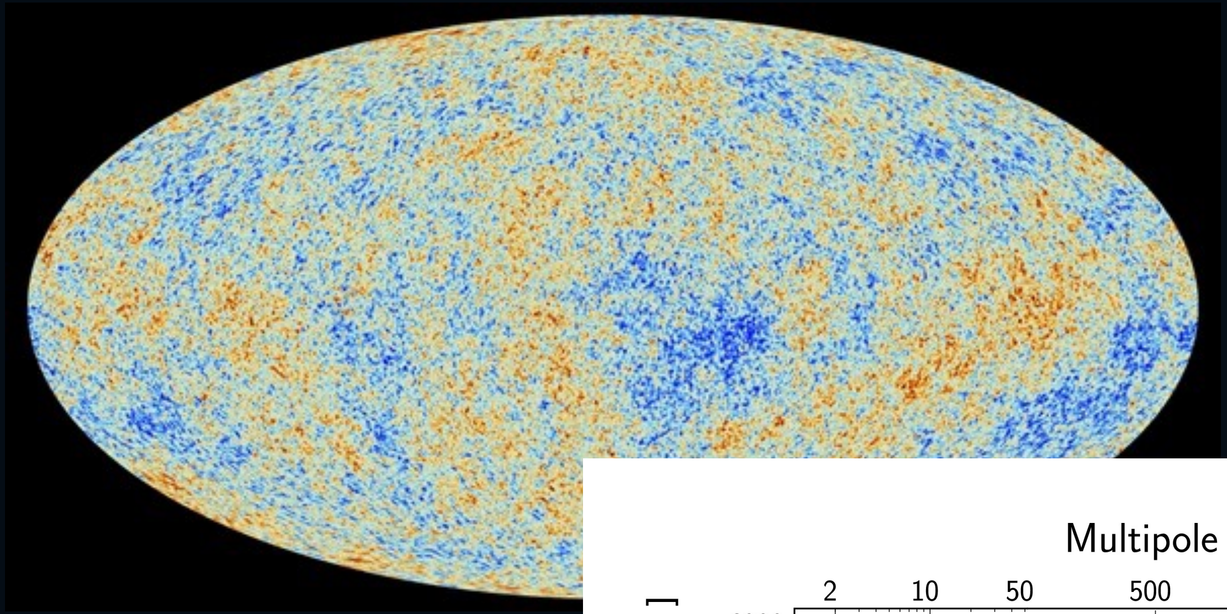
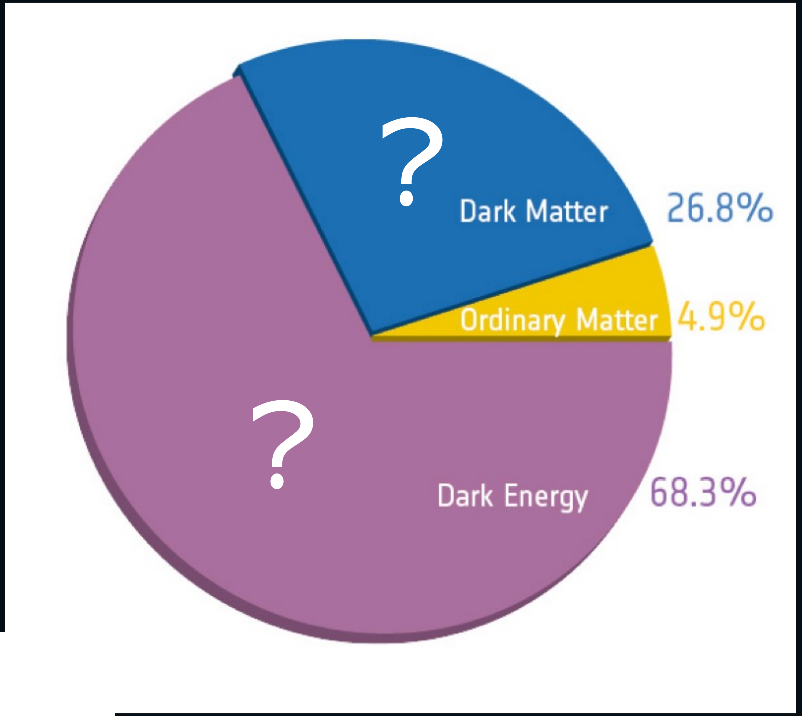
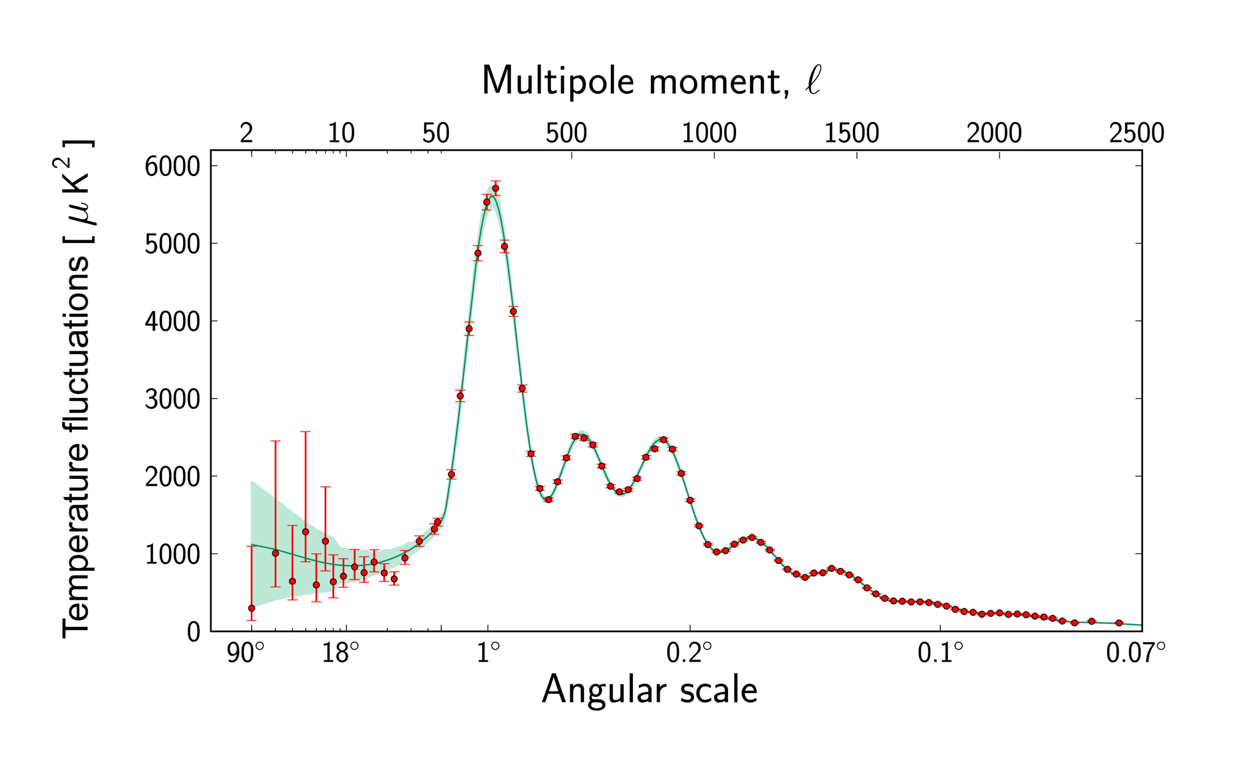
DeepLensing Flow: Score based and flow models for weak lensing statistics

