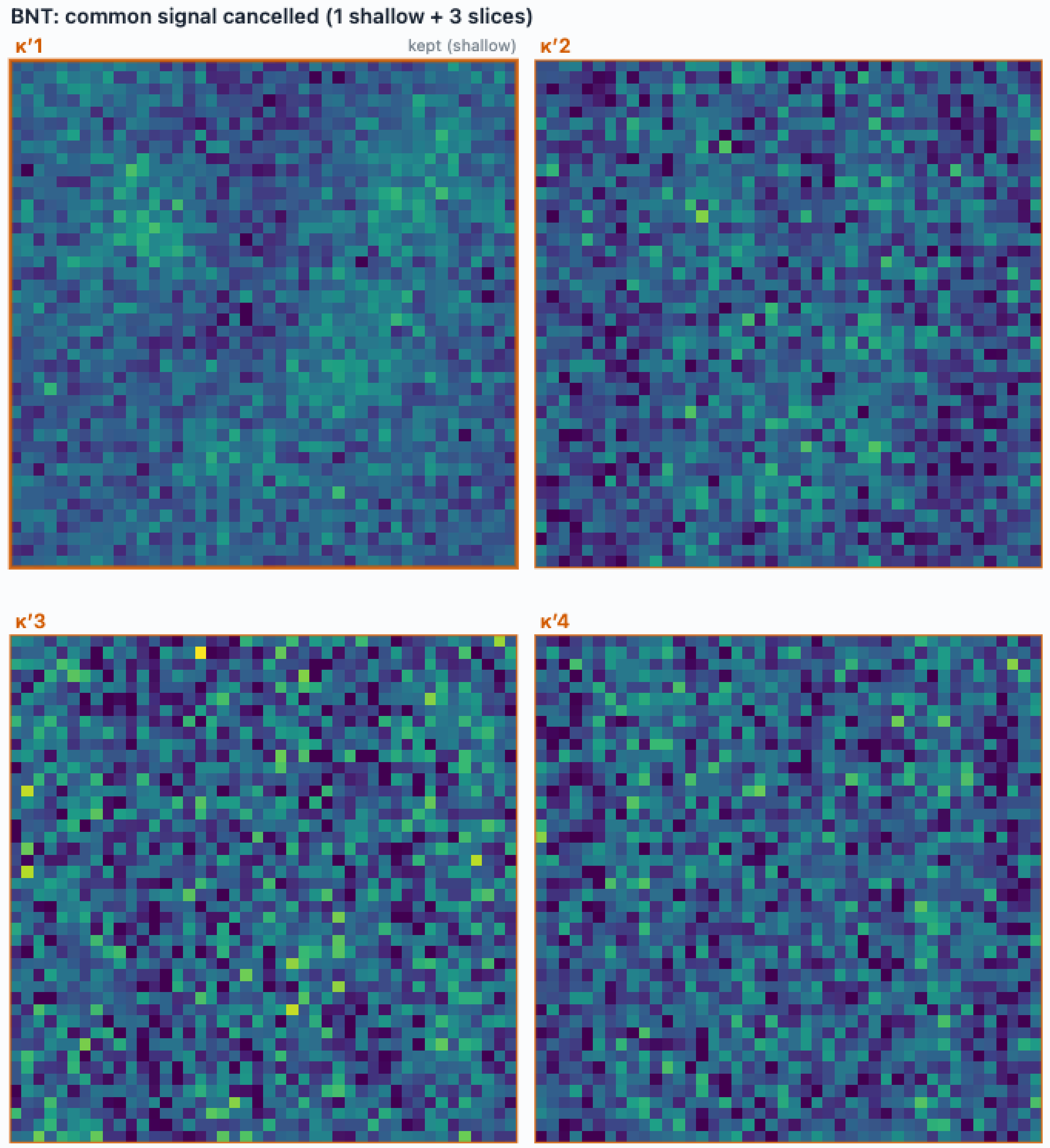


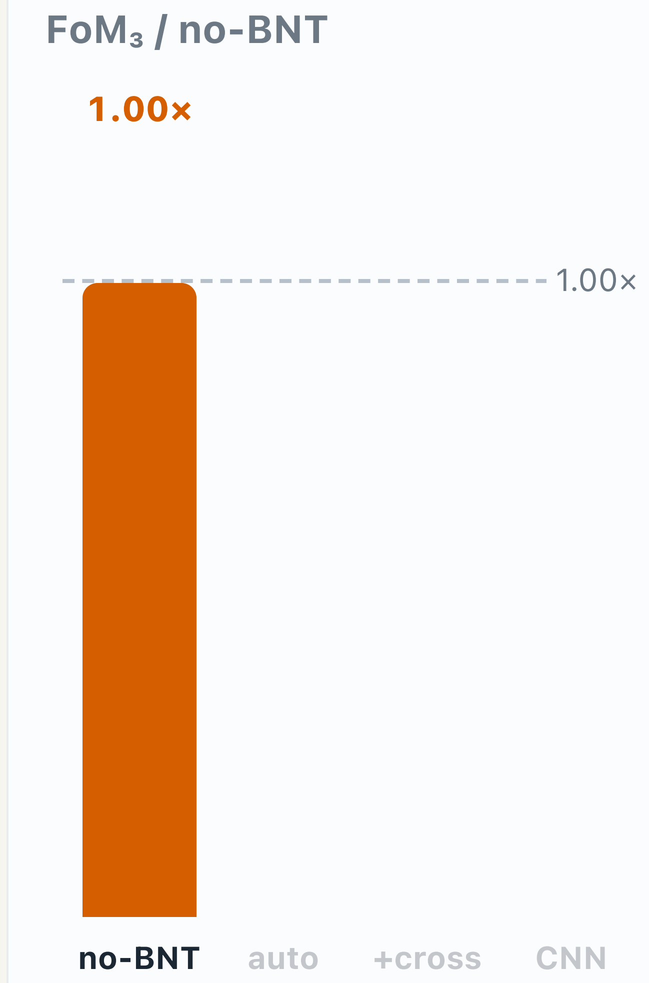


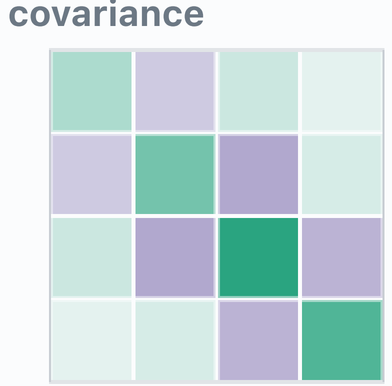
1/5 Four tomographic maps mostly share **one common signal** (plus small increments), so they are strongly **redundant**. Their shape noise is **independent**, equal in every bin.

