Weak lensing cosmology: beyond two-point correlations with the Dark Energy Survey (DES)

Bhuvnesh Jain
University of Pennsylvania

Primary collaborators:
Minsu Park, Kunhao Zhong, Rafael Gomes, Megan Zhao, Sunao Sugiyama, Marco Gatti, Karen Sarmiento; DES

Primary References:
Analytical methods: Zhong+ *2411.04759*; Park, Gatti, BJ *2409.02102*;
3pt and Higher Order Statistics: Gomes, Sugiyama, BJ+ (DES) *2503.03964*; Gatti+ *(DES) 2310.17557*, 2405.10881
Deep learning: Boruah, Jacob, BJ *2502.06687;* Zhong, Gatti, BJ *2403.01368*

- Weak lensing and the S8 tension
- Analytical vs Deep learning methods for mass mapping
- Results from three-point correlations
- Wavelets and Higher Order Statistics
- Topics in Simulation Based Inference (SBI)

The S_8 tension and cosmic puzzles

The S_8 tension: Growth of large-scale structure in the universe is slower than predicted using the baseline model of cosmology, ΛCDM, to extrapolate early time measurements with the CMB to the present.

How serious is the S_8 tension?

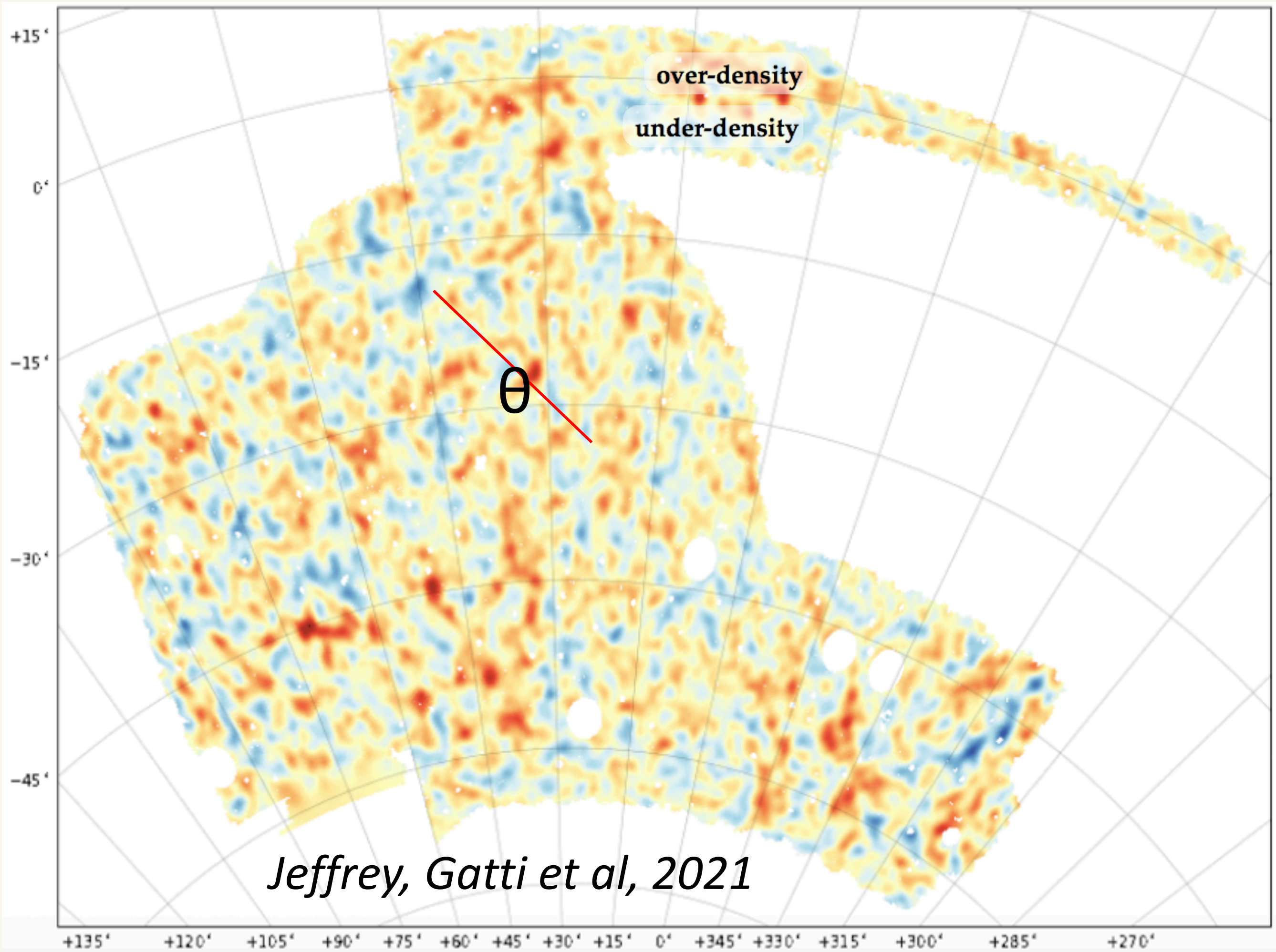
The S_8 tension shows up in multiple lensing, galaxy clustering and cross-correlation measurements but only at ~2-sigma. CMB lensing and DESI P(k) currently show no S_8 tension.

I will not discuss other possibly more interesting cosmic puzzles:

The H_0 tension: The universe is expanding faster than 'predicted'.

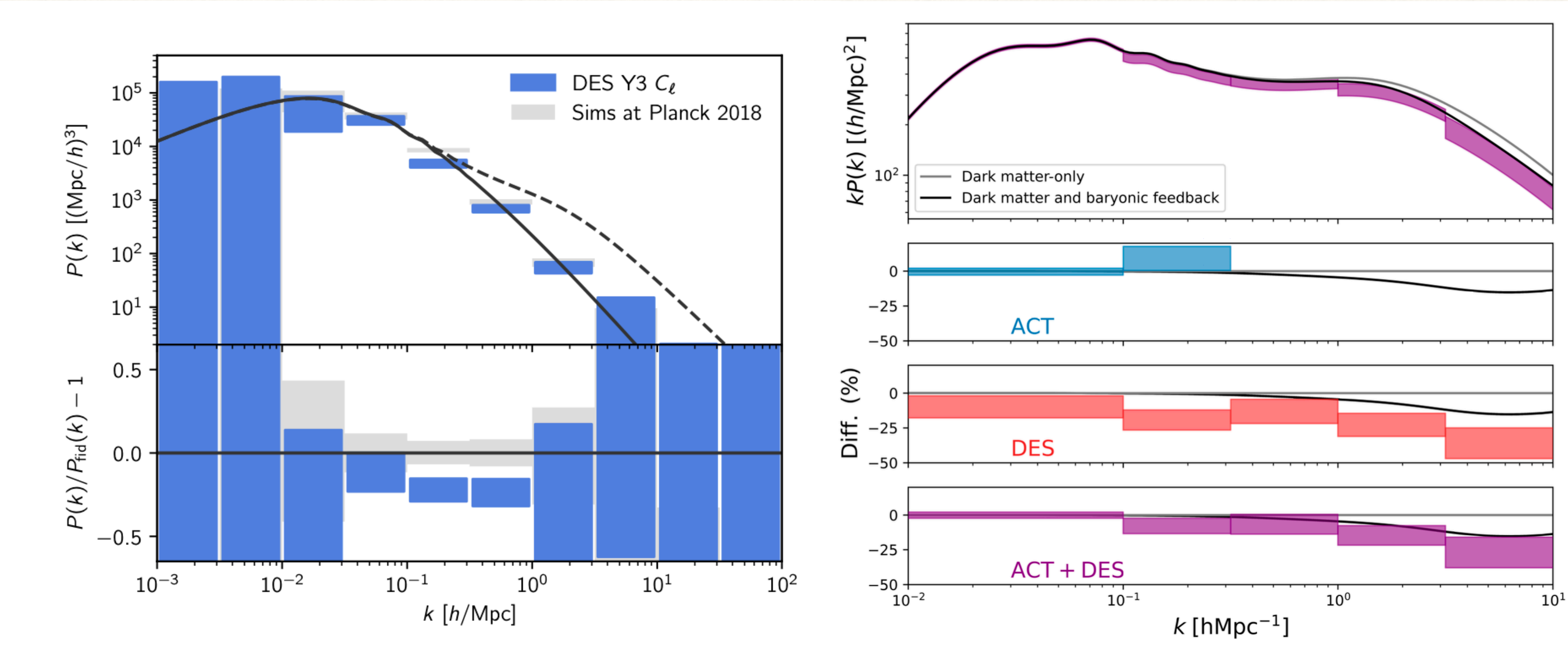
Evolving dark energy and the phantom crossing.

Dark Energy Survey mass map



Lensing mass map: DES Y3, 10 percent of the sky, 100 million galaxies.
DES Year 6: similar footprint on the sky, ~2x as many galaxies.