Joaquin Armijo, Centre for Data-Driven Discovery, Kavli IPMU

Precision cosmology: The ΛCDM model





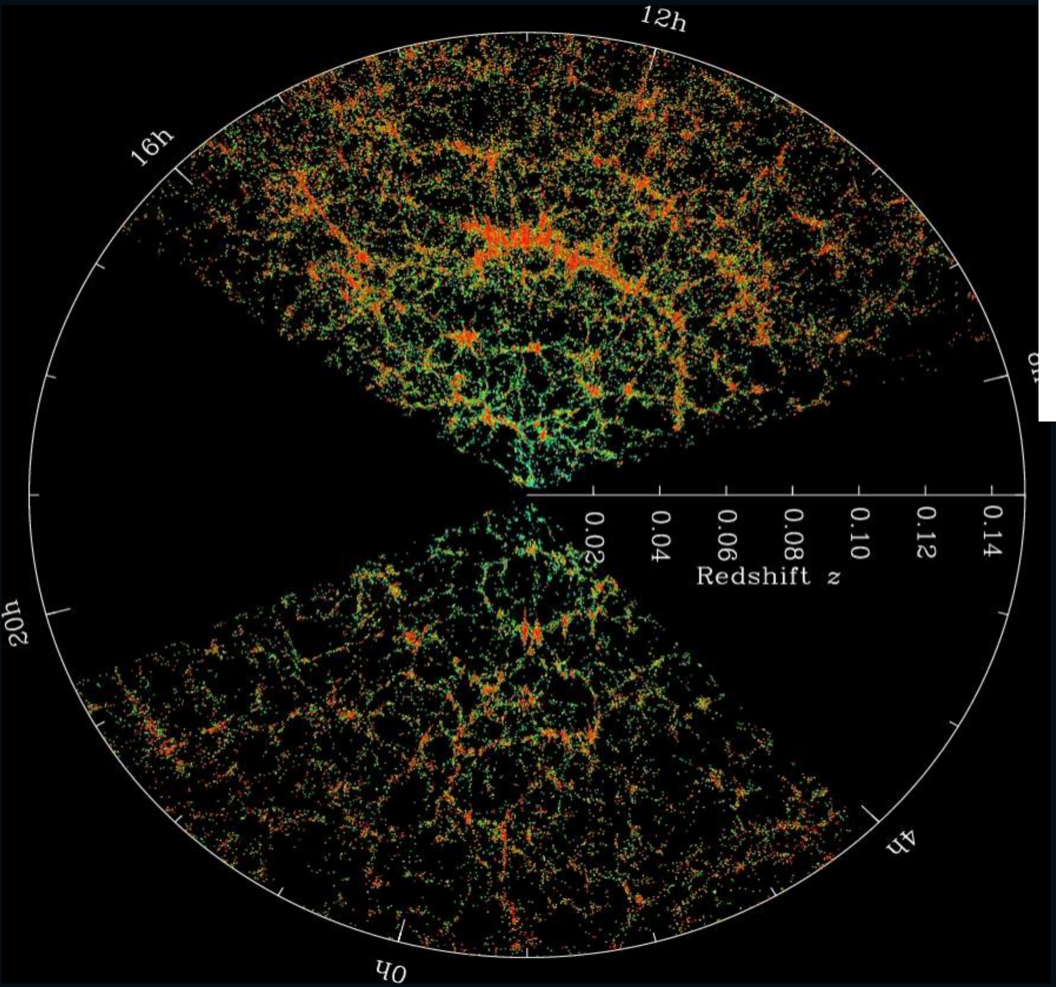


Angular power spectrum

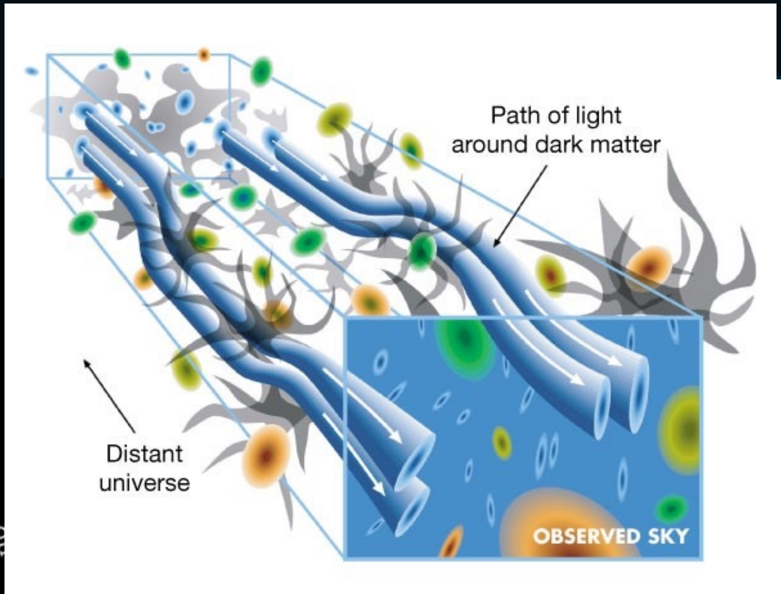
$$C_\ell = \frac{1}{2\ell + 1} \sum_m a_{l,m} a^*_{l,m}$$

Probing the large-scale structure of the Universe

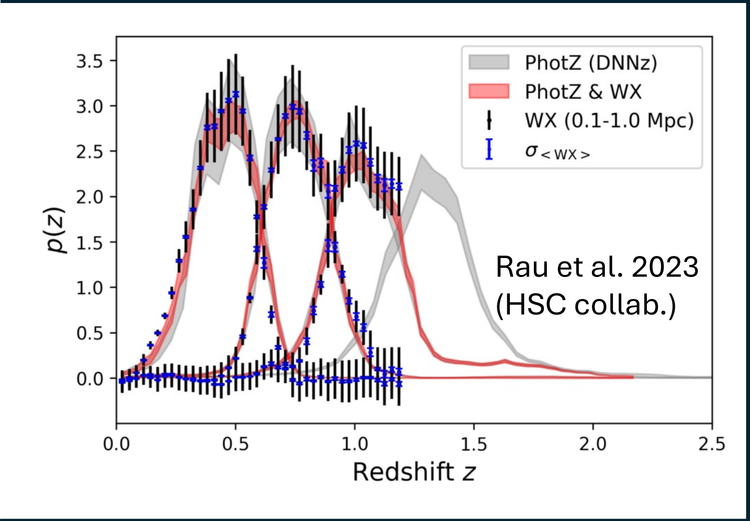
Either assuming a connection between galaxies and the matter field or using gravitational lensing.

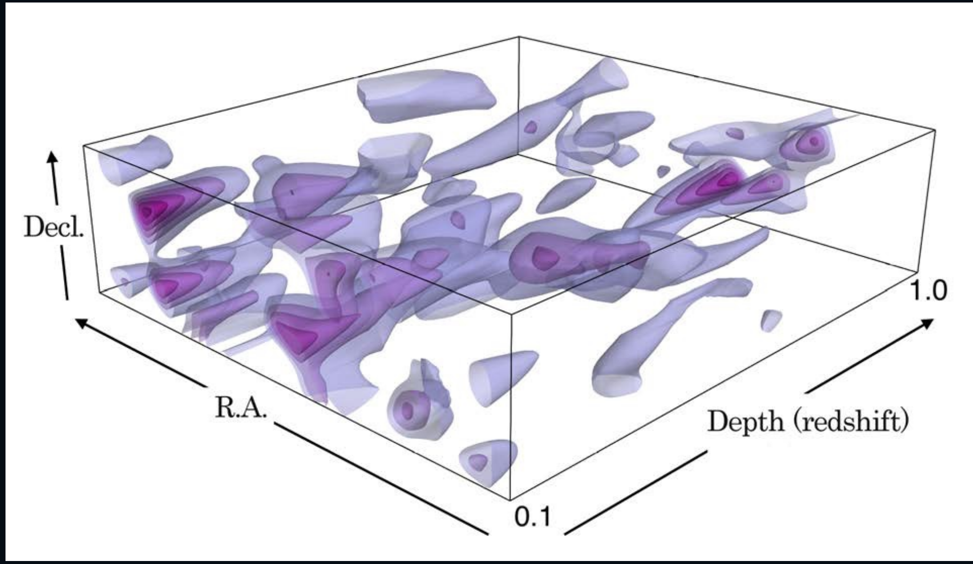


Credits: SDSS



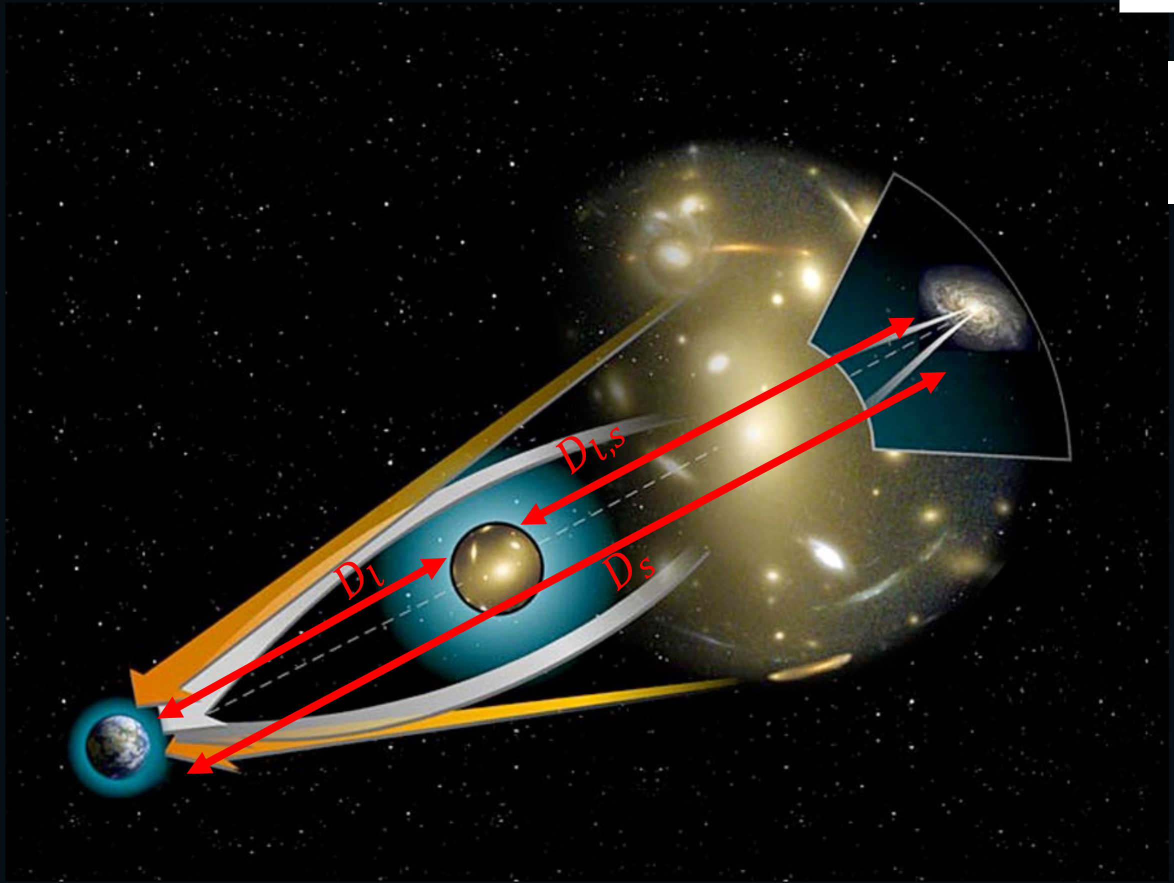
Needs high number density of galaxies!