Signal & noise under BNT: who can read it

↻ replay



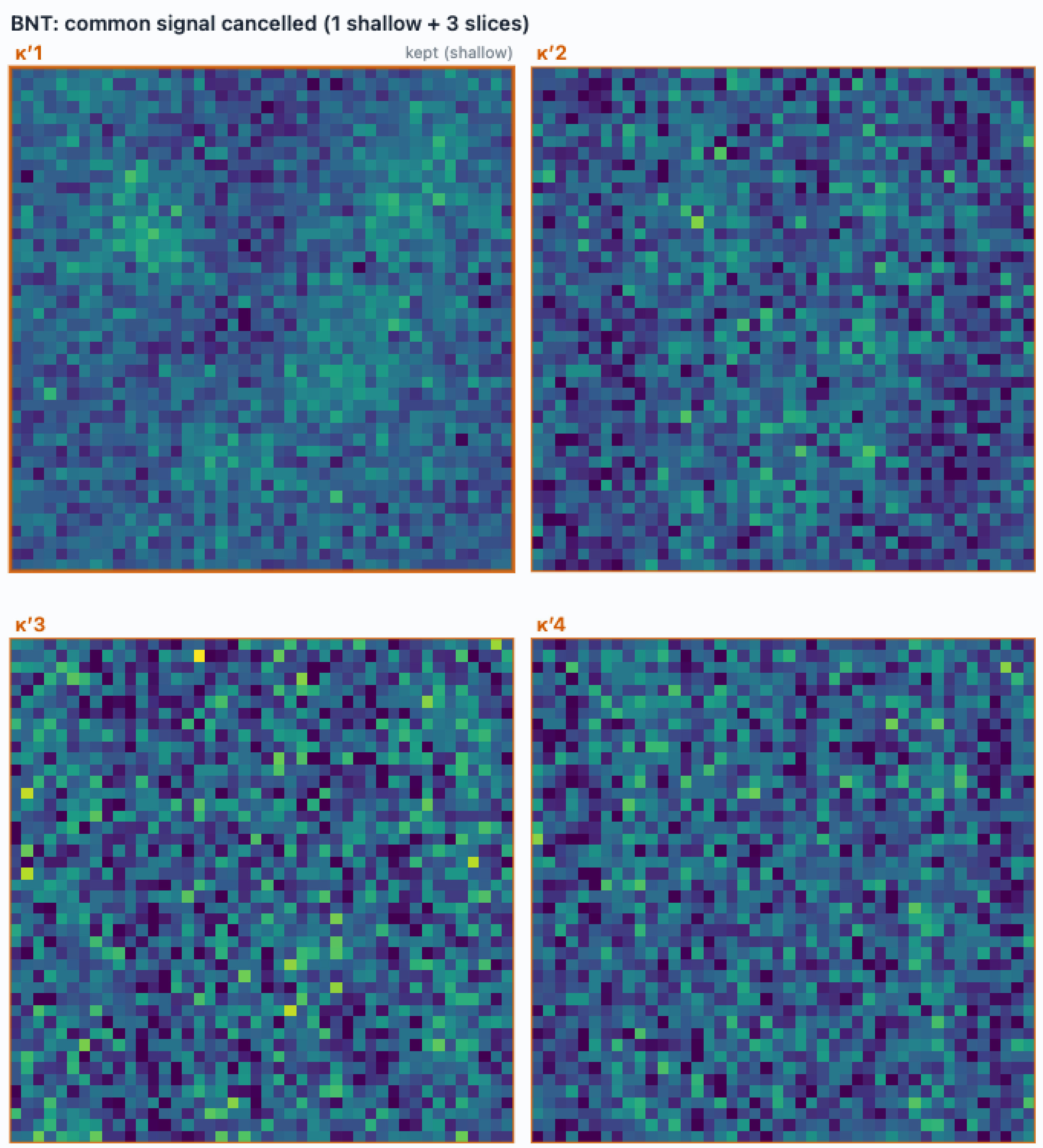


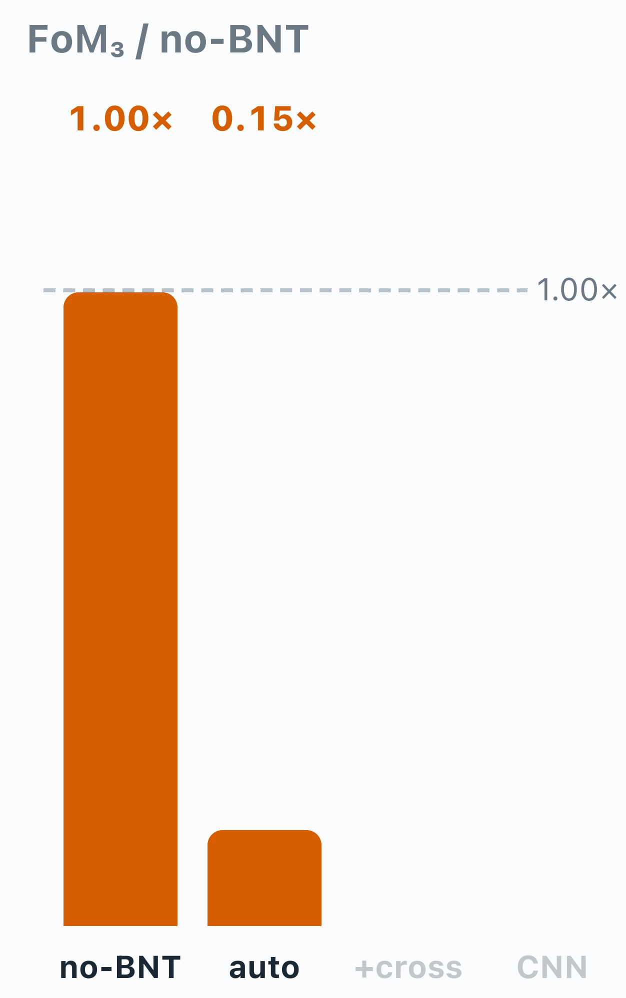


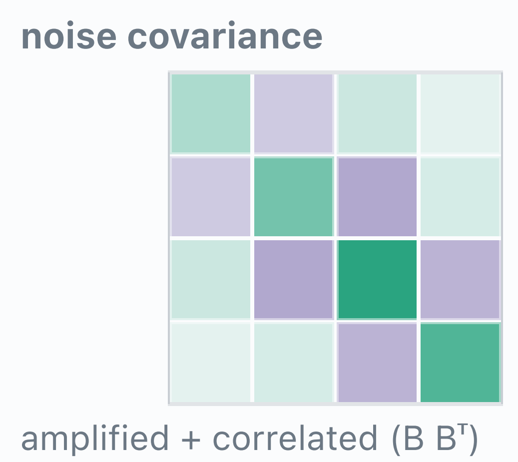
2/5 **BNT** differences them ($\kappa'_i = \Sigma_j B_{ij} \kappa_j$): the common signal **cancels**, leaving one shallow map and three thin **slices**. The structure is now **gone from each map**, not hidden inside it.