Structure mismatch – the S_8 tension



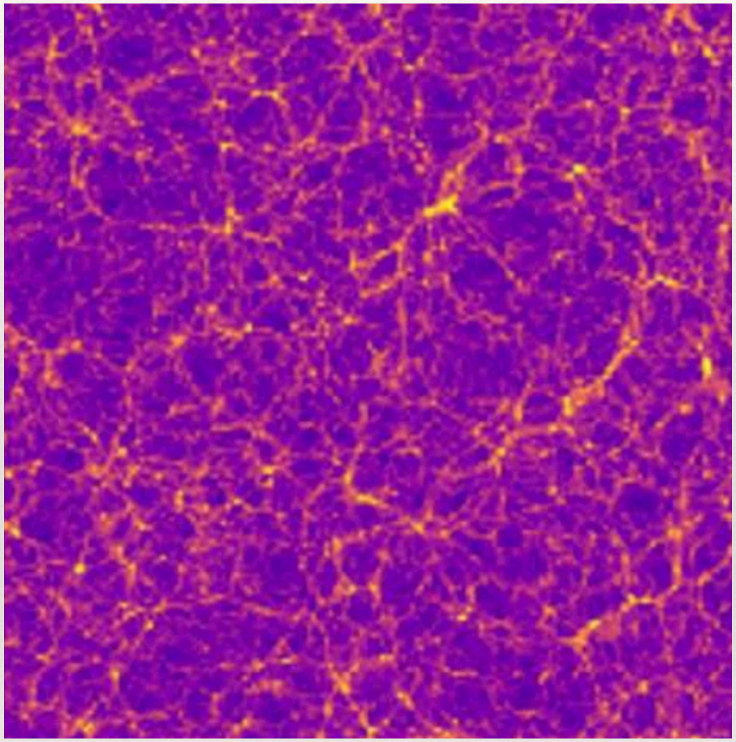
Reconstructing P(k,z)

Up to k~1 h/Mpc, DES requires a lower S8 within LCDM

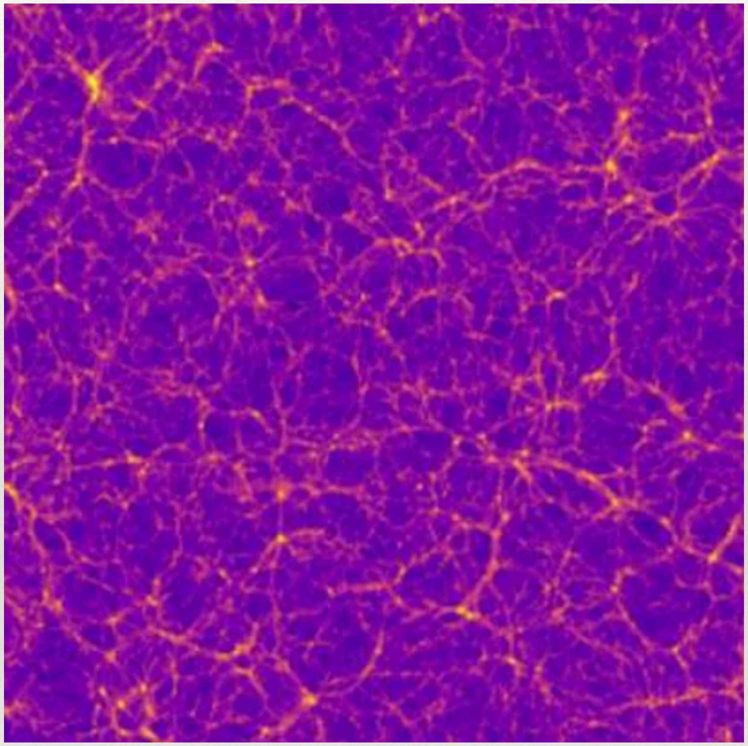
At higher k we see additional suppression.

Doux, BJ+ (DES) 2022; Sarmiento+ 2025

Structure mismatch – the S_8 tension



Predicted by theory



Inferred from lensing data

How can we compare these maps, using all the information they contain?

The non-Gaussian regime

- Beyond 2-point functions is where the action is in WL. A partial list of Higher Order Statistics (HOS):
    - 3-point correlations
    - PDF / CDF
    - Wavelet based statistics
    - Peak statistics
    - Clusters and voids
    - Minkowski functionals
    - Persistent homology
    - Field level inference with DL or BHM
- The 'ultimate' HOS paper? Euclid Paper XXVIII - a cautionary tale

- Weak lensing and the S8 tension
- Analytical vs Diffusion models for mass mapping
- Results from three-point correlations
- Wavelets and Higher Order Statistics
- Topics in Simulation Based Inference (SBI)

Analytical Lensing Maps

Astrophysics > Cosmology and Nongalactic Astrophysics

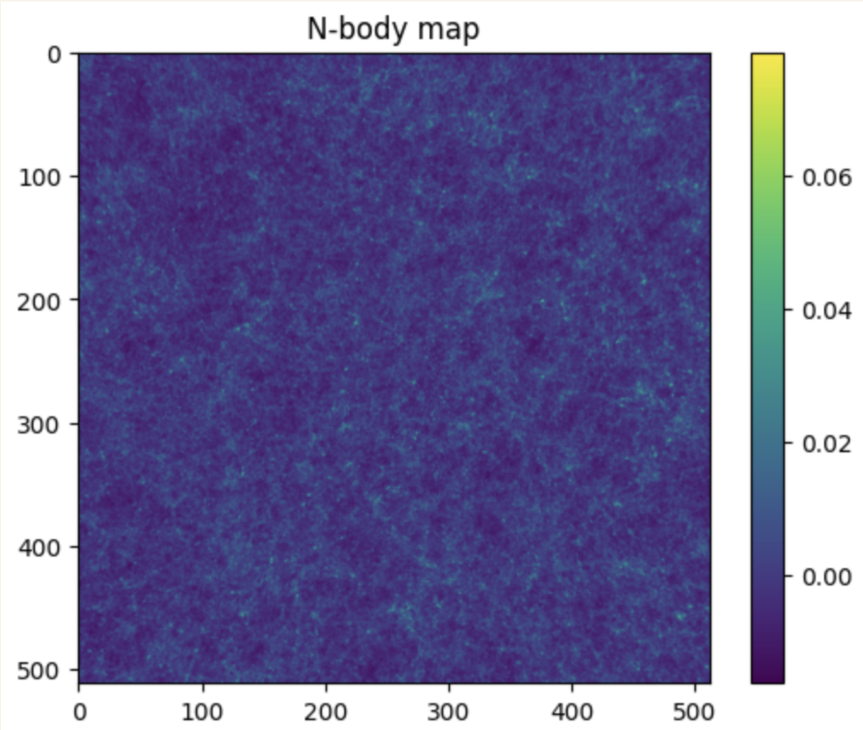
[Submitted on 7 Nov 2024]

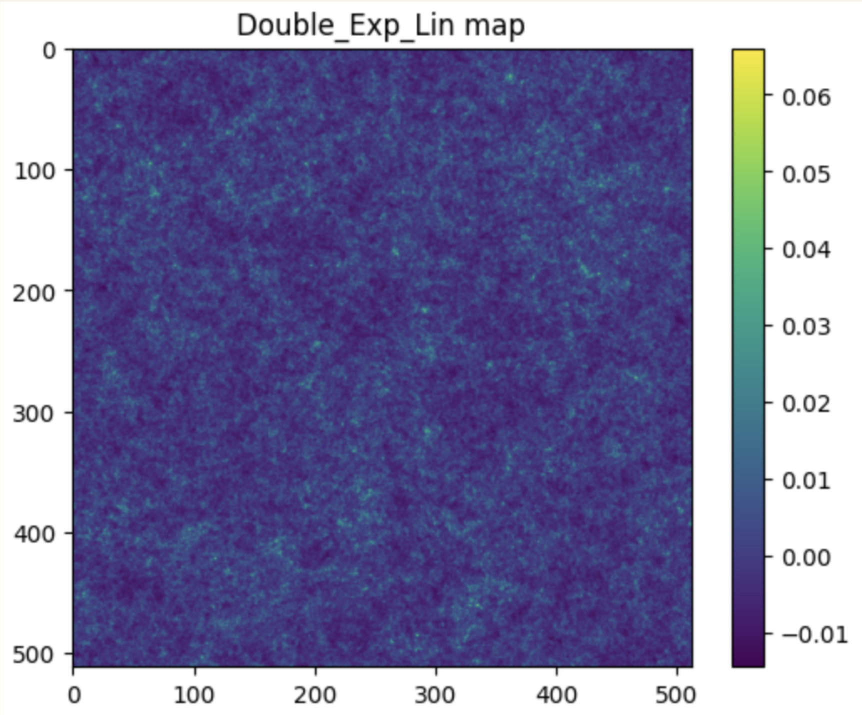
Fast Generation of Weak Lensing Maps with Analytical Point Transformation Functions

Kunhao Zhong, Gary Bernstein, Supranta S. Boruah, Bhuvnesh Jain, Sanjit Kobla

Nonlinear cosmological fields like galaxy density and lensing convergence can be approximately related to Gaussian fields via analytic point transforms. The lognormal transform (LN) has been widely used and is a simple example of a function that relates nonlinear fields to Gaussian fields. We consider more accurate General Point-Transformed Gaussian (GPTG) functions for such a mapping and apply them to convergence maps. We show that we can create maps that preserve the LN's ability to exactly match any desired power spectrum but go beyond LN by significantly improving the accuracy of the probability distribution function (PDF). With the aid of symbolic regression, we find a remarkably accurate GPTG function for convergence maps: its higher-order moments, scattering wavelet transform, Minkowski functionals, and peak counts match those of N-body simulations to the statistical uncertainty expected from tomographic lensing maps of the Rubin LSST 10 years survey. Our five-parameter function performs 2 to 5× better than the lognormal. We restrict our study to scales above about 7 arcmin; baryonic feedback alters the mass distribution on smaller scales. We demonstrate that the GPTG can robustly emulate variations in cosmological parameters due to the simplicity of the analytic transform. This opens up several possible applications, such as field-level inference, rapid covariance estimation, and other uses based on the generation of arbitrarily many maps with laptop-level computation capability.