Weak lensing fields

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{\text{crit}}} \longrightarrow \kappa(\boldsymbol{\theta}) = \int_0^{r_{\text{hor}}} dr\, w(r)\delta(x(r)\boldsymbol{\theta}, r)$$



$$\Sigma_{\text{crit}} \propto \frac{D_s}{D_l D_{l,s}}$$

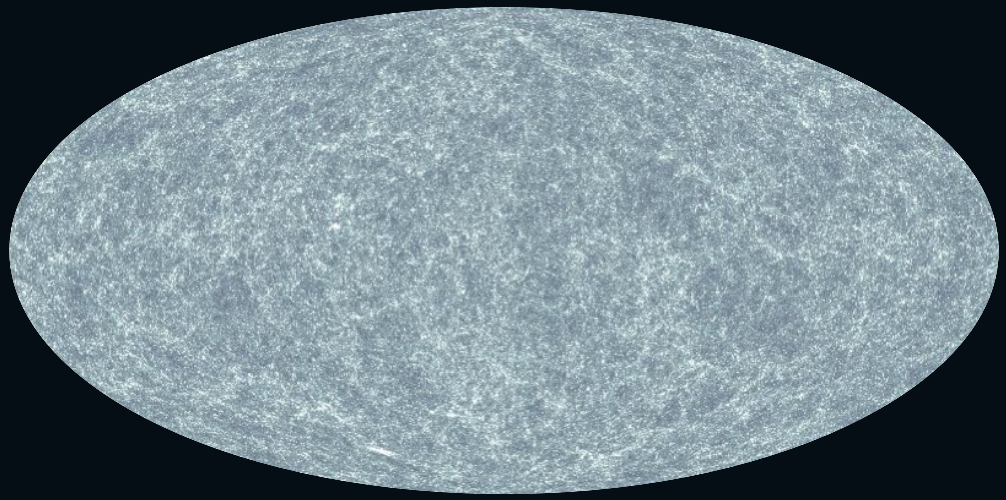
κ is a weighted measurement of the density field.

- The presence of any mass bends the light passing, including the galaxies we observe in the field.
- Weak lensing is an intrisecally **statistical measurement**. It gives information about the matter field.
- Convergence (magnifies size) and shear (tangentially stretches).
- To see some WL results check Jia, Daniela, Joachim, Sihao talks using stage-III surveys.

Credits: H. Miyatake

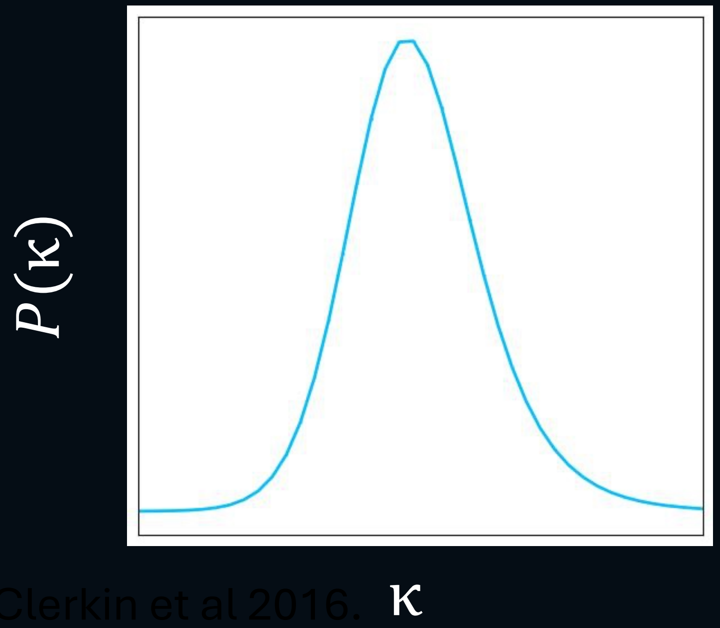
Weak lensing κ-maps: a machine learning problem

CosmoGrid data (Kacprzak++2023)



- Simulations of WL may be **expensive to run when a complete statistical analysis is required** (e.g covariance matrix).
- **Extract information from the non-linear scales**, where gravity and baryonic physics have an impact requires high-resolution maps.
- The convergence field is well approximated by a log-normal distribution (also the density field). **It can be described by a few parameters**.

Can we encode/decode this information and replicate high-fidelity maps with a limited number of simulations?



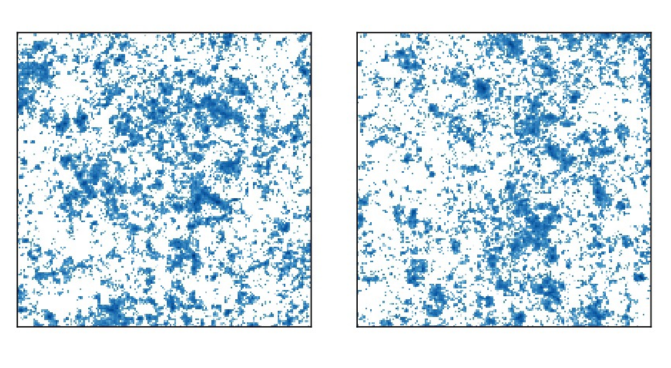
Clerkin et al 2016.

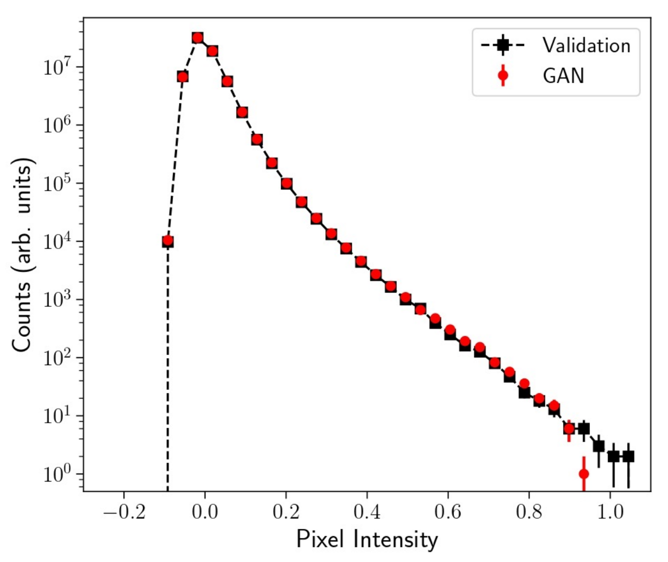
$$P(\kappa) = \begin{cases} \dfrac{\exp\left[-\dfrac{(\ln(\kappa_0+\kappa)-\mu)^2}{2\sigma^2}\right]}{\ln(\kappa_0+\kappa)\sqrt{2\pi}\sigma} & \text{for } \kappa > -\kappa_0, \\ 0 & \text{otherwise,} \end{cases}$$