Signal & noise under BNT: who can read it

↻ replay



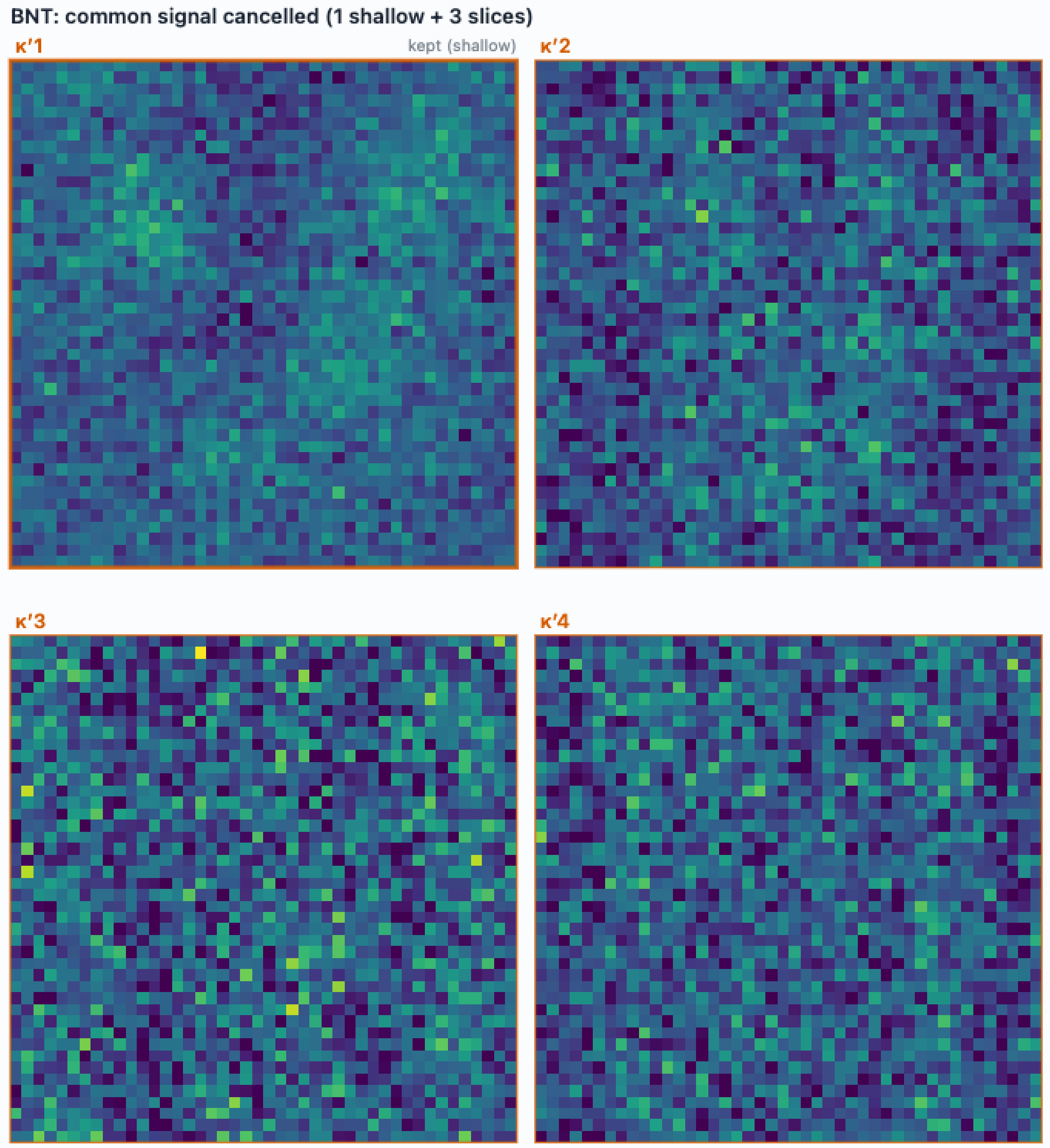


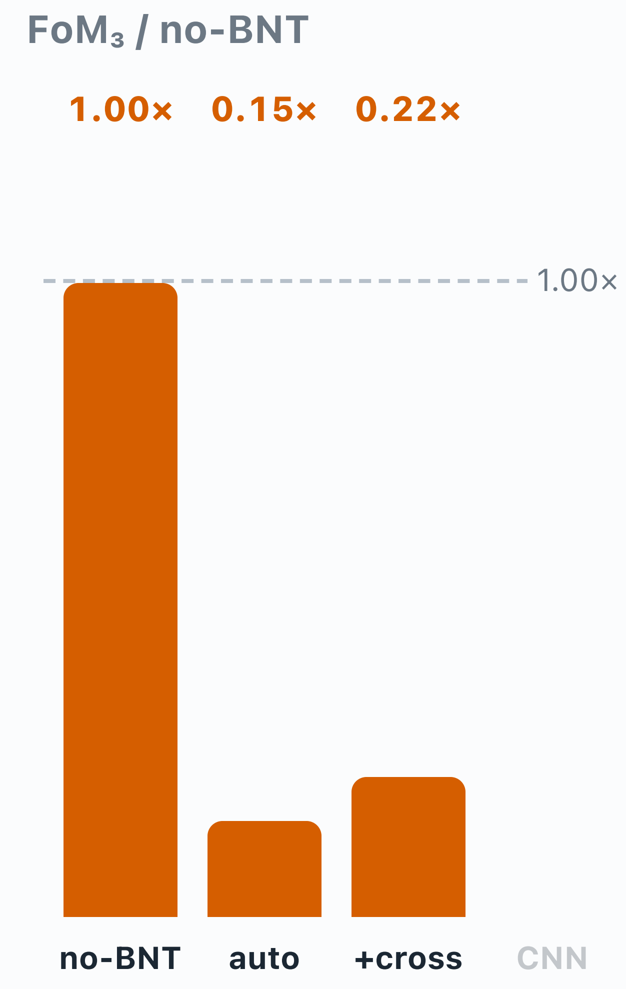


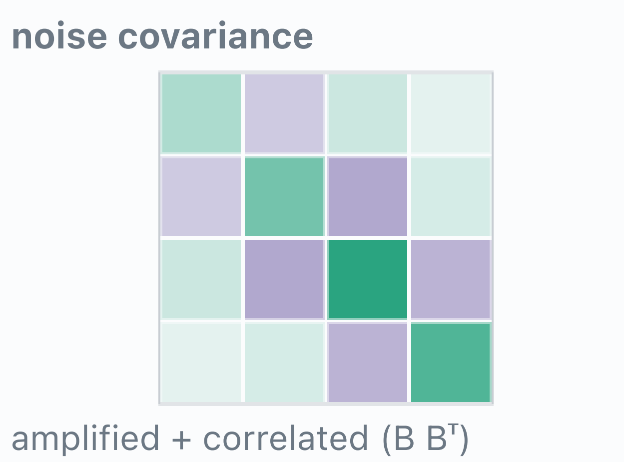
3/5 The noise is mixed too: **amplified** (×1, 1.4, 1.8, 1.6) and **correlated** (−0.71). Each nulled map is tiny signal under big noise, so a per-map **ℓ1** sees almost nothing (**0.15×**).