Kappa maps via inverse Gaussianization





We can pick up some subtle differences by eyes, but ~all statistical tests pass at the LSST level. Applications follow from the ability to generate millions of ‘analytic’ kappa maps on a laptop

- Aid SBI in various ways
- Fine tune CNNs and ViTs
- Embedding vector for cosmological maps <-> physical parameters of the theory

Lognormal transform

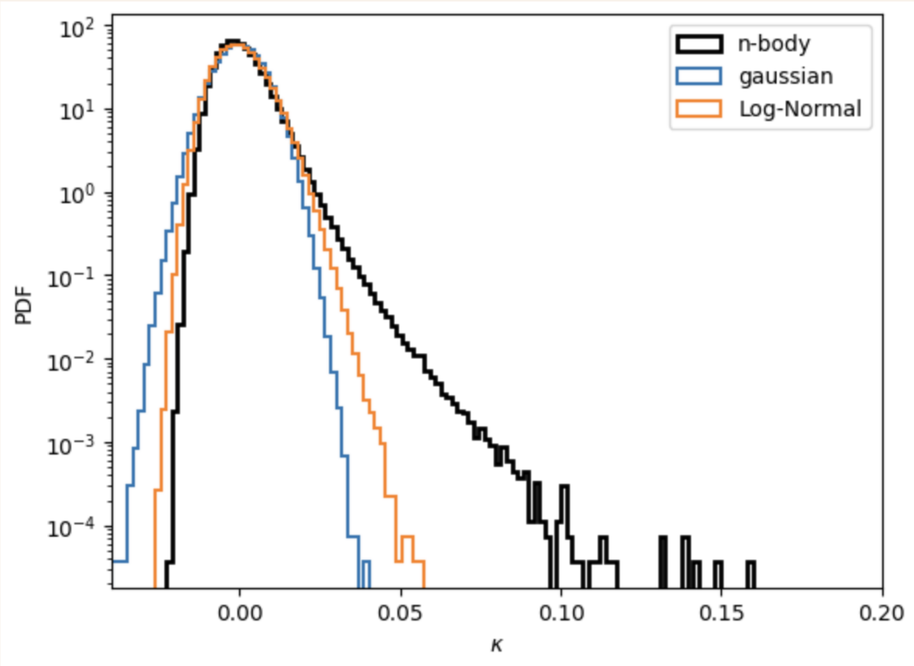
Write the observed field **Y** in terms of a Gaussian variable ***x~N(0,1)***

$$Y=Exp[x]$$

The PDF of this non-gaussian variable **Y** is

Known as **Log-Normal distribution:**

$$P(x) = \frac{1}{x\,\sigma\sqrt{2\,\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\,\sigma^2}\right).$$

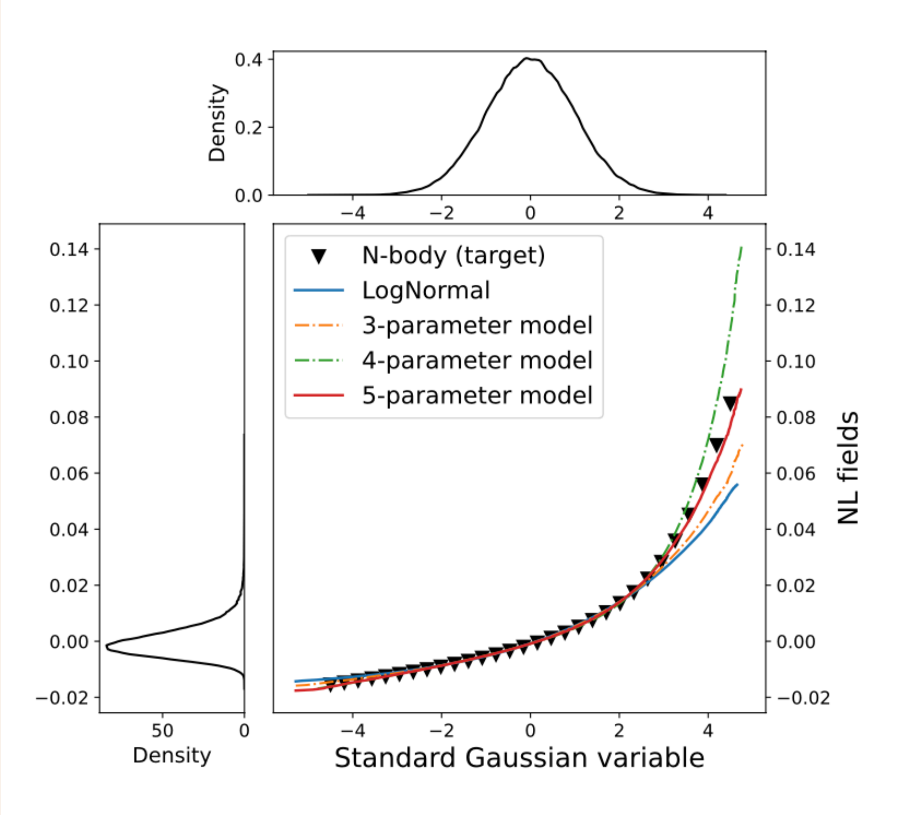


A ‘first order’ approximation to the nonlinear PDF, widely used but fails in the tails

General Point-Transformation:

- ‘Converting’ a general PDF to a Gaussian PDF can be done by CDF matching, or:

$$
\begin{aligned}
G(y_{\text{sim}}) &= \text{ppf}^{\text{norm}}\left(\text{cdf}(y_{\text{sim}})\right) \\
&= \sqrt{2}\,\text{erf}^{-1}\left(2\,\text{cdf}(y_{\text{sim}}) - 1\right),
\end{aligned}
$$

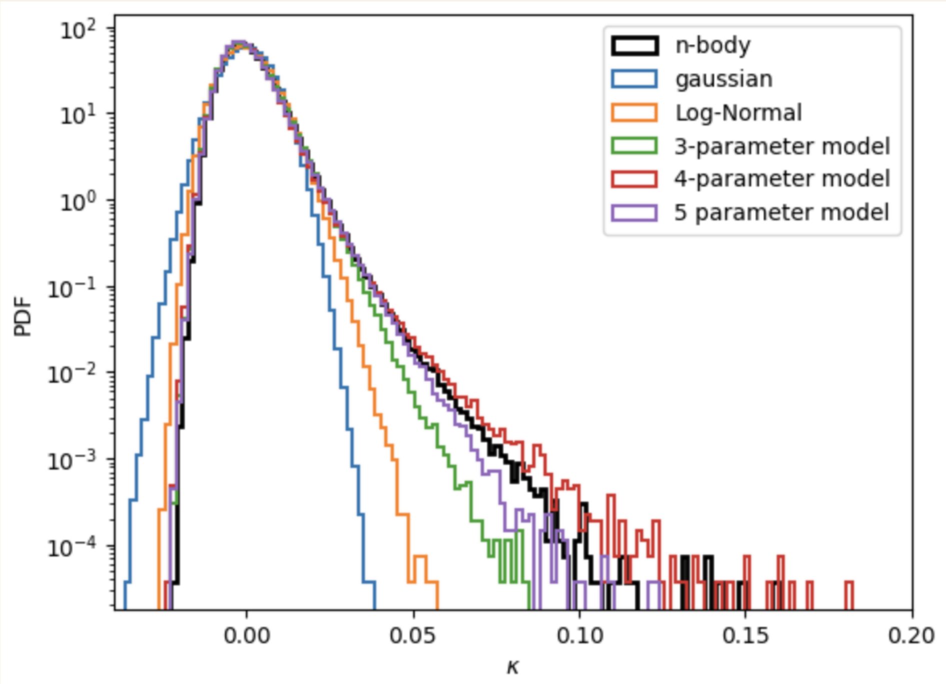


- Log-Normal models the nonlinear-linear curve as a straight line in log space, but it has significant curvature.

- Two straight lines? Close!

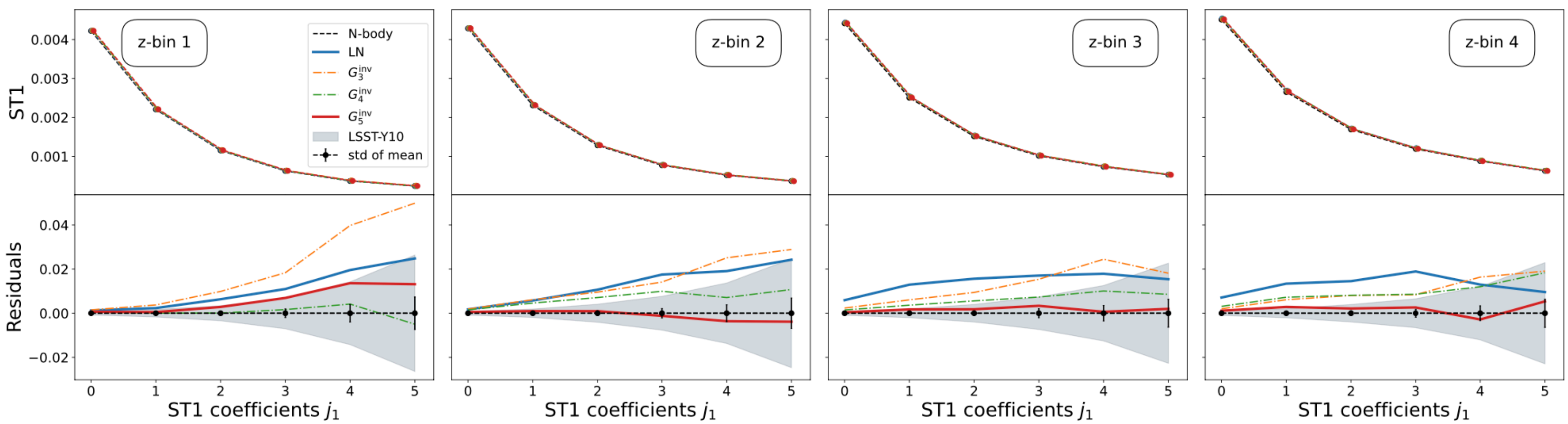
Accurate PDFs

- With 4-5 parameters, the PDF can be fit very well deep into the tails
- The power spectrum is 'exactly' matched
- All HOS match to within LSST accuracy! (above ~5 arcminutes)
  - 3rd moments
  - Peaks
  - Wavelets: ST1, ST2
  - Minkowski functionals



Name	Point Transform Function	Features
LN ($G_2^{\rm inv}$)	$\beta \exp(\alpha x - \alpha^2/2) - \beta$	The PDF and relation of correlation functions to the Gaussian field are analytical.
$G_3^{\rm inv}$	$n\left(e^{ax-a^2/2} + bx + c\right) - 1$	Best function suggested by Symbolic Regression. Fits individual tomographic bins, but failed to model cross-bin correlations.
$G_4^{\rm inv}$	$ne^{a_1 x - a_1^2/2}\left(1 + e^{(x-x_0)t}\right)^{(a_2-a_1)/t} - 1$	Functional form motivated by the limiting behavior of Gaussianization curve, previously tested on density shells with Quijote simulations [97].
$G_5^{\rm inv}$	$n\left(e^{a_1 x - a_1^2/2} + bx\right)\left(1 + e^{(x-x_0)t}\right)^{(a_2-a_1)/t} - 1$	Functional form that combines the two functions above. The two tails are flexible enough to accurately model the κ maps with different cosmology.