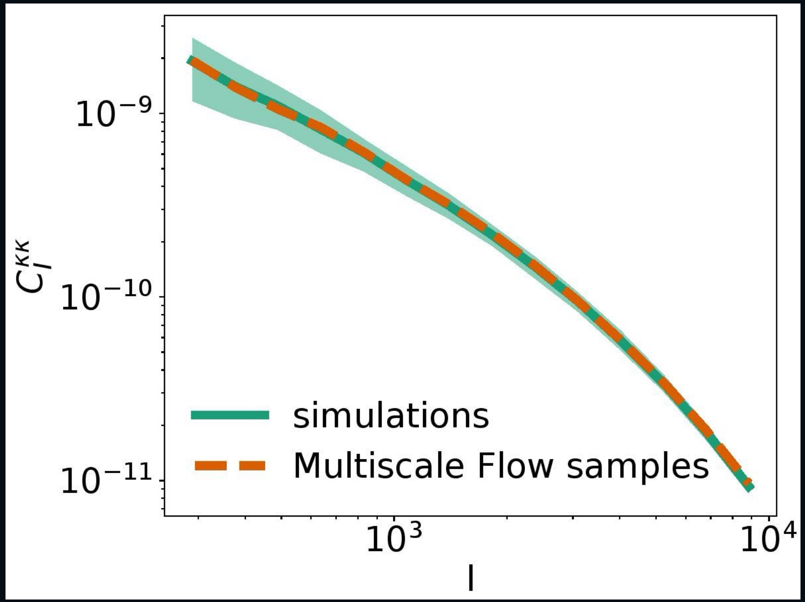
ML for field statistics

CosmoGAN
Mustafa++2019

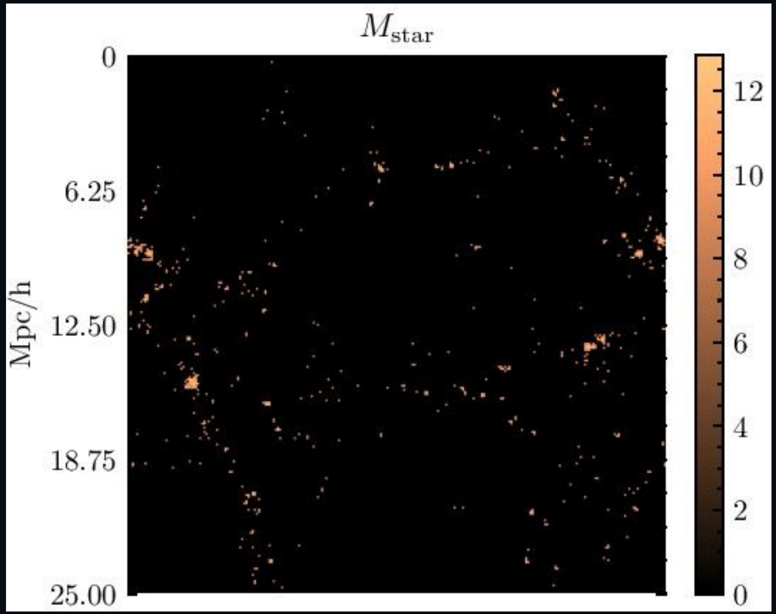
κ-maps



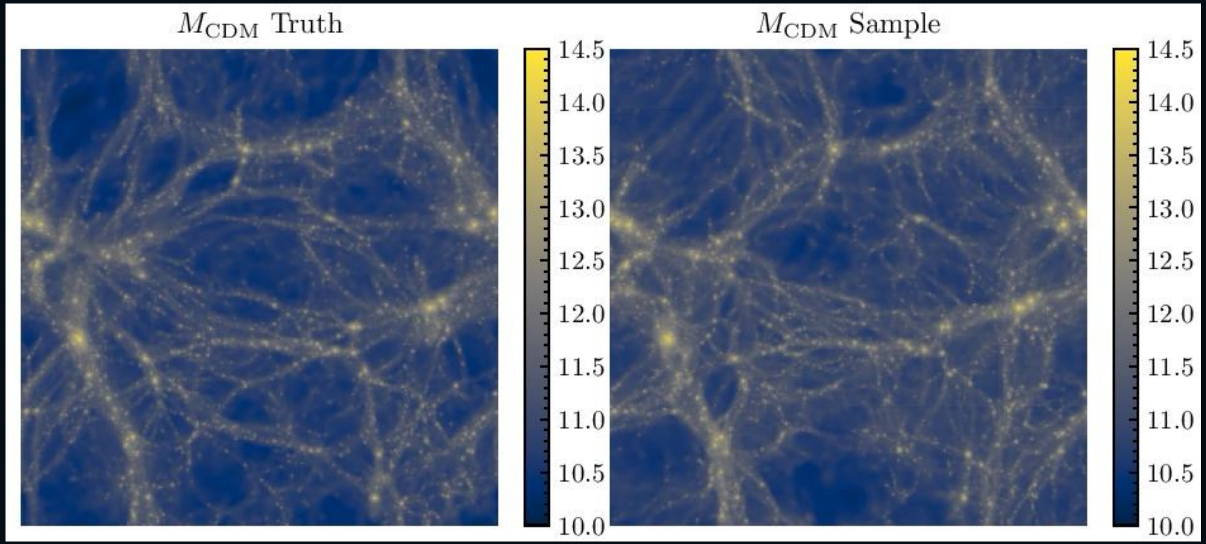




Multiscale flow
Dai & Seljak 2023

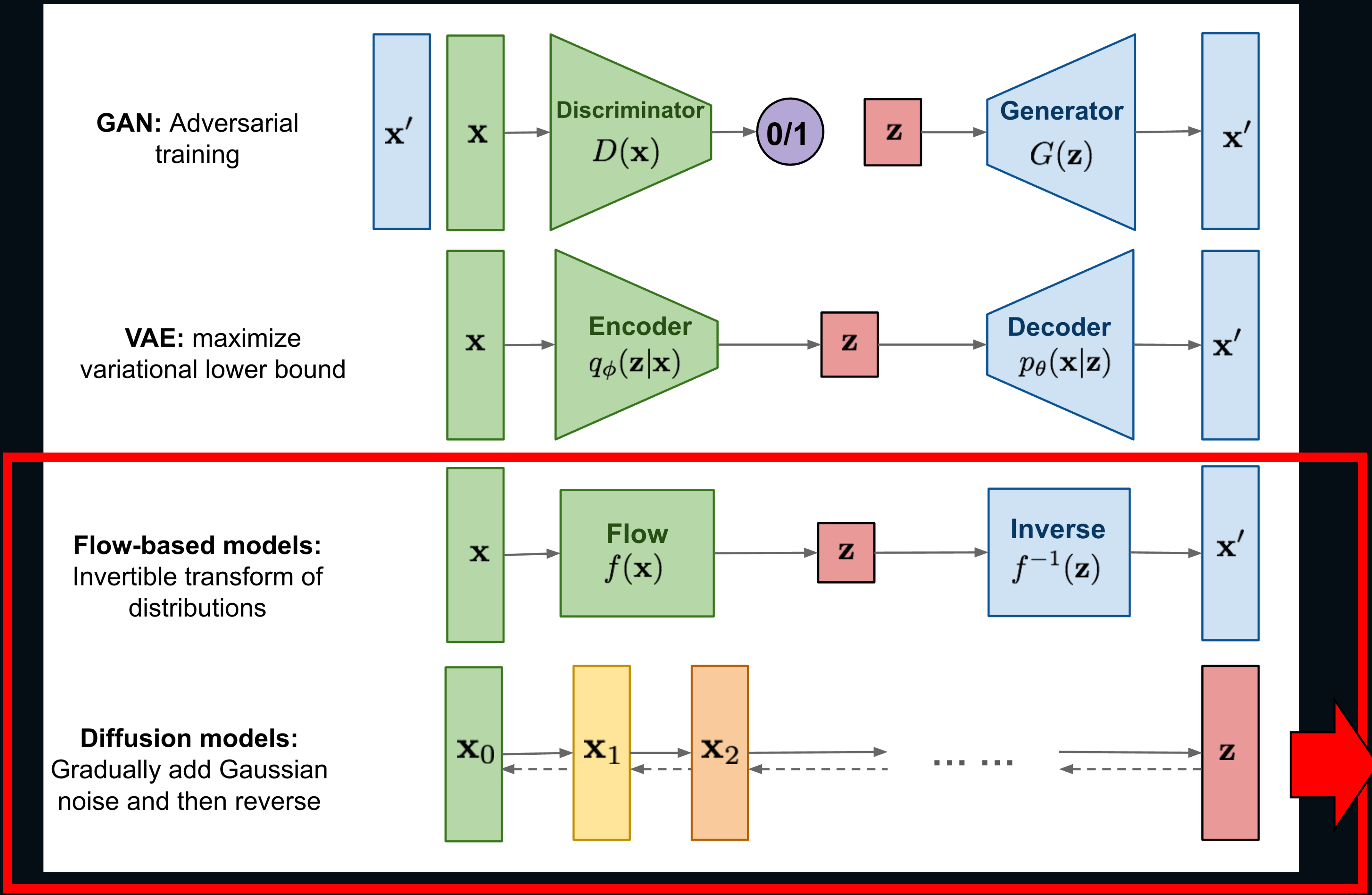


Park++ 2023



And more examples!!