Signal & noise under BNT: who can read it

replay



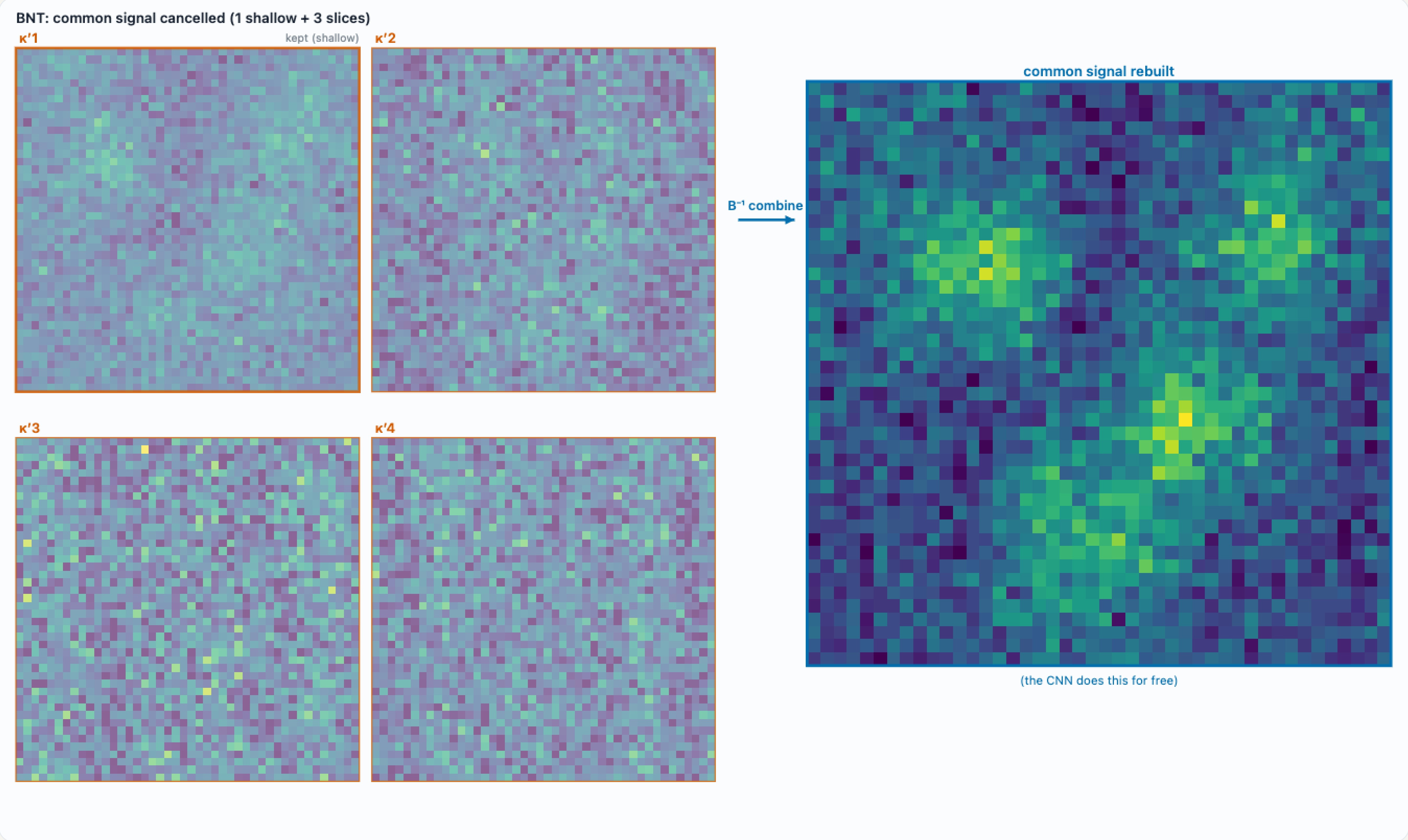


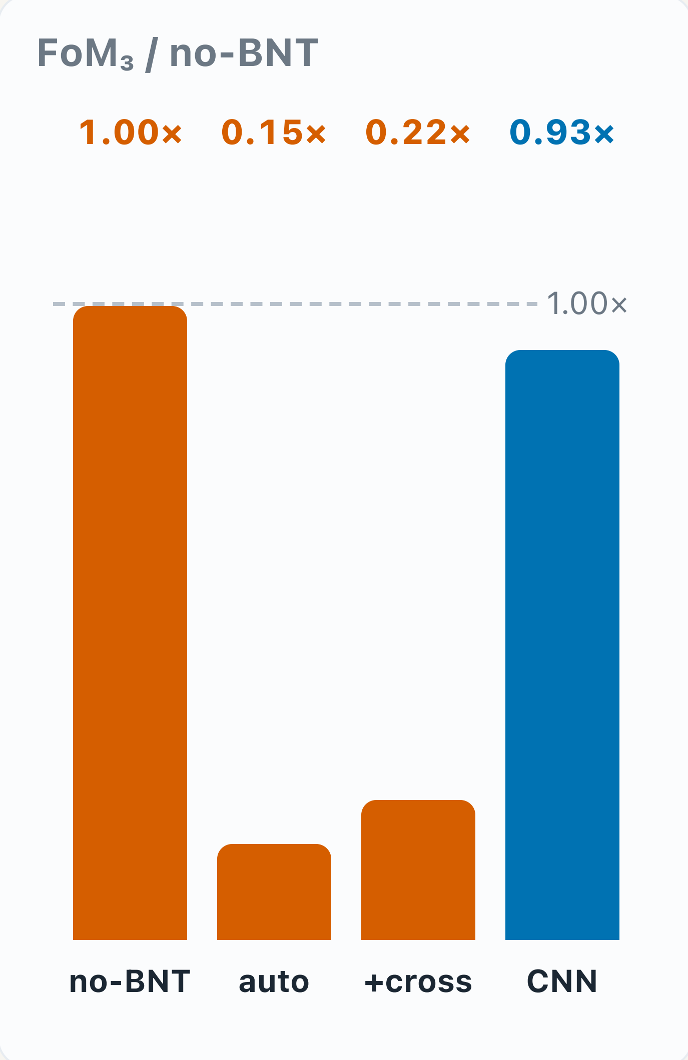


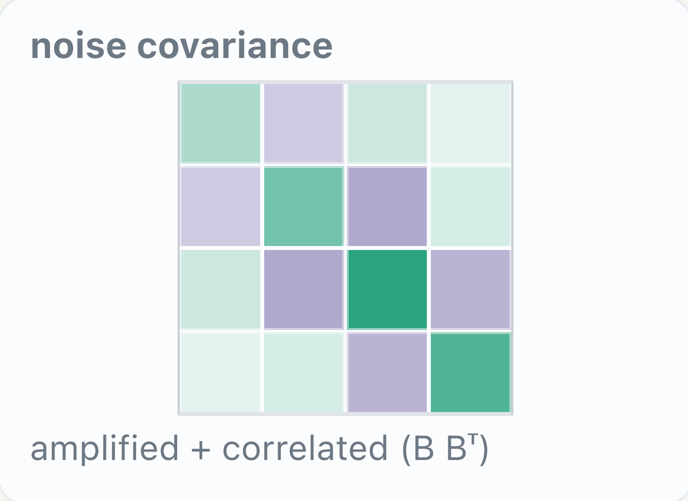
4/5 **Cross maps** help only partly (**0.22×**): they are fixed **pairwise** products (product = zero-lag joint, conv = 2-point), recovering the pairwise share, not the higher-order multi-bin info. The ℓ1 is **nonlinear**, so no finite set of products is complete; capturing all of it needs the full tree of higher moments (triplets, quadruplets, ...).

Signal & noise under BNT: who can read it

↻ replay



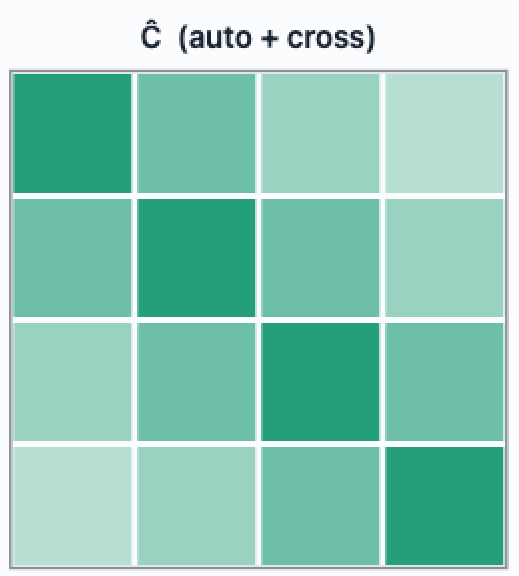




5/5 But BNT was **linear**, so you need none of that tree: **recombine** the maps ($\kappa = B^{-1}\kappa'$) and the common signal **returns**. The **CNN** does exactly this in its first layer, for free (**0.93×**); any clean frame (**whitening**) gives **1.06×**. The information was in the **joint**, not any single map.

What survives BNT: the 2-point rule

replay

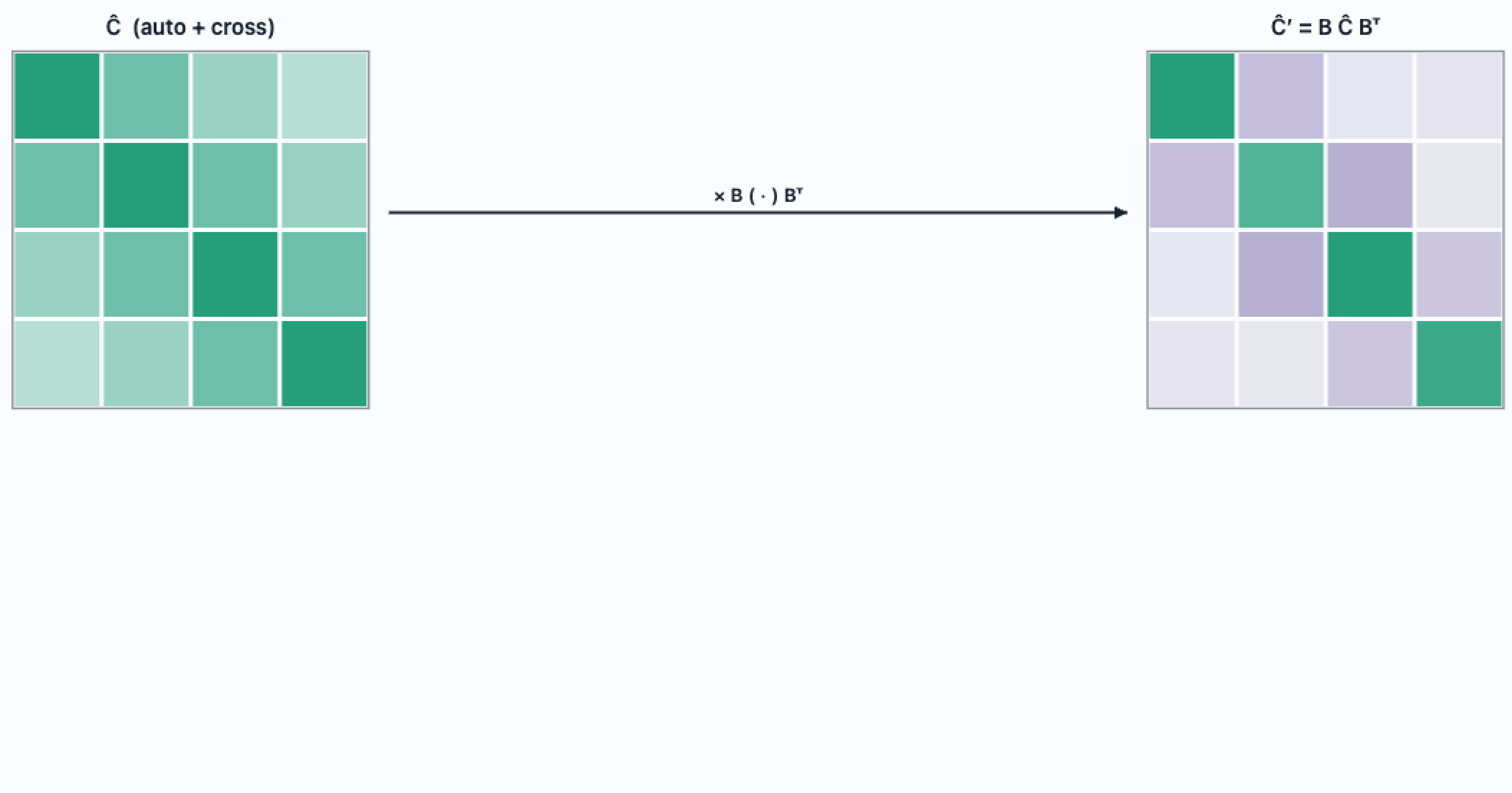


teal + / purple –

1/5 The **two-point** information is the full set of **auto- and cross-spectra**, a 4×4 matrix **Ĉ**.