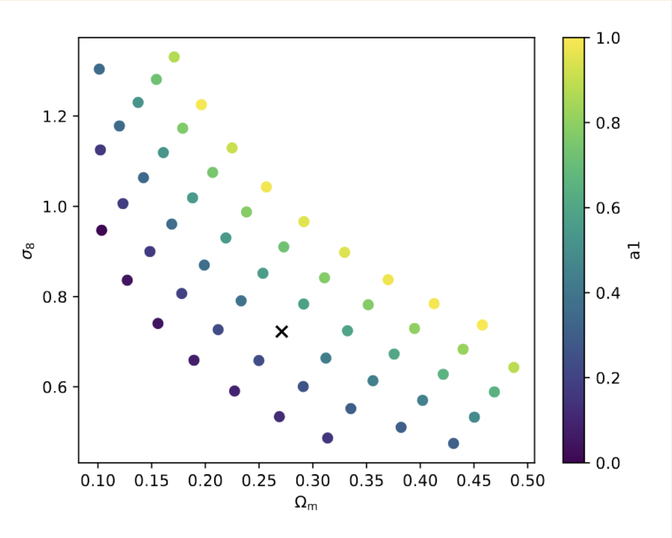
How good is the analytical mass map?

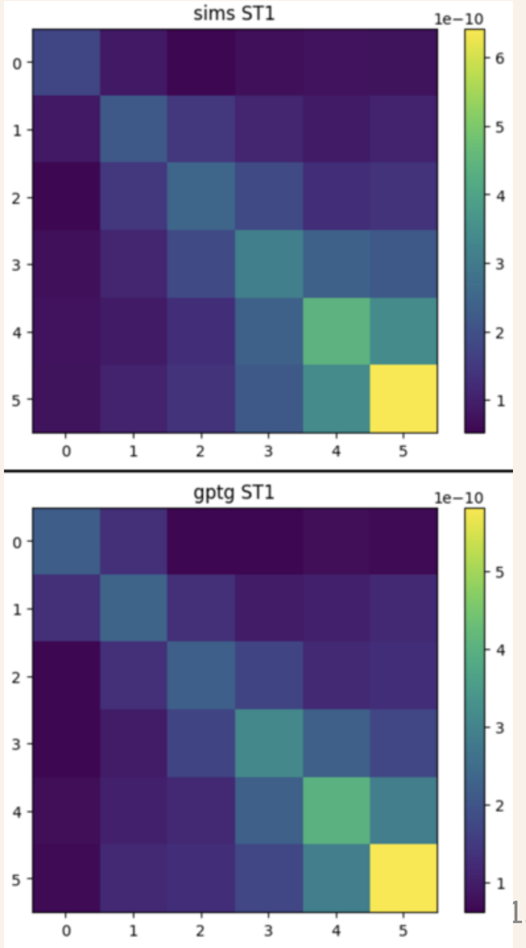


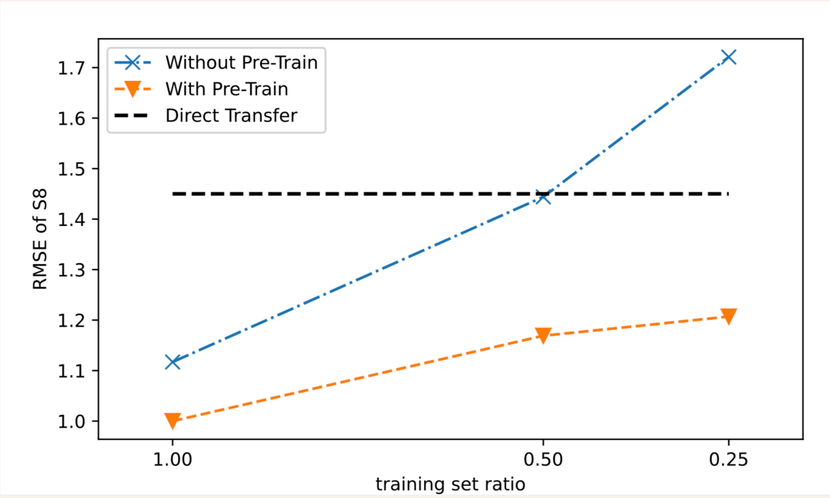
Predicted vs Simulation scattering transform: percent level agreement.
Gray band - expected LSST uncertainty bands.
PDF, Moments, peak counts, Minkowski functionals — within LSST uncertainty though there are deviations above simulation error bars.

Why another (mediocre) emulator?

- 5 parameter fits can be easily generalized: interpolate in parameter space; extended cosmo models; avoid overfitting…).
  - *Interpolation is one of the many hidden issues with SBI.*
- Rapid covariance estimation
- Pretraining CNNs and ViTs
- Transfer learning
- Various ways to aid SBI/FLI



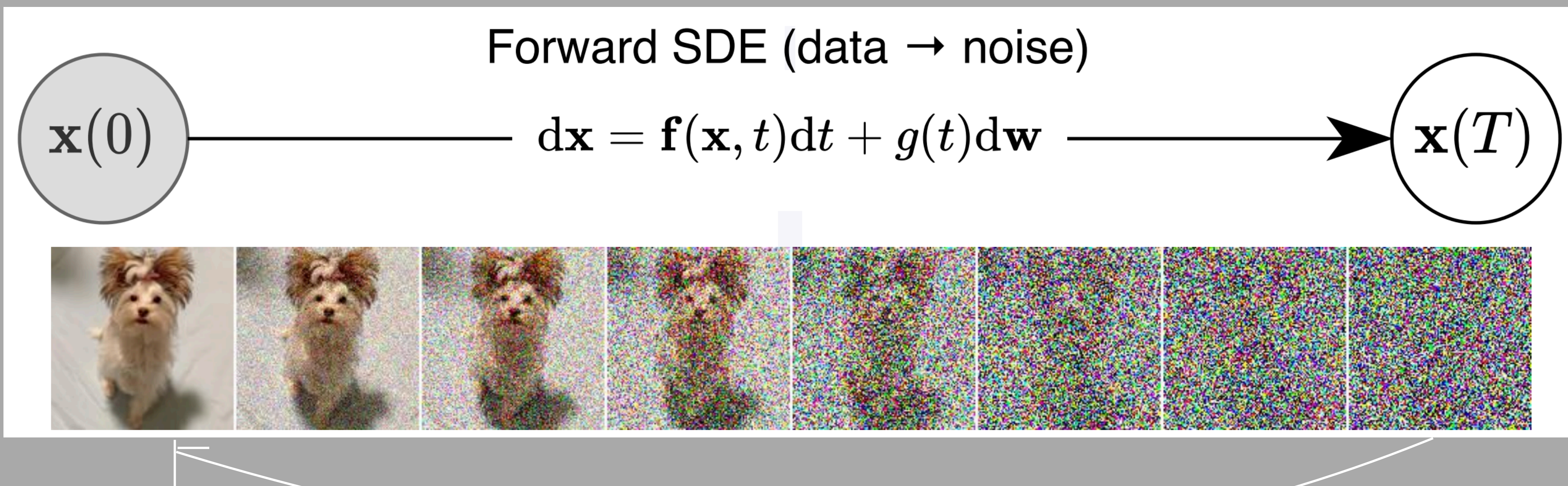




- When you start a Machine Learning project, you may end up writing an ‘anti-ML’ paper

Diffusion — the main idea

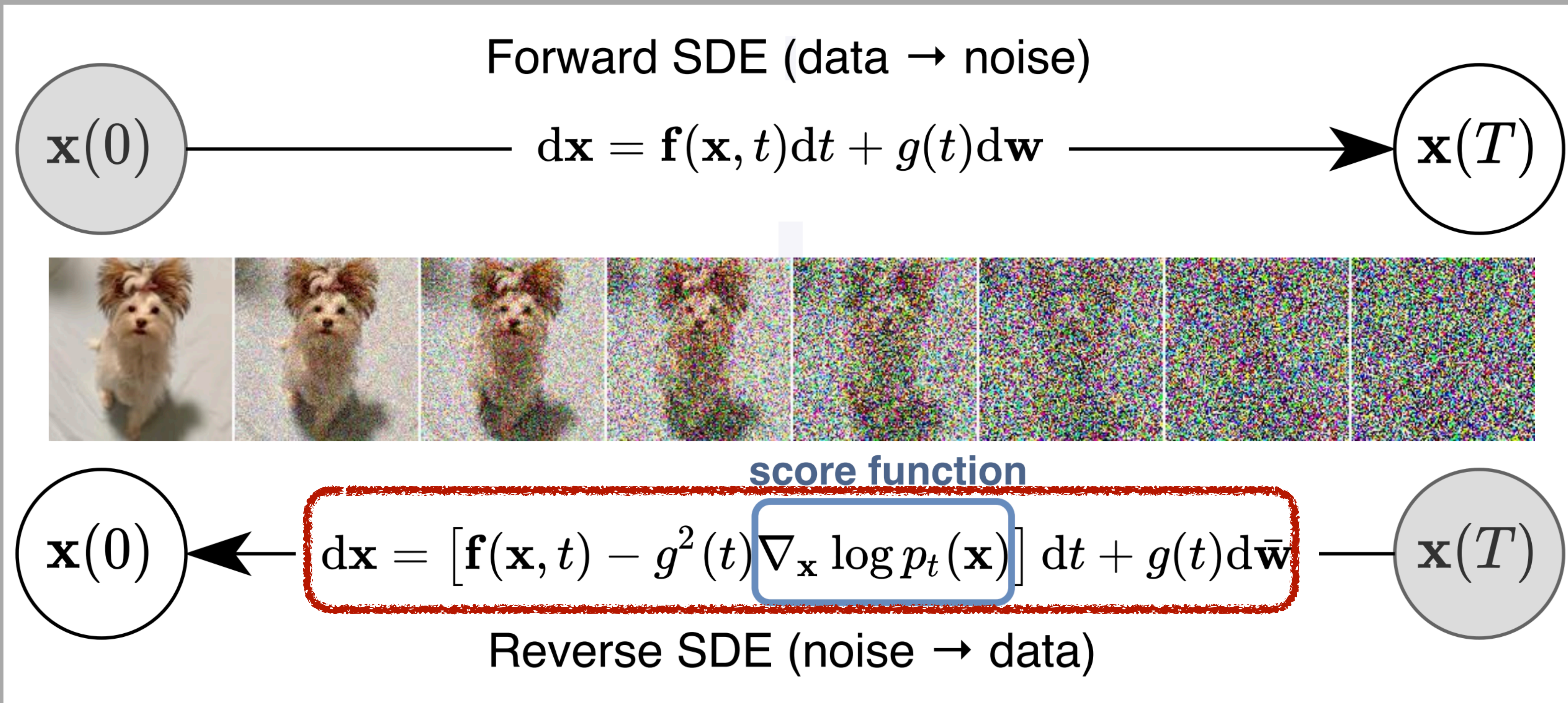
Progressively add noise to the image until we just have white noise
(We completely control this step via a Stochastic Differential Equation)



Slide from Supranta Bouruah

Why diffusion is useful for generative modeling?