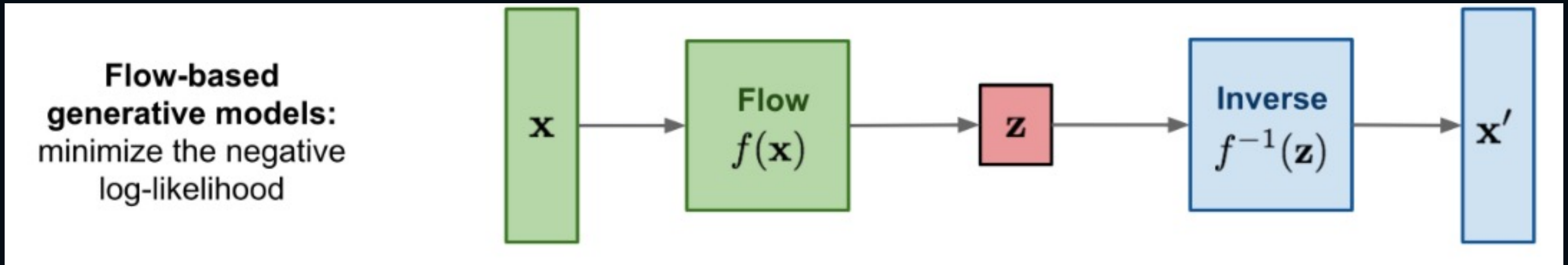
Generative models: A novel option



Used by most AI image generators

https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

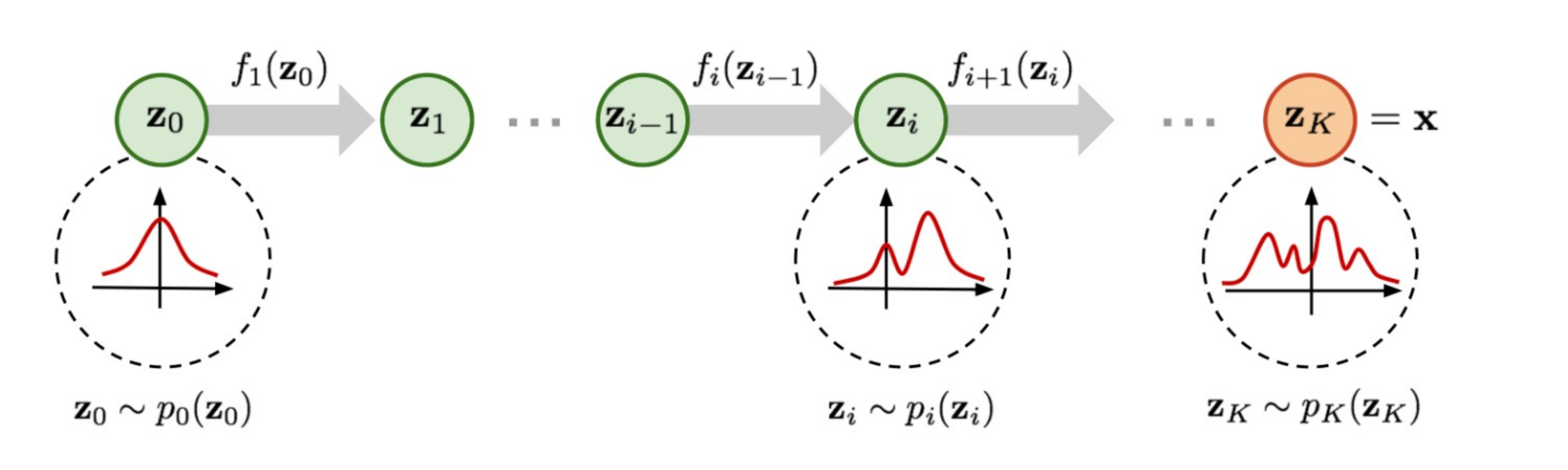
Generative models: The normalizing flow



$$\int p_x(x)dx = \int p_z(z)dz = 1 \qquad \text{(by definition of a probability distribution)}$$

$$\Leftrightarrow p_x(x) = p_z(z)\left|\frac{dz}{dx}\right| = p_z(f(x))\boxed{\left|\frac{df(x)}{dx}\right|}$$

Determinant of the Jacobian matrix

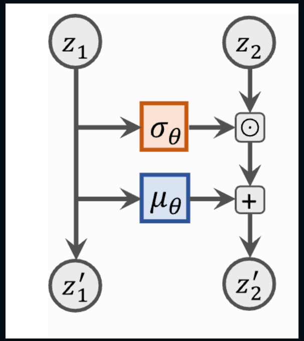


NF reproduce the PDF of the real data

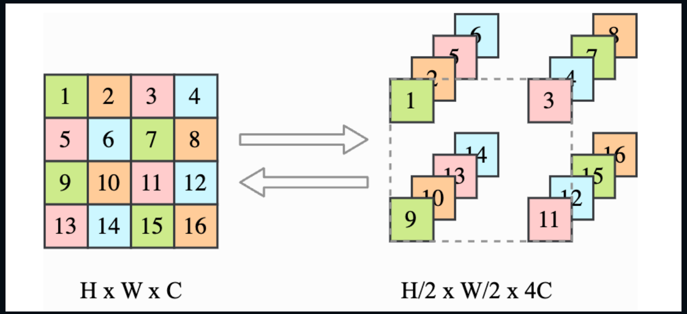
https://colab.research.google.com/github/phlippe/uvadlc_notebooks/blob/master/docs/tutorial_notebooks/tutorial11/NF_image_modeling.ipynb
https://lilianweng.github.io/posts/2018-10-13-flow-models/

Normalizing flow architecture

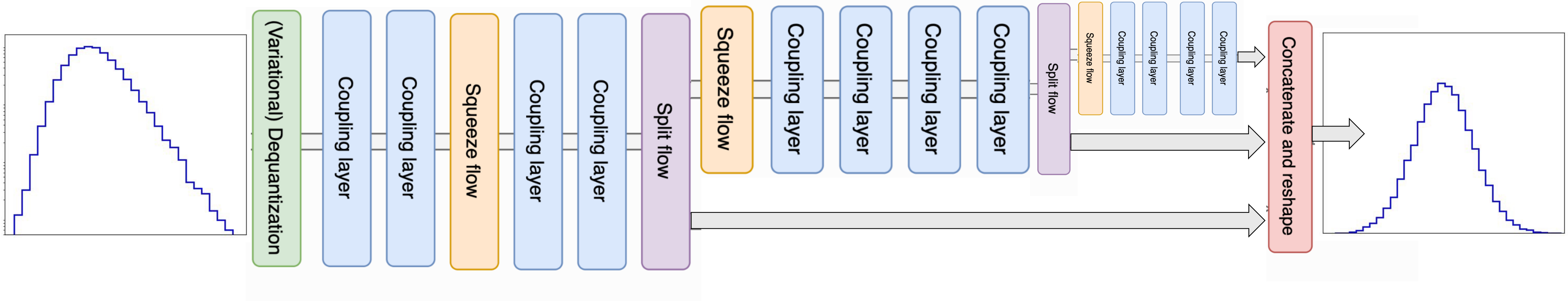
Coupling layers:
(scale, rotate..)

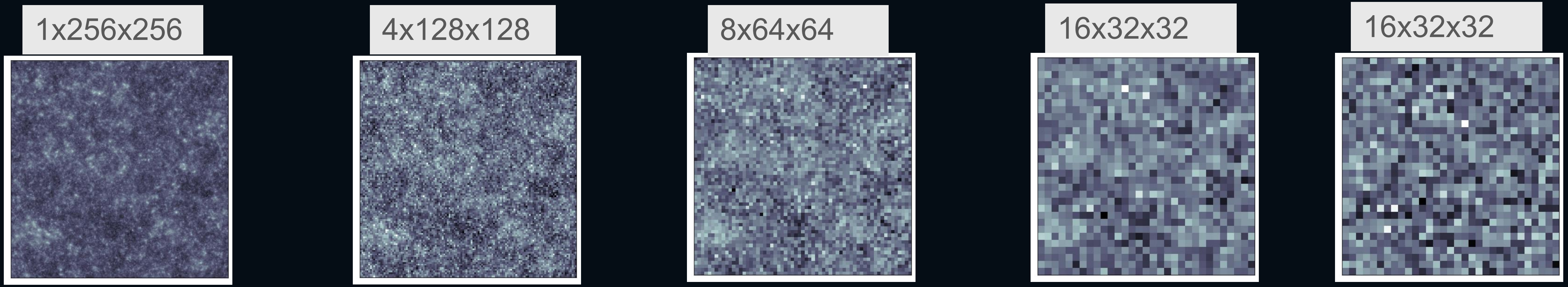


Squeeze and split:



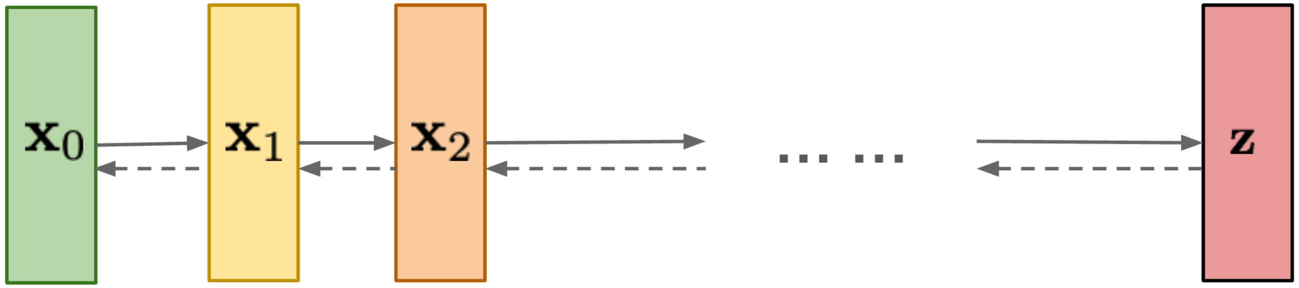
Similar to a super resolution CNN!