What survives BNT: the 2-point rule

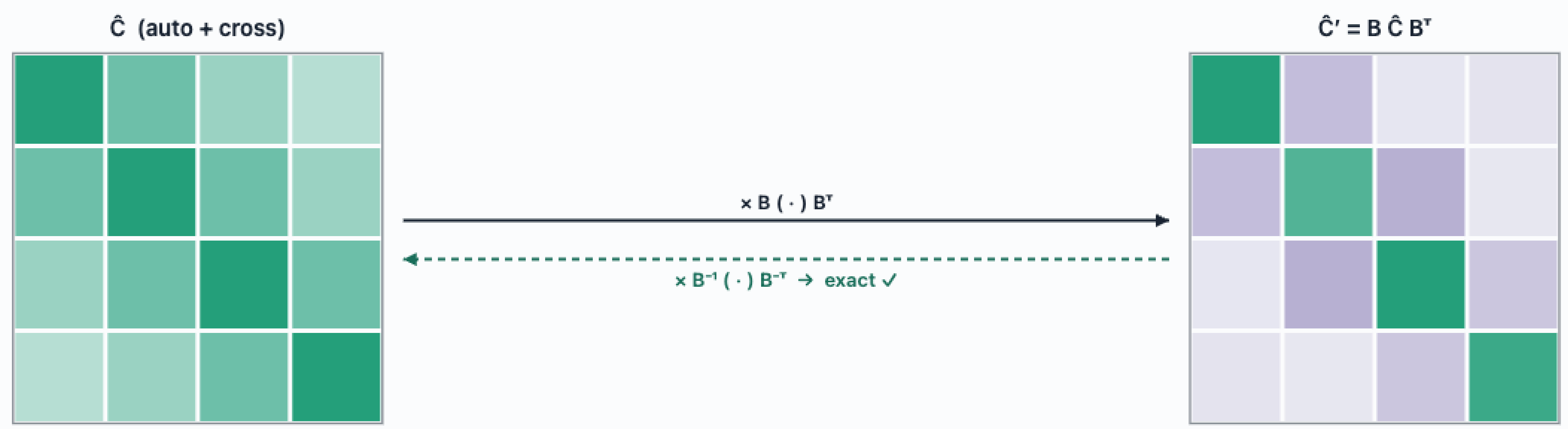
↻ replay



2/5 BNT is linear: it sends **$\hat{C} \to B\,\hat{C}\,B^{T}$**. A different matrix, but a **known, invertible** map.

What survives BNT: the 2-point rule

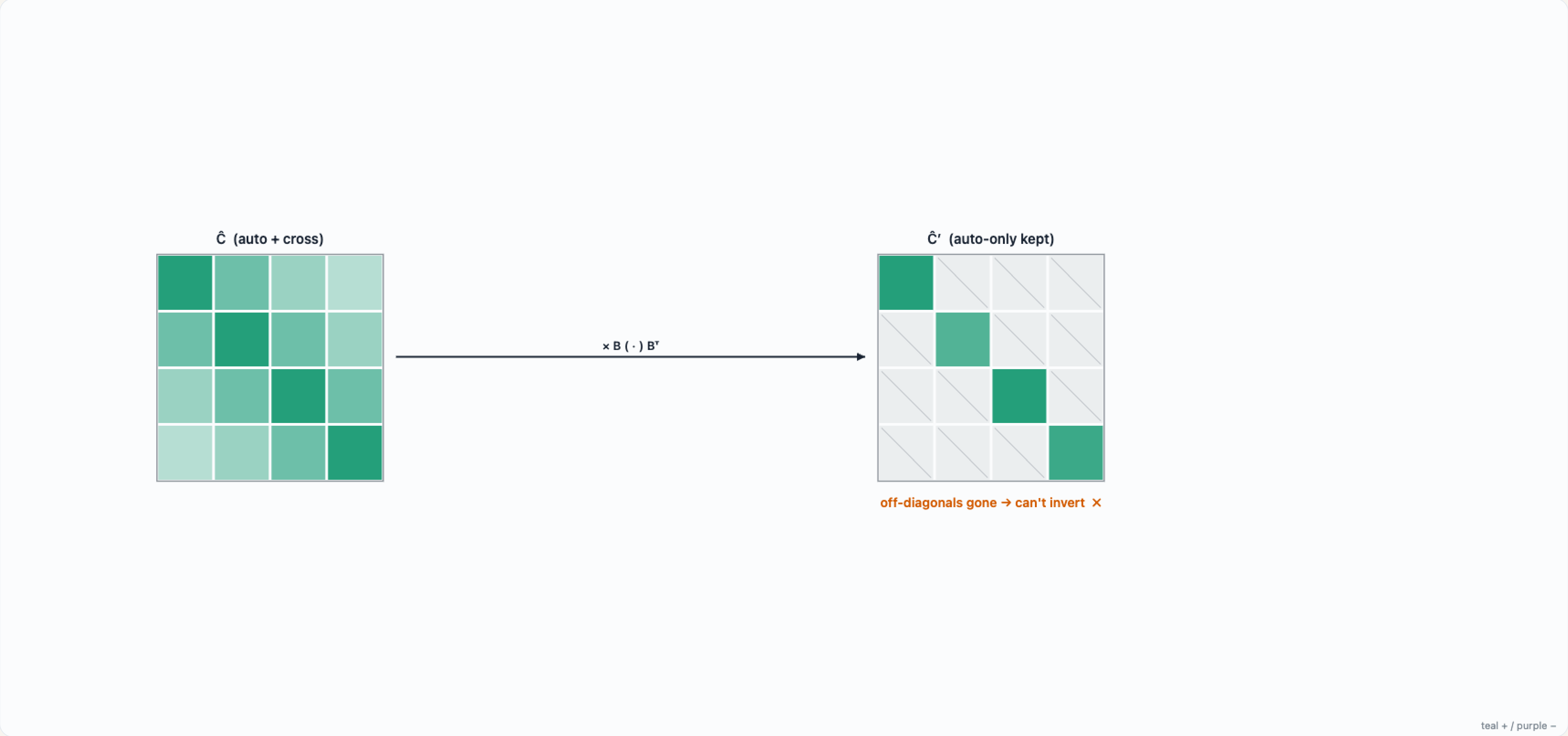
replay



3/5 So multiply back: $\hat{C} = B^{-1}\,\hat{C}'\,B^{-T}$. **Nothing is lost**: **auto+cross power spectra are exactly BNT-invariant** (identical posteriors).

What survives BNT: the 2-point rule

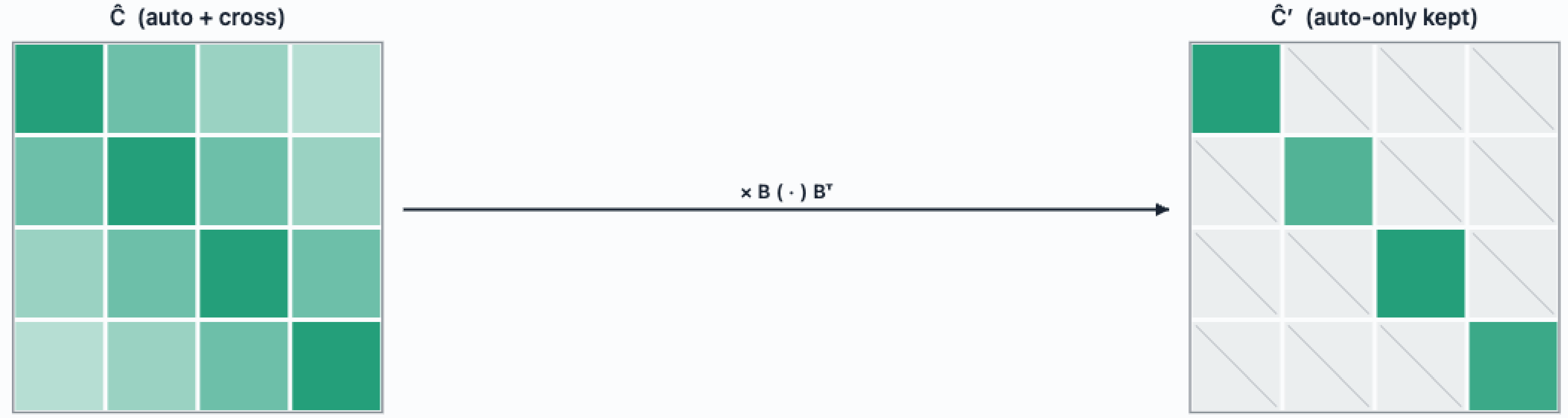
↻ replay



4/5 Keep only the **diagonal** (auto-spectra alone) and you discard the off-diagonals, so **you cannot rebuild Ĉ**. Auto-only 2-pt fails by **incompleteness** (cheap to fix: add the cross-spectra back).

What survives BNT: the 2-point rule

↻ replay



teal + / purple –

5/5 **The rule:** survives BNT ⇔ you can reassemble it. **Auto+cross 2-pt: yes**, it is **linear and complete** (B just relabels it, $\hat{C} \mapsto B\hat{C}B^{T}$). The **ℓ1** / peaks: **no**, because B mixes the maps and then the histogram throws away the joint (a **mix-then-marginalize**); no per-channel set closes. That is the collapse.

§2 Is the +7% worth it? The honest cost of going neural (an open question for the round table)

THE BENEFIT

~7%

FoM3 gain of the optimized CNN over the analytical ℓ1, both calibrated.