Important result: Any SDE of this form can be reversed!



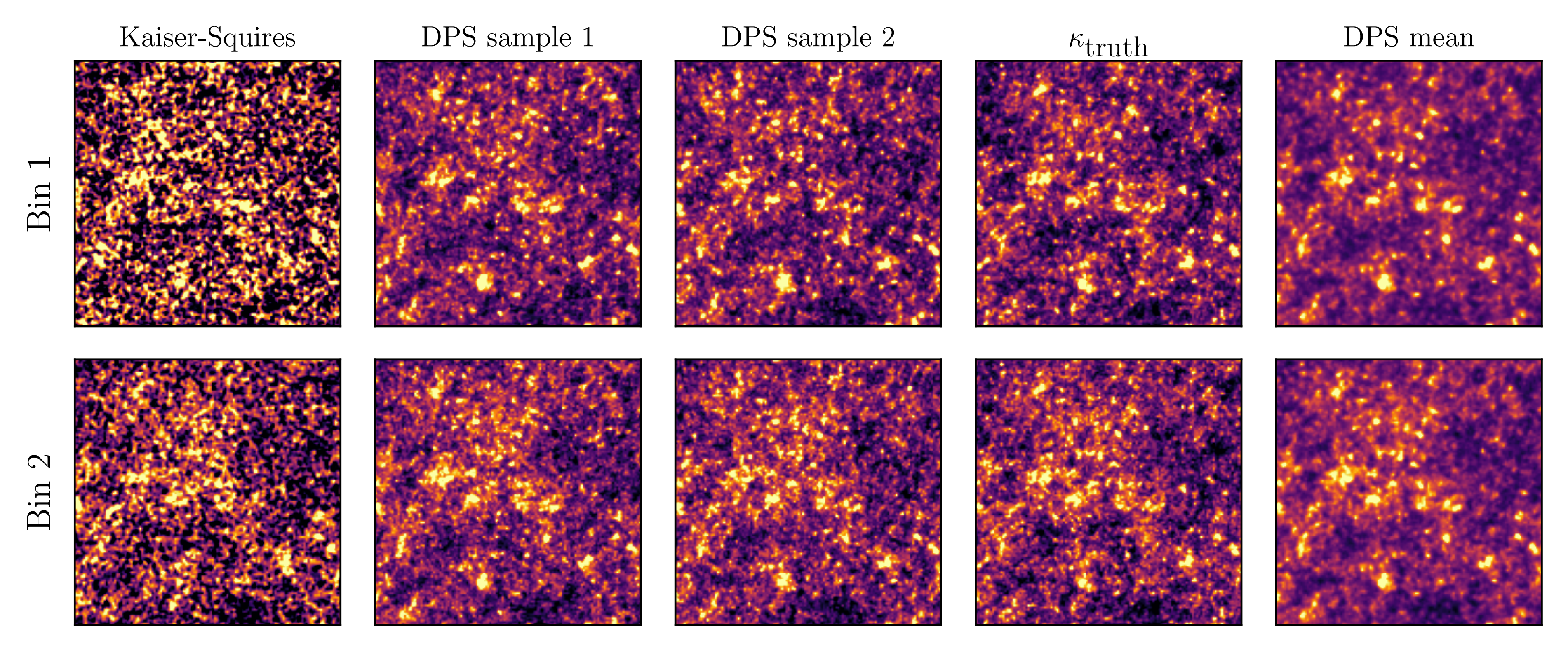
Reverse stochastic differential equation requires the gradient of the log probability, a.k.a ***score***

Train neural networks to predict the score, at different noise level

Once trained, neural networks can be used to transport latent space noise to samples from the data distribution

Slide from Supranta Bouruah

Diffusion models: noisy data ➡ underlying field



- Learns the 'prior' distribution of mass maps from simulations
- Reconstruct mass map from noisy survey data with no additional training!
- High resolution maps with **uncertainty** estimates — fast and robust
- Uses the simulation prior and shared information across z bins to achieve arc-minute scale resolution

Bouruah, Jacob, BJ 2025, arXiv:2502.04158

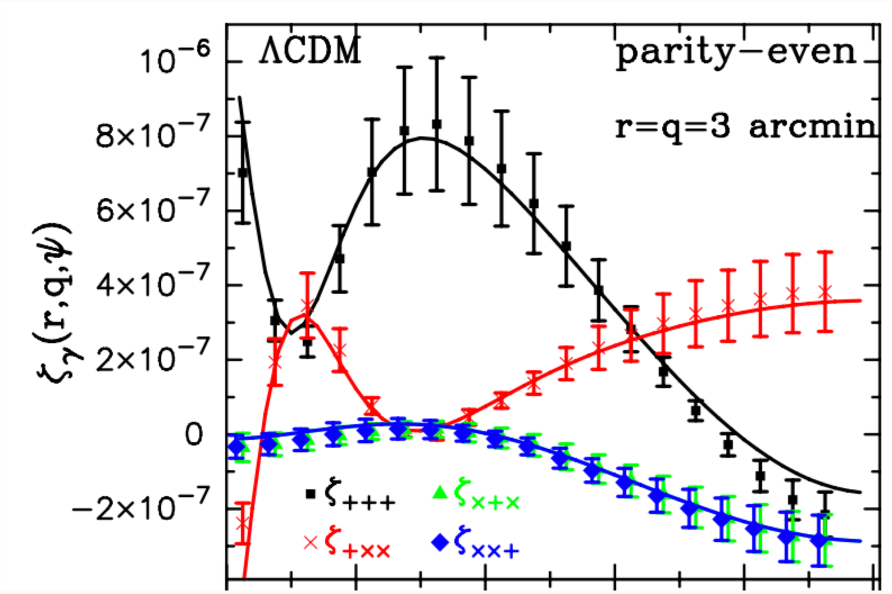
Mass mapping summary

- Our analytical Gaussianizing method is effective on scales above the 1-halo regime
  - applications for covariance estimation, data compression and fine tuning of deep learning models
- Diffusion models help generate high resolution mass maps **given** a simulation based prior
  - Additional applications in reconstructing mass maps from noisy data: finding interesting LSS features and comparing to optical or SZ maps

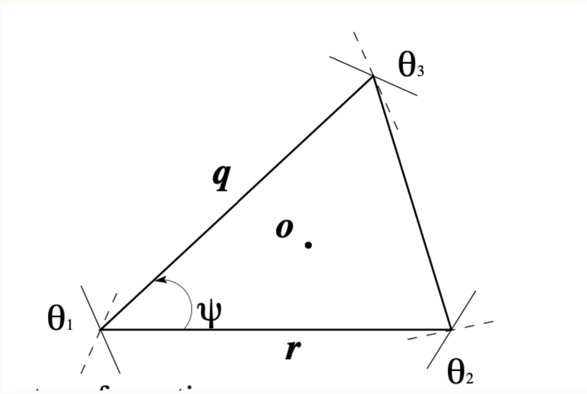
- Weak lensing and the S8 tension
- Analytical vs Diffusion models for mass mapping
- Results from three-point correlations
- Wavelets and Higher Order Statistics
- Topics in Simulation Based Inference (SBI)

Three-point correlations

Four 3-point functions of shear



0 $\psi \rightarrow$ π



Distinct signatures of quasilinear gravitational clustering

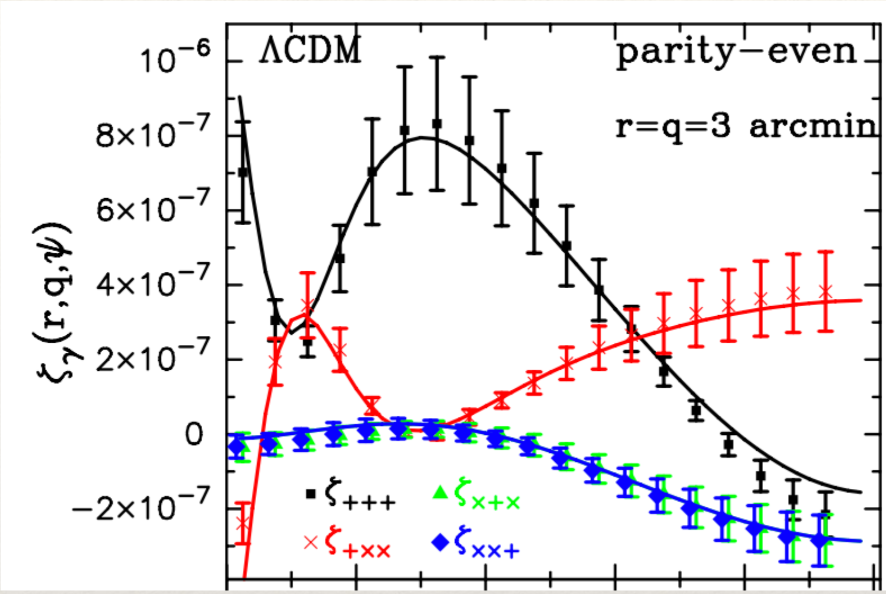
Takada & BJ 2003a,b; Zaldarriaga & Scoccimarro; Ho & White 2003