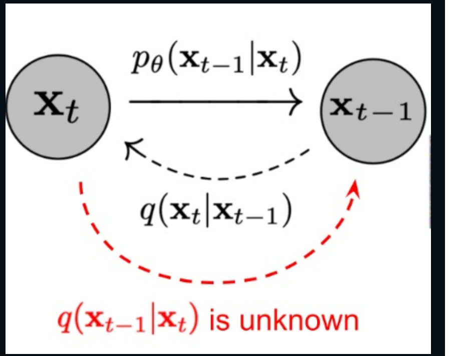


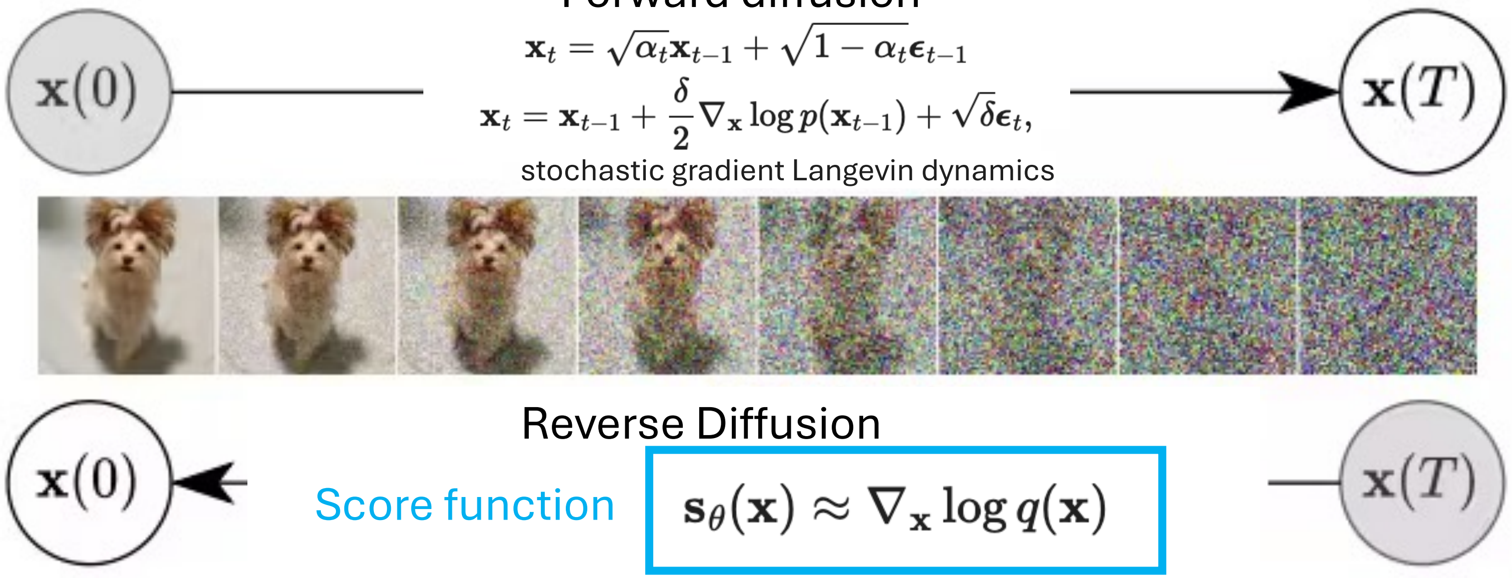


Score-Based Generative Models

Diffusion models:
Gradually add Gaussian noise and then reverse







$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1}$$

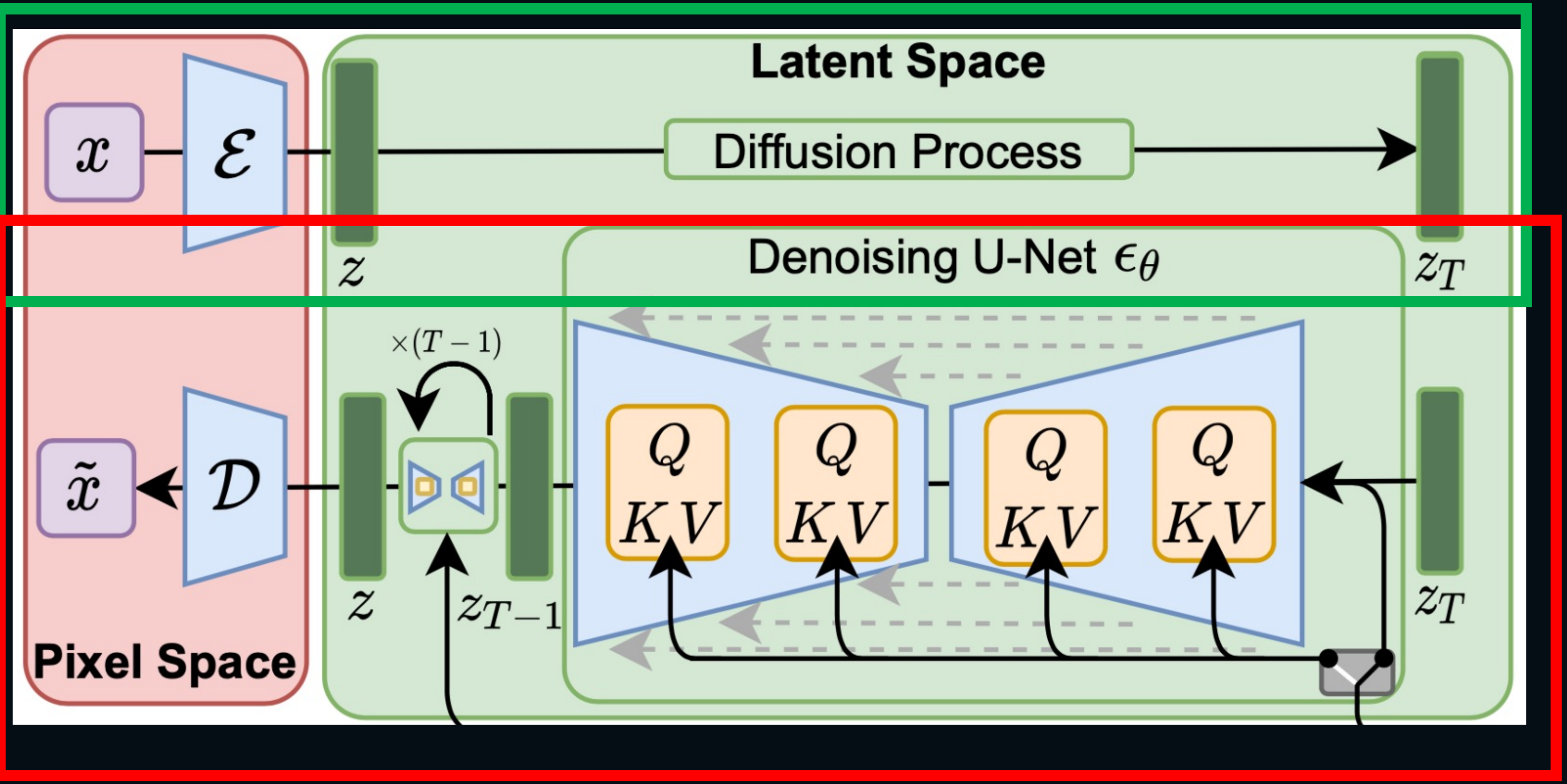
$$\mathbf{x}_t = \mathbf{x}_{t-1} + \frac{\delta}{2}\nabla_\mathbf{x} \log p(\mathbf{x}_{t-1}) + \sqrt{\delta}\boldsymbol{\epsilon}_t,$$

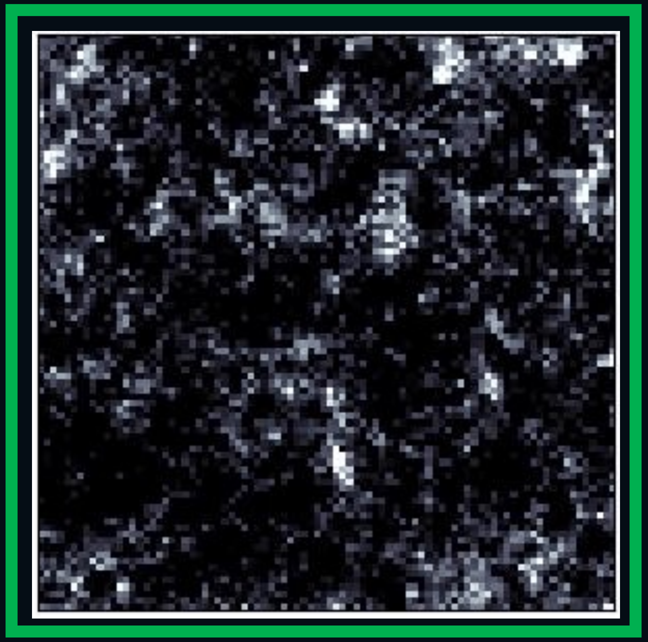
$$\boxed{\mathbf{s}_\theta(\mathbf{x}) \approx \nabla_\mathbf{x} \log q(\mathbf{x})}$$