THE COST

- extensive architecture and hyperparameter search
- a very large dataset (899 cosmologies, ~324k maps); VMIM needs the scale or it biases
- unphysical-information traps (patch geometry, map-mean / mass-sheet mode, 20° projection features) that tighten contours dishonestly
- **and these largely escape TARP and SBC**: the contours look calibrated and are still wrong

THE THUMB ON THE SCALE

- ℓ1 is simple, interpretable, inspectable; CNNs are powerful but treacherous
- Where the CNN earns its keep: **BNT**, the channel-mixing win
- For the panel: is a ~7% gain worth that cost and that risk?

§3 Do baryons break HOS? No.

PART 1 · BARYONS

Usable non-Gaussian information persists on baryon-safe scales (the **ℓ1-norm beats P(k) ~3×**), cleaned by a single scale cut.

PART 2 · LEARNED VS ANALYTICAL

The hand-built ℓ1-norm **nearly matches the optimal learned summary (~7%)**, and both are calibrated.

BNT

The apparent BNT break is a **frame artifact**: a channel-mixing compressor, or one fixed rotation, recovers it.

THE BIGGER QUESTION

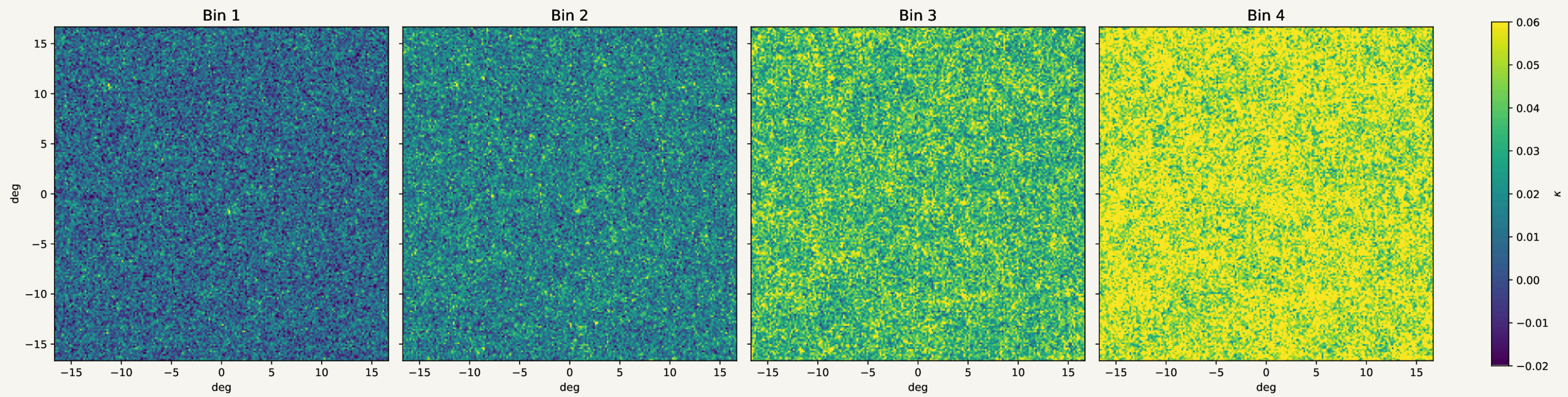
And can we trust higher-order weak lensing? **Getting there**...

APPENDIX

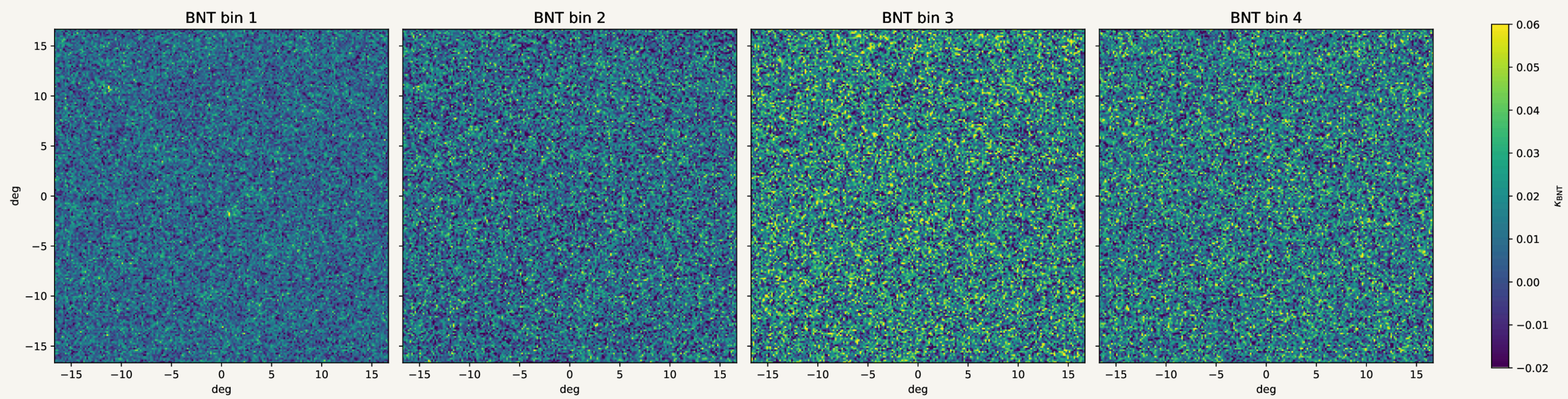
Backup

Supporting slides, for questions.

§1 You can see it in the maps: BNT trades deep signal for amplified, correlated noise