Krause, Eifler, Schneider 2012; Linke+ 2022; Burger+ 2024

For galaxy distribution: Wang, Jeong+ 2024

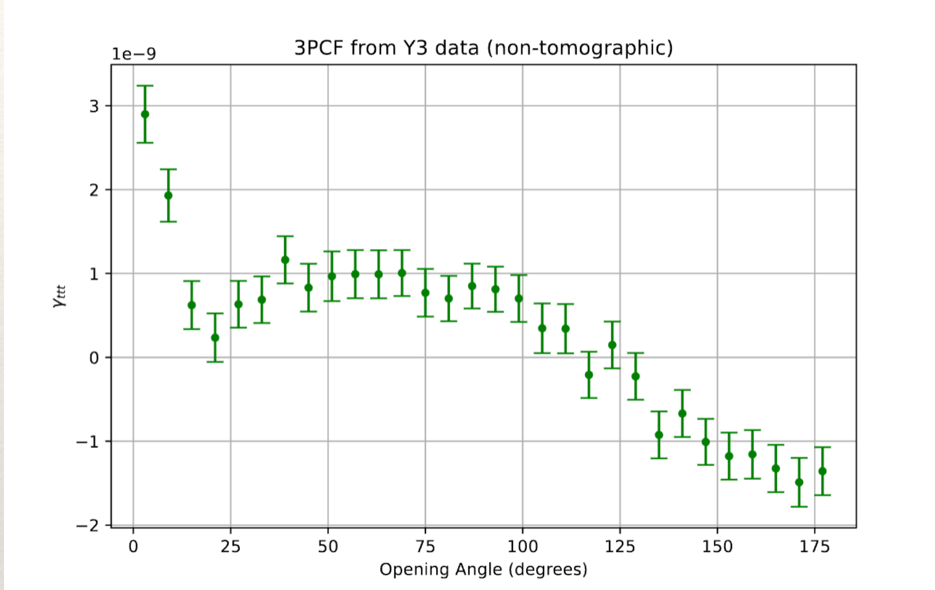
3-point function: DES lensing

Halo model 1-halo term prediction



0 $\psi \rightarrow$ π

DES measurement



Isosceles triangles, versus opening angle

Shape dependence: enables checks on systematics

Much of the information from the quasilinear regime

Secco, Jarvis, BJ+ (DES) 2022

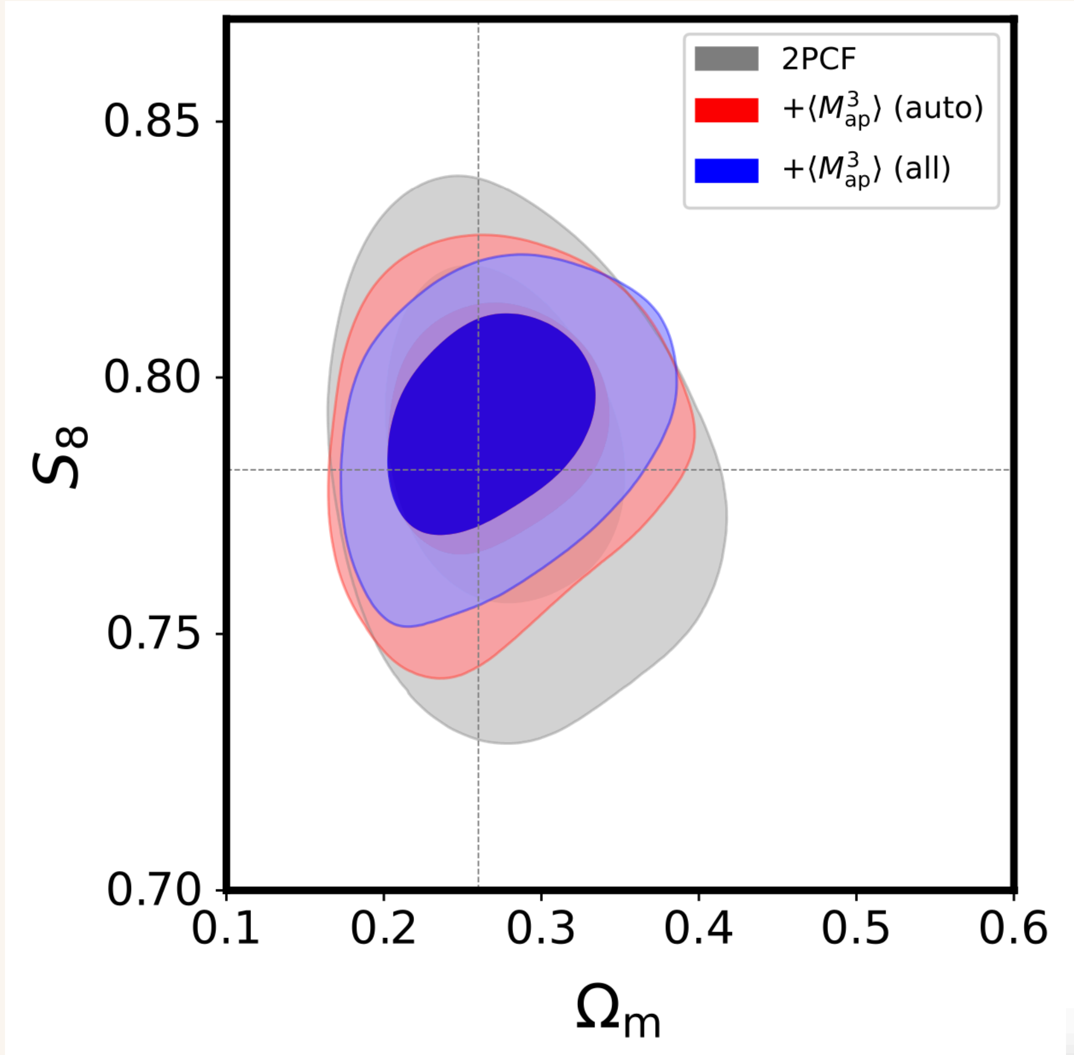
Gomes, Sugiyama, BJ+ (DES) 2025a,b

Cosmology with 2+3pt functions: Simulations

1. All 3-point functions: 100,000 element datavecor
2. Compress into Map^3 -> ~100 numbers (90% of the information)
3. All tomographic bins (20)
4. Model IA (NLA) and mitigate other systematics.

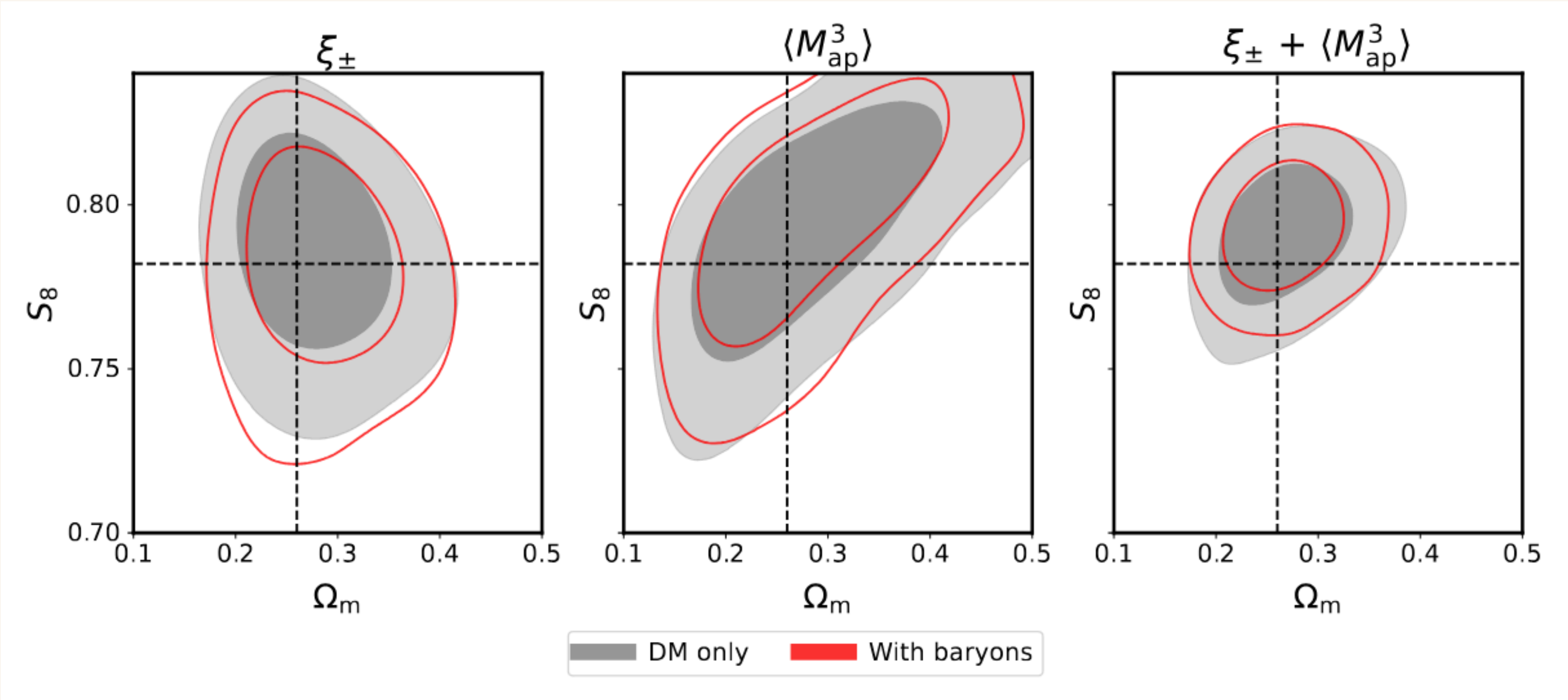
Gain ~80% on Figure of Merit (FOM)

See also Linke+ 2023 and Burger+ (KIDS) 2024 analysis



Gomes, Sugiyama,BJ + DES 2025; Data analysis next

Robustness to baryonic feedback



Possibility of different effects of baryonic feedback on 2 and 3pt correlations: net result is more robust. But depends on analysis details.