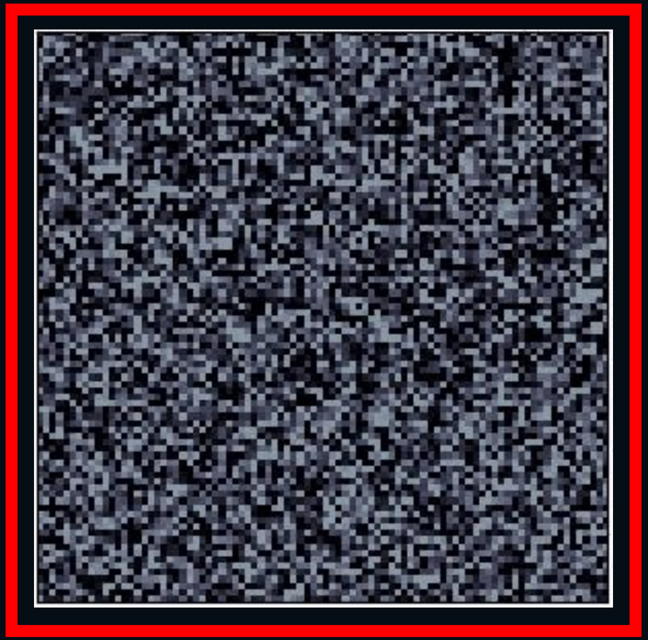
If we can learn the gradient and reverse the above process, we will be able to recreate the true sample from a Gaussian noise input.

https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

Stable diffusion



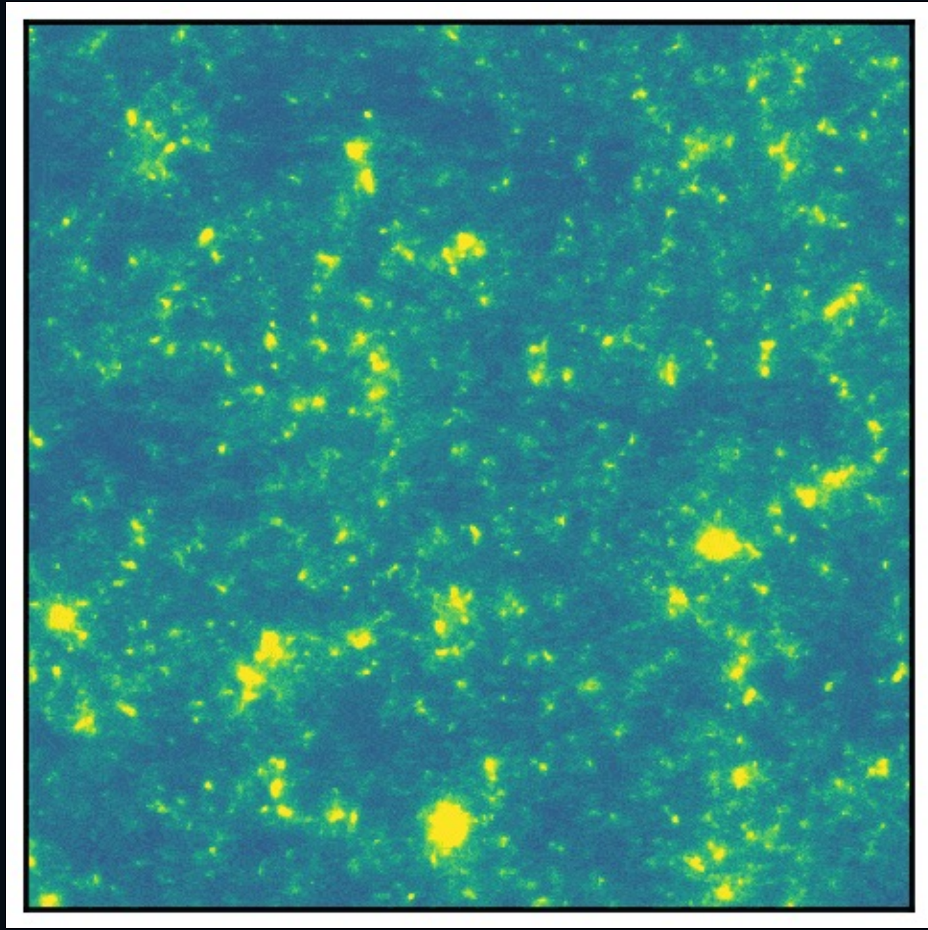


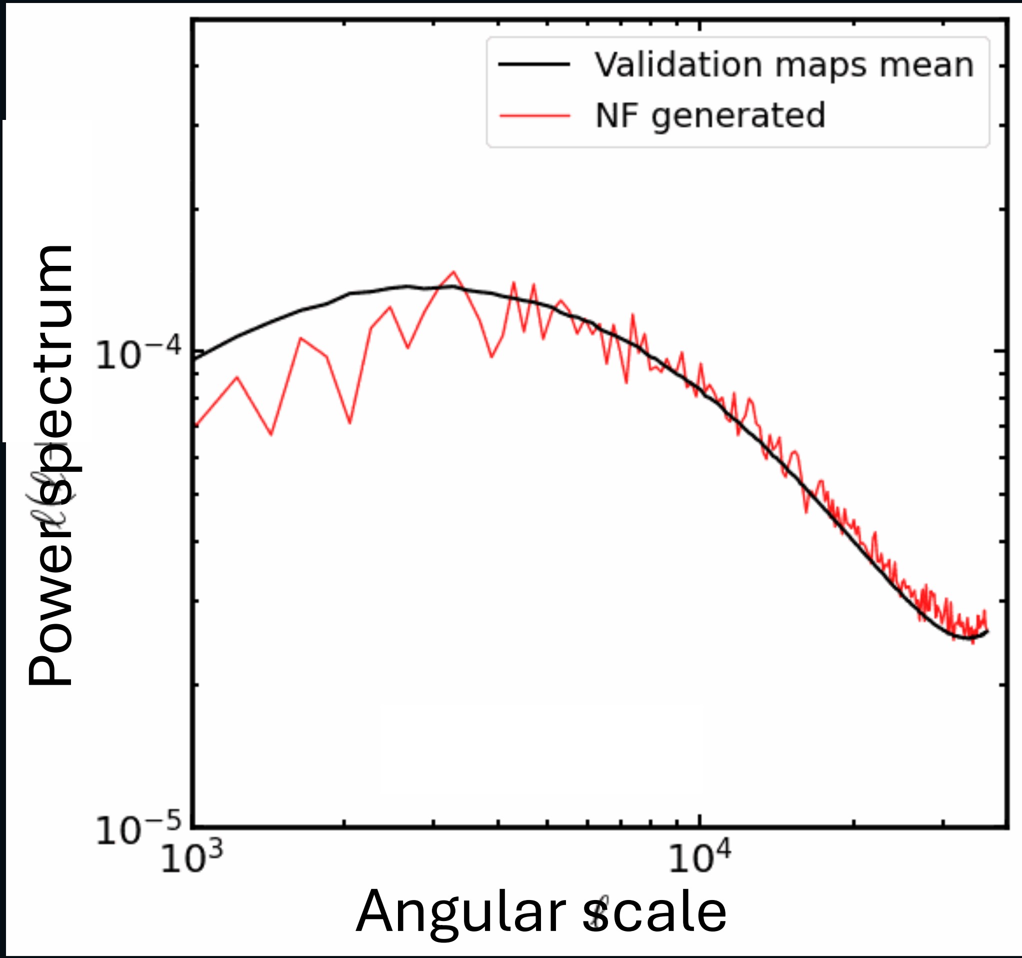


Generating new κ-maps

Trained on SLICS simulations ~1000 pseudo-independent 10x10 sq. deg maps for covariance calculation.
We use data augmentation to expand the data set to 60000 5x5 sq. deg maps.