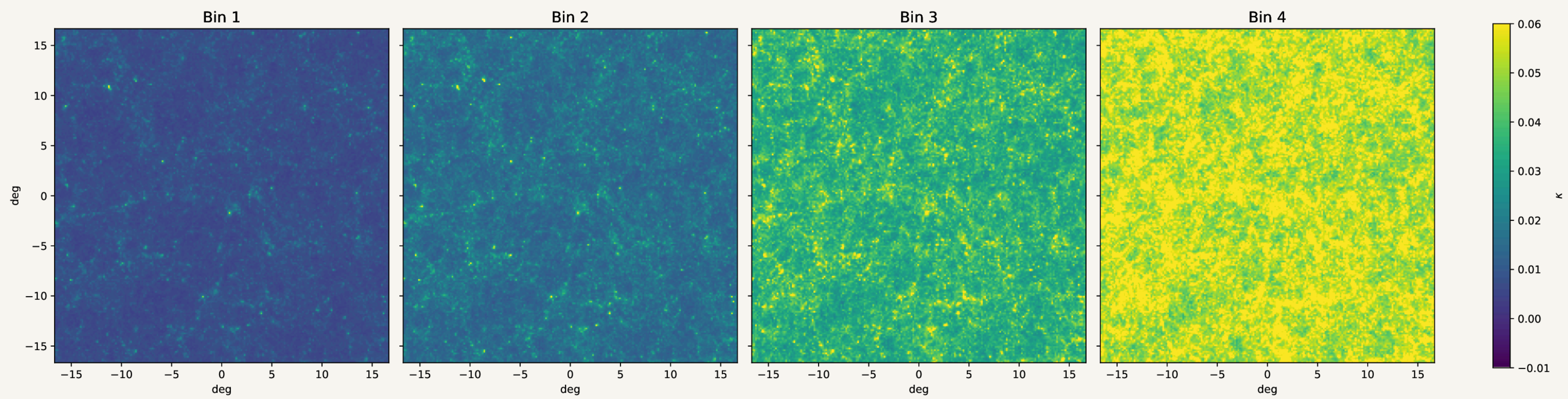


before BNT

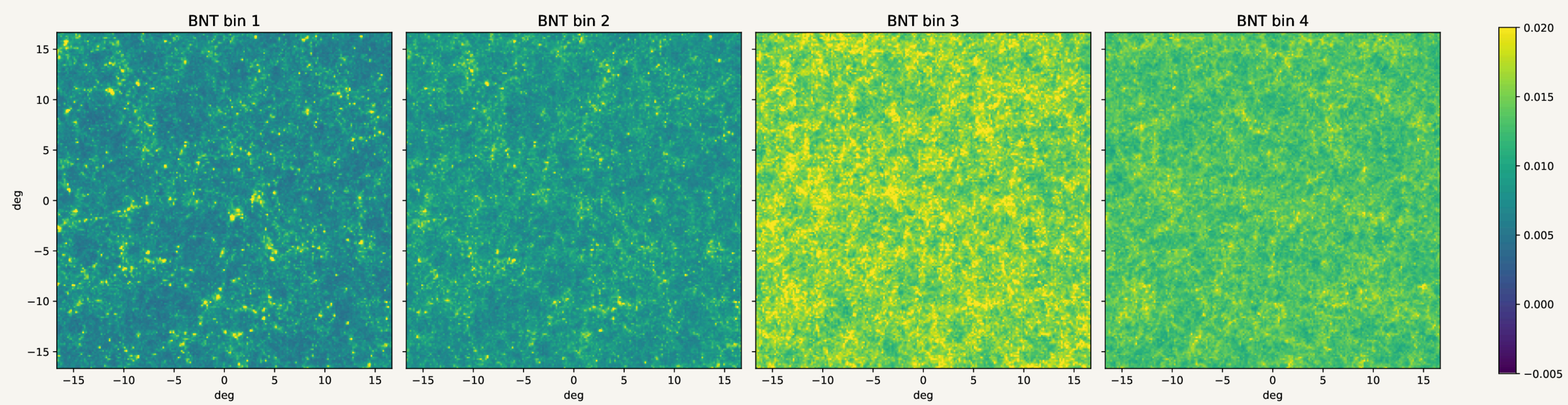


after BNT: the SNR collapses

§1 The same maps, noiseless: BNT cleanly redistributes the signal



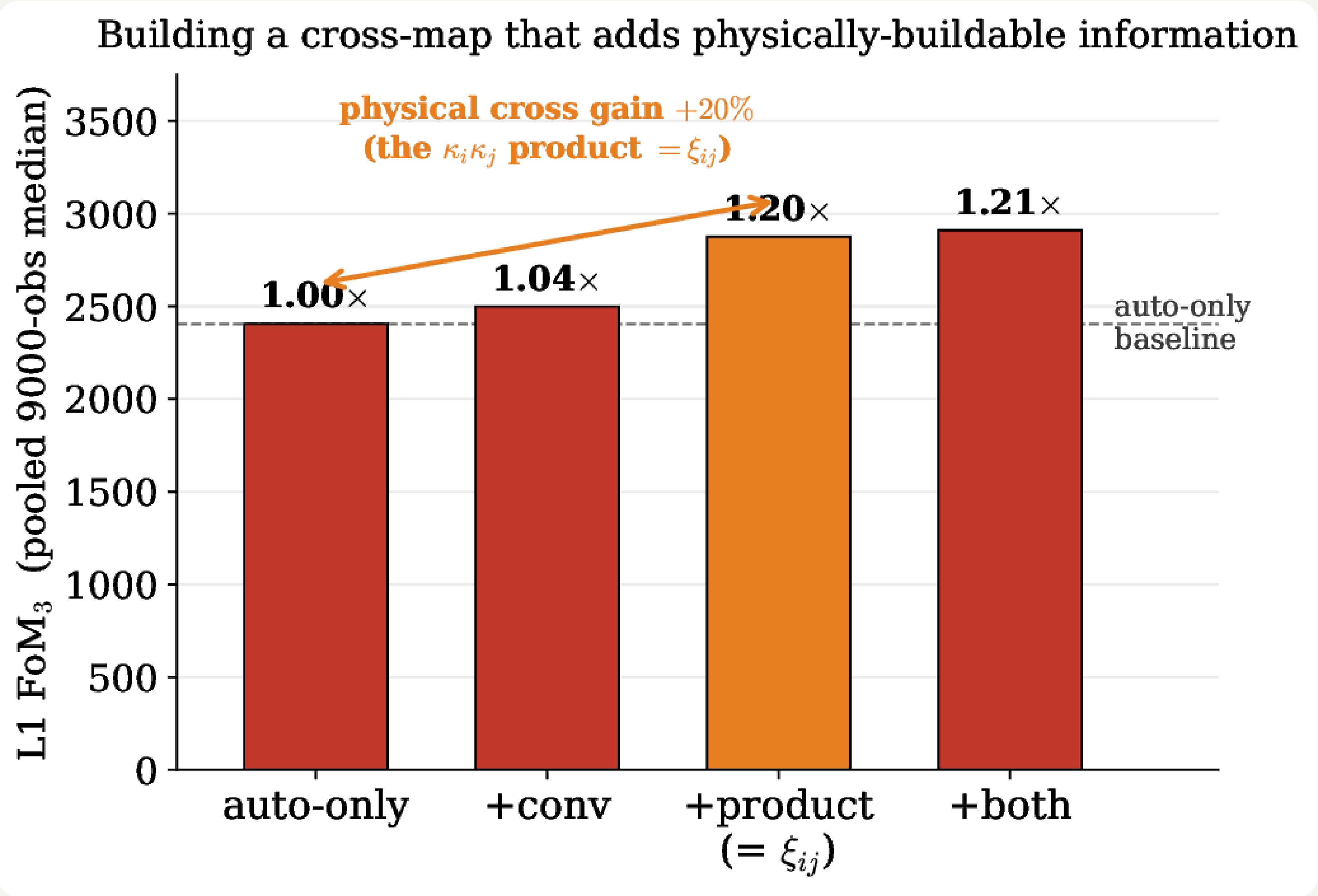
before BNT (noiseless)



after BNT (noiseless)

> Without shape noise, BNT is a clean, invertible redistribution of the signal (the deep common mode becomes one shallow map plus thin slices). The contour inflation comes from the correlated noise it introduces, not from any lost signal.

§2 Where the cross-bin information lives: the $\kappa_i\kappa_j$ product buys ~20% (and the full-sphere 4× was leakage)

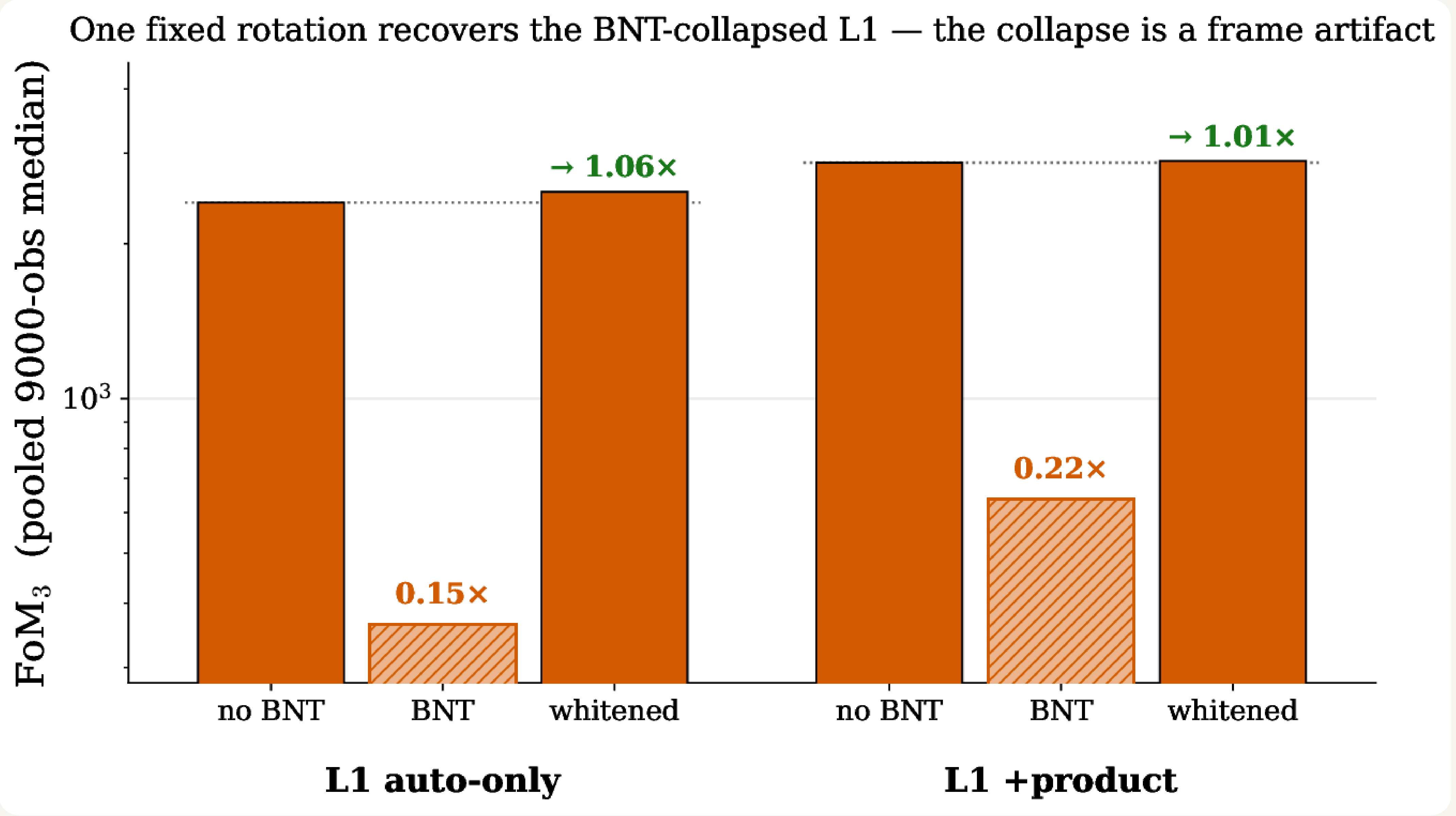


ADD THE CROSS-BIN PHYSICS CAREFULLY

- Product $\kappa_i\kappa_j$ (= ξ_{ij}): **+20%**
- A convolution buys ~0; both together, +21%
- The robust, physical cross-bin gain is **~+20%**

Community caution: a full-sphere cross construction would inflate this to ~4×, but ~92% of that is leakage (each cross-patch pixel is a global functional of the whole sky). We use only the physically buildable flat-sky arms.