Wavelet based nonlocal, nonlinear transforms

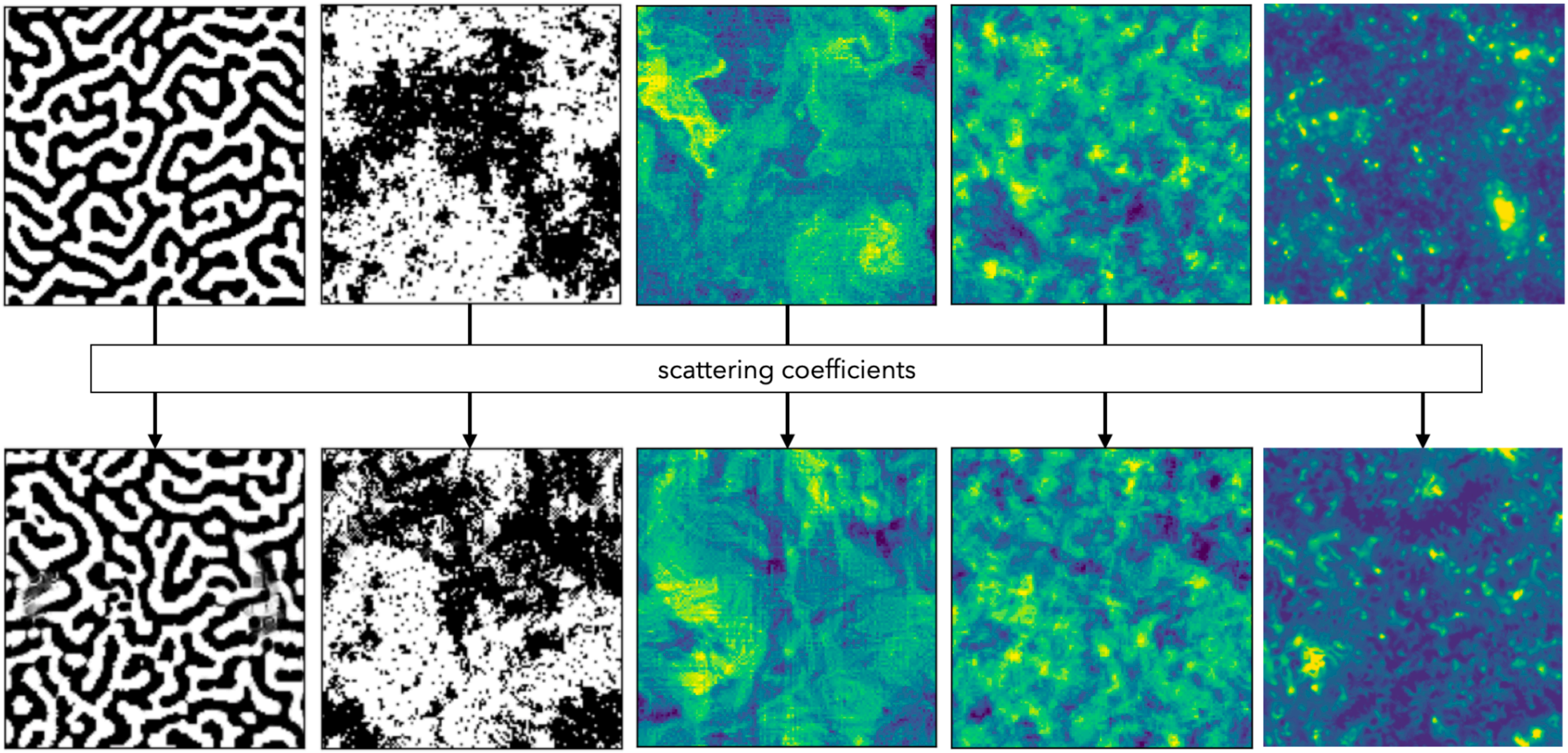
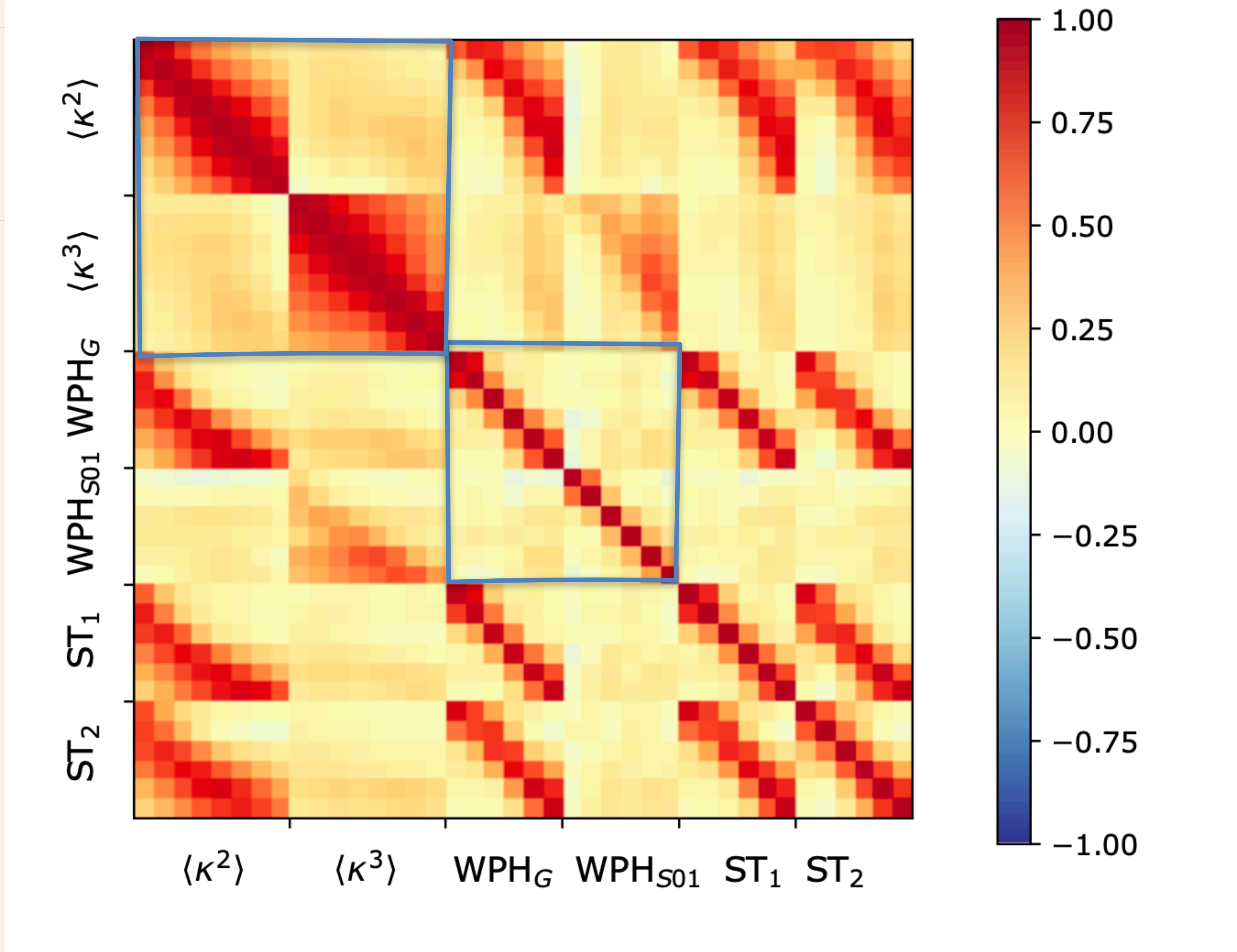


Figure 2. Texture synthesis using the scattering transform for a variety of physical fields: Turing pattern, Ising model, ocean turbulence, solar surface, cosmic matter density. The upper panels show input 2-D fields from simulations or observations. The lower panels show randomly generated fields with scattering coefficients matching their upper counterparts.

Wavelet can represent non-Gaussian stochastic fields pretty well. *Cheng & Menard 2021*

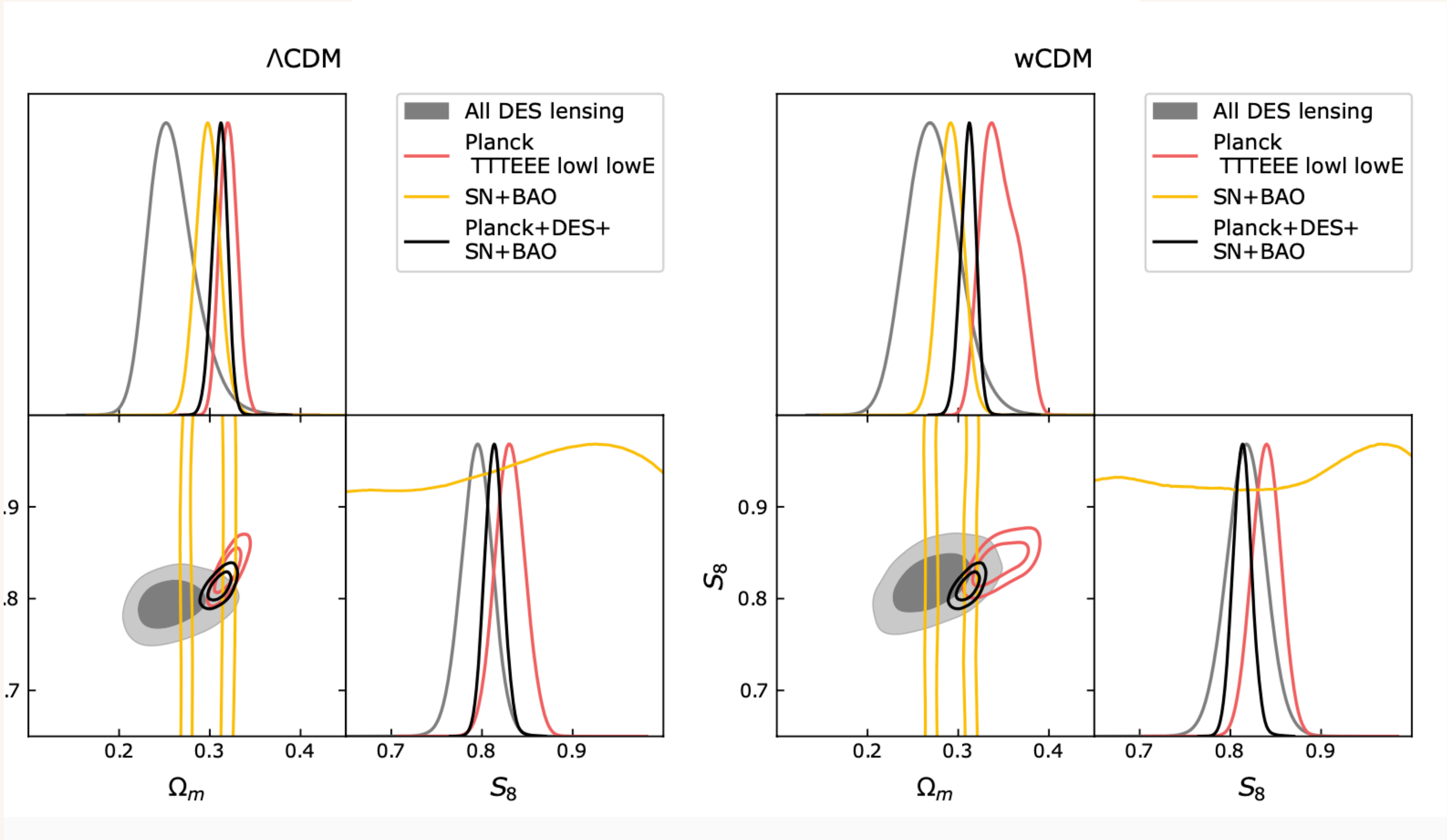


Happiness is…a band diagonal covariance matrix!

The wavelet-transformed field has most of the information in its **low order** statistics.

Gatti+ 2023.

SBI (Simulation Based Inference, aka LFI) with wavelets: DES Year 3 analysis



SBI steps: Simulation, Summary statistic measurement, Data compression, Emulation, Posterior estimation, Tests of confidence intervals

2-sigma tension with Planck (but consistent with w=-1)

Gatti+ 2023,2024 (DES collaboration): arXiv:2310.17557, arXiv:2405.10881

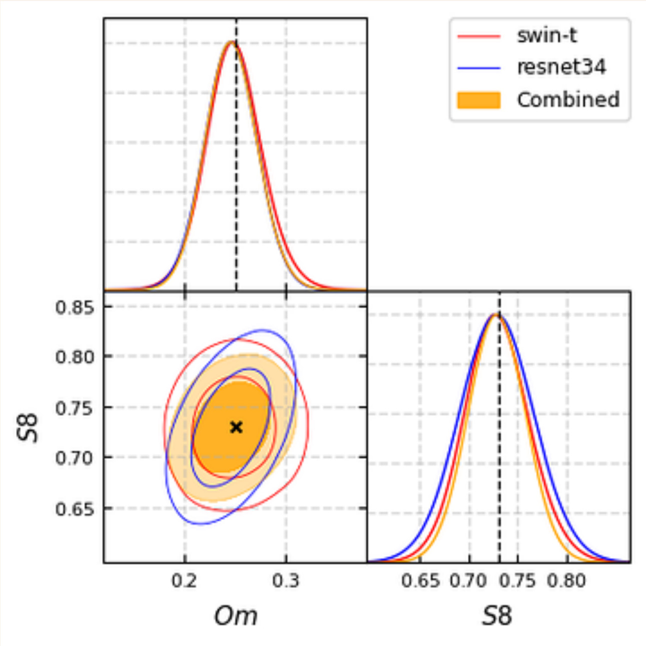
Summary Statistics vs Field Level Inference

- Simulation Based Inference is well matched for weak lensing: fully nonlinear field with uncertain physics and systematics ~amenable to simulations. Which is not to say it is easy!
- Deep learning aims to extract all the information in mass maps (though most of the new information comes from the smallest scales)
- Results so far are promising but a long path ahead to validate and interpret the extra information gained
- *See: Sharma, Dai, Seljak papers.*

Is Interpreting Deep Learning a fantasy?

Do we even Interpret 'analytical' methods anymore?

Assorted deep learning topics



- A light-weight CNN for cosmology with regularization methods that avoid over-fitting. *Zhong, Gatti, BJ, arXiv:2403.01368*
- AI x Science collaboration: sum-of-parts interpretation. *w/ Weiqiu You & Eric Wong at Penn*

- Weak lensing and the S8 tension
- Analytical vs Diffusion models for mass mapping
- Results from three-point correlations
- Wavelets and Higher Order Statistics
- Dimensionality reduction for Simulation Based Inference (SBI)

Astrophysics > Cosmology and Nongalactic Astrophysics

[Submitted on 3 Sep 2024 (v1), last revised 6 Sep 2024 (this version, v2)]

Dimensionality Reduction Techniques for Statistical Inference in Cosmology

Minsu Park, Marco Gatti, Bhuvnesh Jain

We explore linear and non-linear dimensionality reduction techniques for statistical inference of parameters in cosmology. Given the importance of compressing the increasingly complex data vectors used in cosmology, we address questions that impact the constraining power achieved, such as: Are currently used methods effectively lossless? Under what conditions do nonlinear methods, typically based on neural nets, outperform linear methods? Through theoretical analysis and experiments with simulated weak lensing data vectors we compare three standard linear methods and neural network based methods. We propose two linear methods that outperform all others while using less computational resources: a variation of the MOPED algorithm we call e-MOPED and an adaptation of Canonical Correlation Analysis (CCA), which is a method new to cosmology but well known in statistics. Both e-MOPED and CCA utilize simulations spanning the full parameter space, and rely on the sensitivity of the data vector to the parameters of interest. The gains we obtain are significant compared to compression methods used in the literature: up to 30% in the Figure of Merit for Ω_m and S_8 in a realistic Simulation Based Inference analysis that includes statistical and systematic errors. We also recommend two modifications that improve the performance of all methods: First, include components in the compressed data vector that may not target the key parameters but still enhance the constraints on due to their correlations. The gain is significant, above 20% in the Figure of Merit. Second, compress Gaussian and non-Gaussian statistics separately -- we include two summary statistics of each type in our analysis.

Dimensionality Reduction

- Linear Methods : Construct transformation matrices (U) such that data (vector) compression is a linear operation $c = Ut$
  - PCA
  - MOPED, e-MOPED
  - CCA
- Non-linear methods : Train a neural network such that the compression is a neural network transformation $c = f_\phi(t)$
  - NN-MSE, IMNN, VMIM

MOPED, e-MOPED

- MOPED is ‘Optimal’ for a Gaussian linear model with parameter independent noise
- Computing fiducial covariance and parameter derivatives is computationally expensive
  - Especially so for higher order statistics
- What if we could compute them without extra simulations? Use linear regression to find the implied Jacobian (derivatives)
- With this linear model, we can “shift” the simulated data vectors to a fixed point in parameter space for a ‘fiducial covariance’ estimate
- It’s “easier”: e-MOPED!

Canonical Correlation Analysis

- Identify linear combinations of data vector and parameters of maximum correlation, can also be understood as maximizing mutual information between parameter and data vector given a Gaussian linear model
- Boils down to a generalized eigenvalue problem

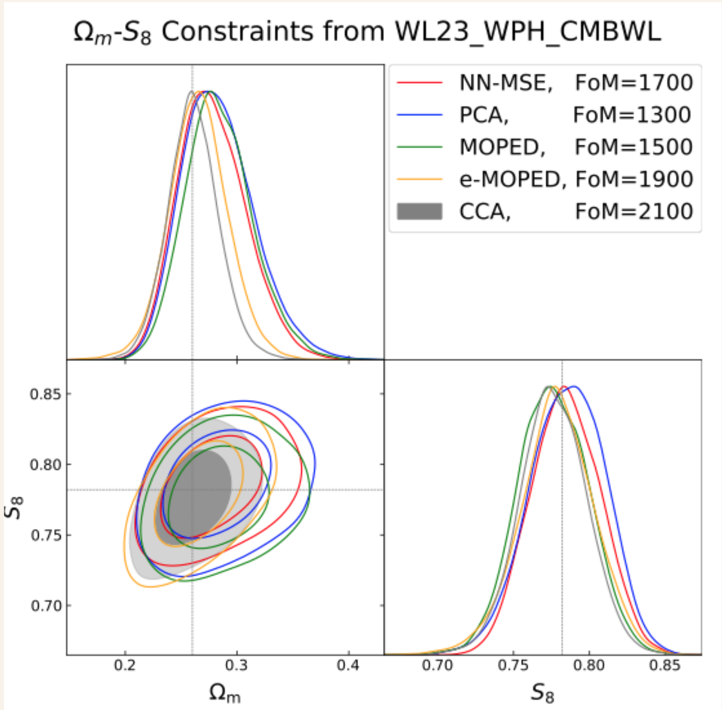
NN-MSE and Optuna

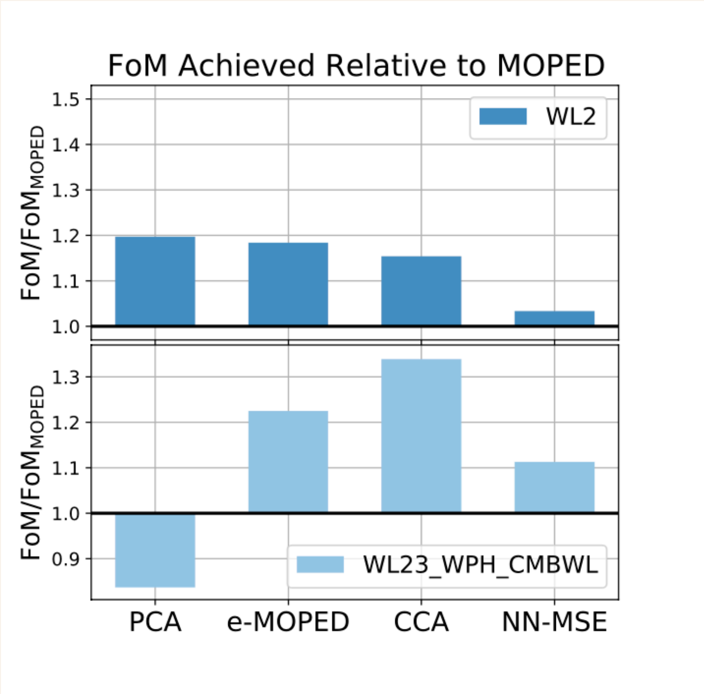
- Train neural networks to infer parameter values from data vector

$$\min_{\phi} \sum_{i} (f_{\phi}(t_i) - p_i)^2$$

- The (optimally) inferred parameter values serve as the compressed DV
- Each parameter has a different optimal architecture for inference
  - Tune architecture for each parameter with Optuna (# of layers, layer width, learning rate)

Which Compression Method Should We Use?





- **For current data and simulation resources: CCA or e-MOPED**
- PCA fails with HoS
- NN-MSE worse than linear in practice
- CCA and e-MOPED computationally feasible and better than MOPED

Conclusions

- Analytical point transforms generate reliable mass maps on scales above ~2 Mpc
- Diffusion models generate and reconstruct high resolution mass maps
- 3-point correlations and wavelet based statistics are powerful Beyond 2-pt statistics.
- We are developing a path to selecting a realistic set of summary statistics.
- Improvements in SBI
  - Data compression
  - Interpolating simulations in parameter space
  - Build and regularize a CNN for cosmology
  - Fine tuning for deep learning