§2 The clincher: a frame artifact, not lost information (one rotation recovers ℓ1, 1.06×)

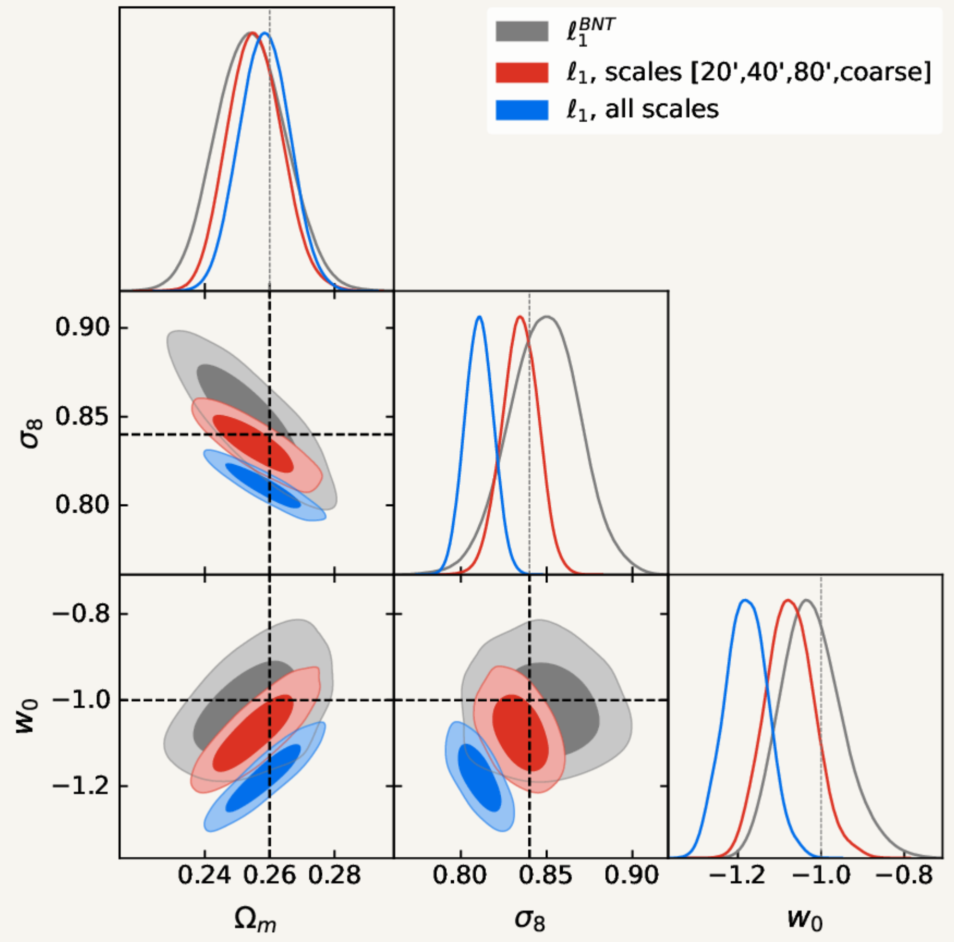


THE INFORMATION WAS NEVER LOST

- One fixed *whitening* rotation Q recovers the full no-BNT FoM3 for **ℓ1** too (**1.06×**)
- The collapse is a per-channel *frame* artifact: mix the bins, or re-rotate once
- Confirms the intuition block; closes the Vinciguerra loop

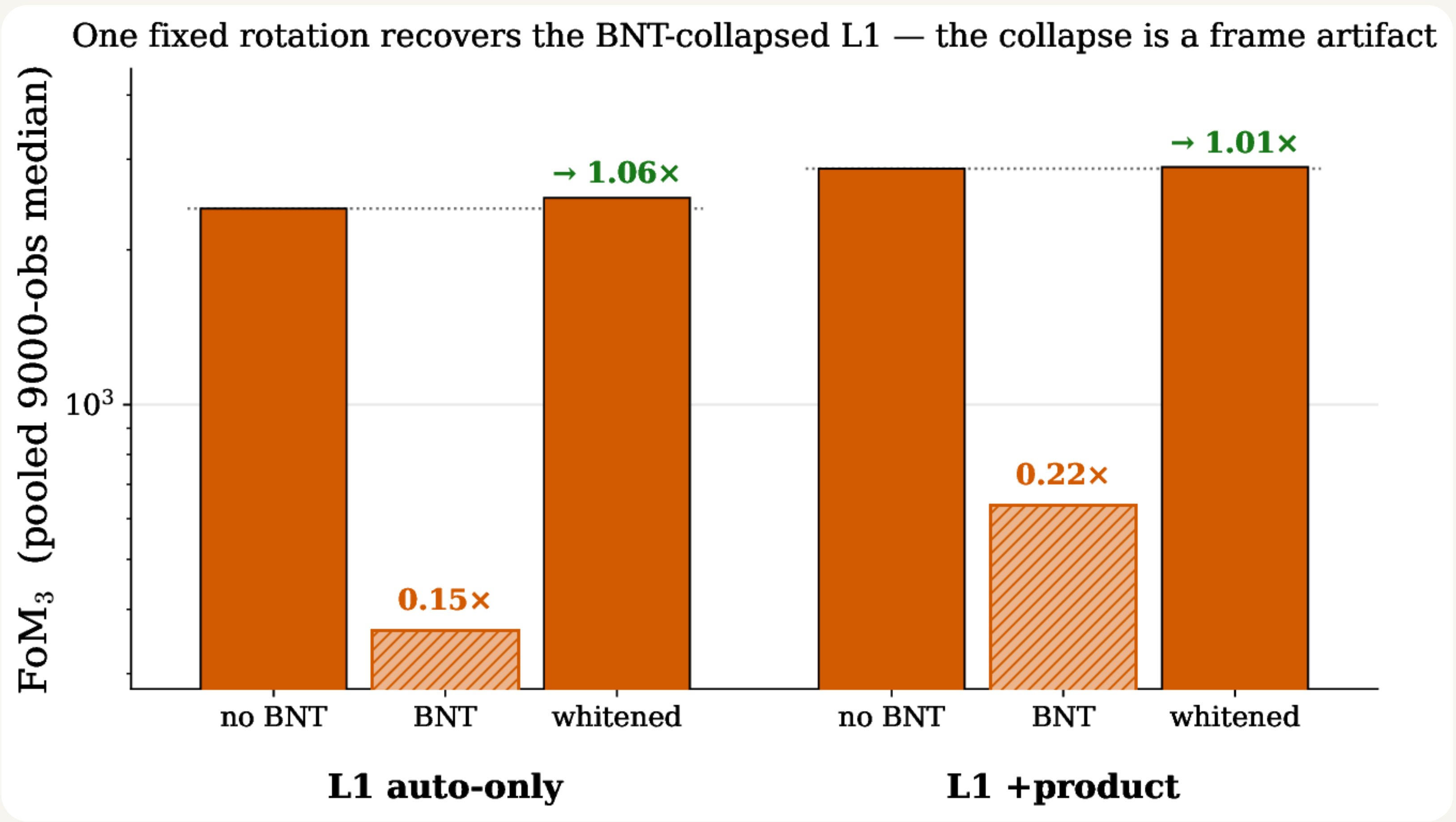
Closes the Vinciguerra loop: their forecast said recovering the BNT SNR for HOS is "highly non-trivial"; here it is, in one fixed rotation. A frames result: a one-point statistic's information content is basis-dependent, and BNT is simply a poor frame for a per-channel statistic.

§3 One story: the optimal tomographic strategy (BNT becomes viable once the summary mixes bins)



problem: the per-bin M1 inflates under BNT

→



resolution: mix the bins, or re-rotate

ONE ESCALATING STORY ABOUT CROSS-BIN INFORMATION

- P(k) → **ℓ1** (much more, even on safe scales) → **learned** (a bit more, calibrated)
- Per-bin statistics cannot access cross-bin info (break under BNT); a channel-mixing compressor can (BNT-lossless)
- BNT becomes viable once the summary mixes bins

Forward-looking: a route to baryon-robust, non-Gaussian SBI that keeps BNT's clean per-bin scale cuts without the contour-inflation tax. A next step, not a finished end-to-end measurement.