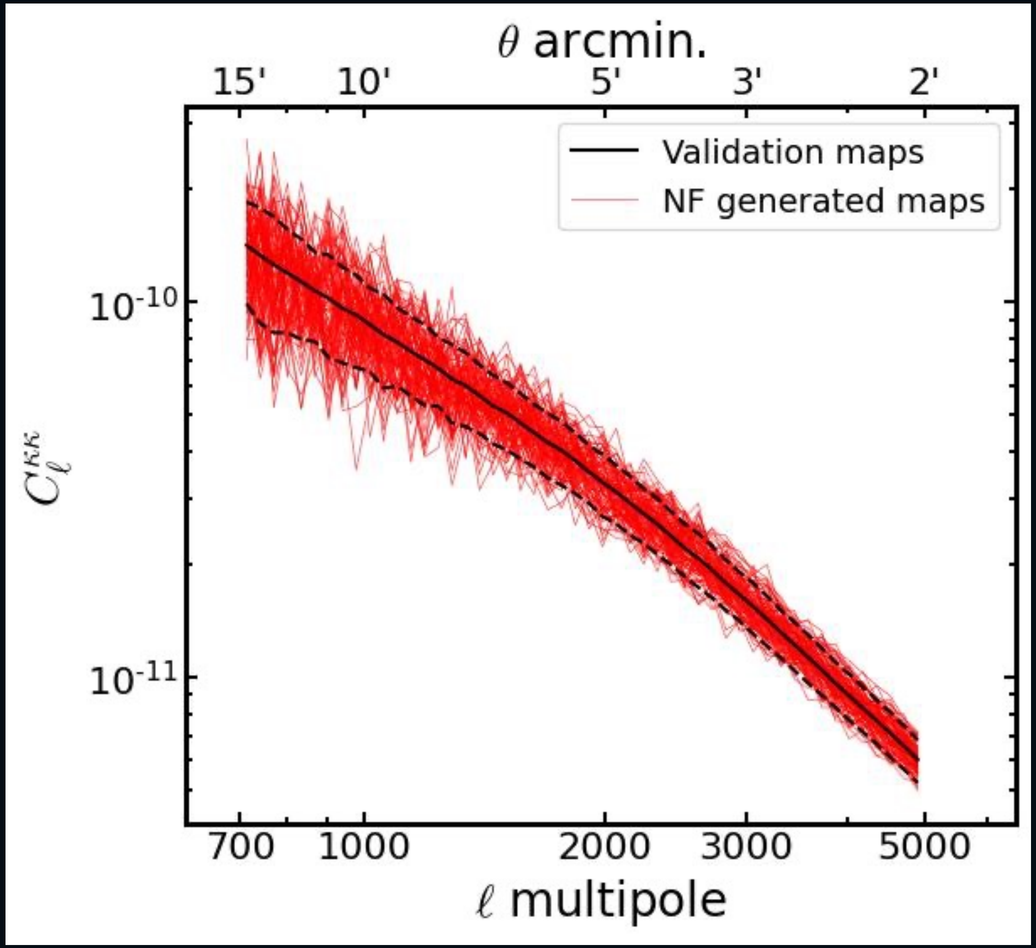
Normalizing flow generated maps

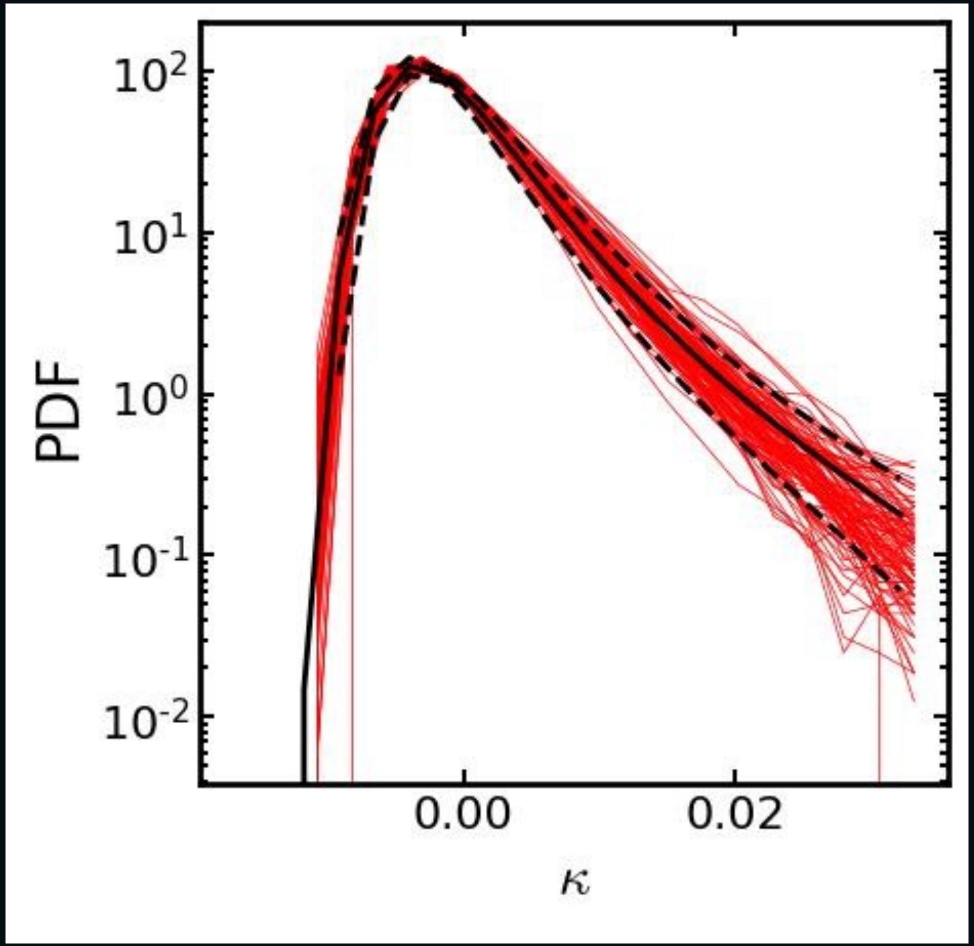




Summary statistics

ML models can recover the mean values of the original maps

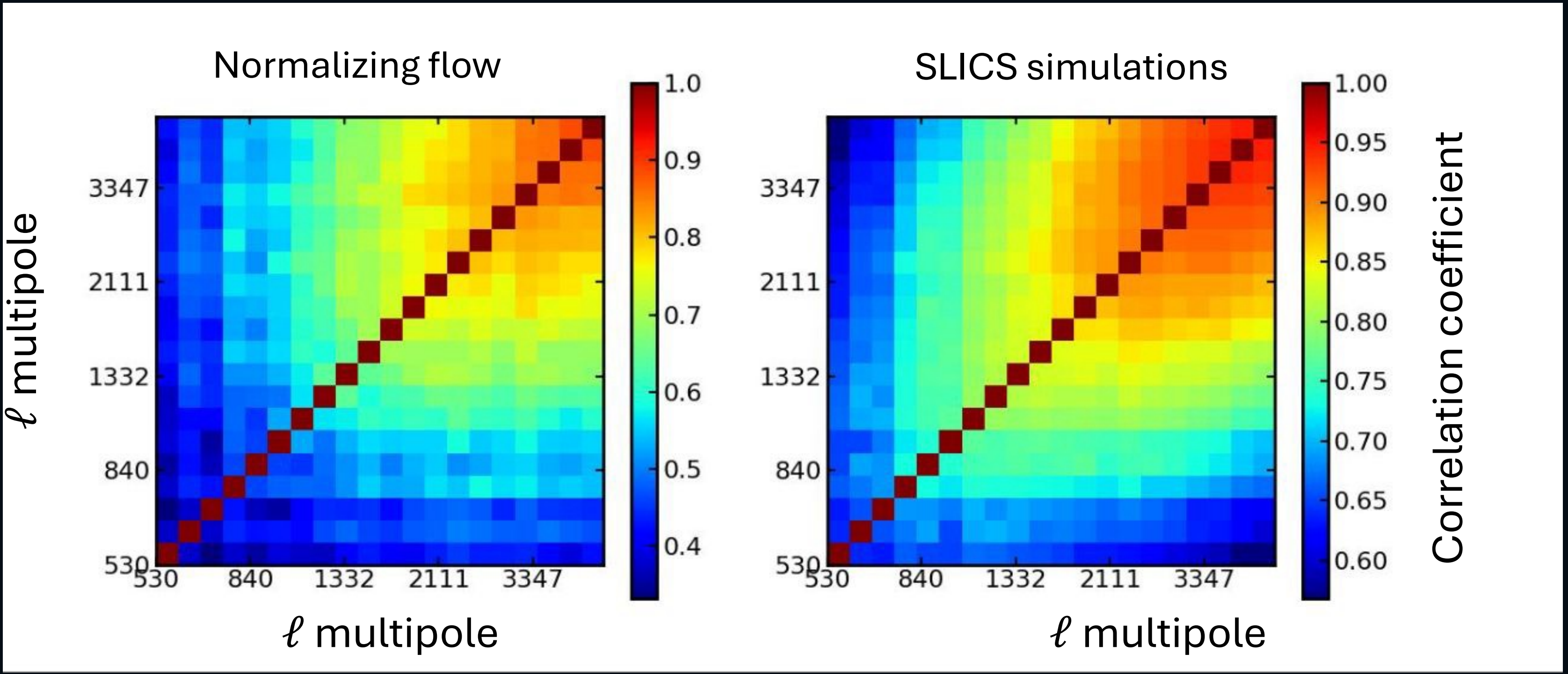


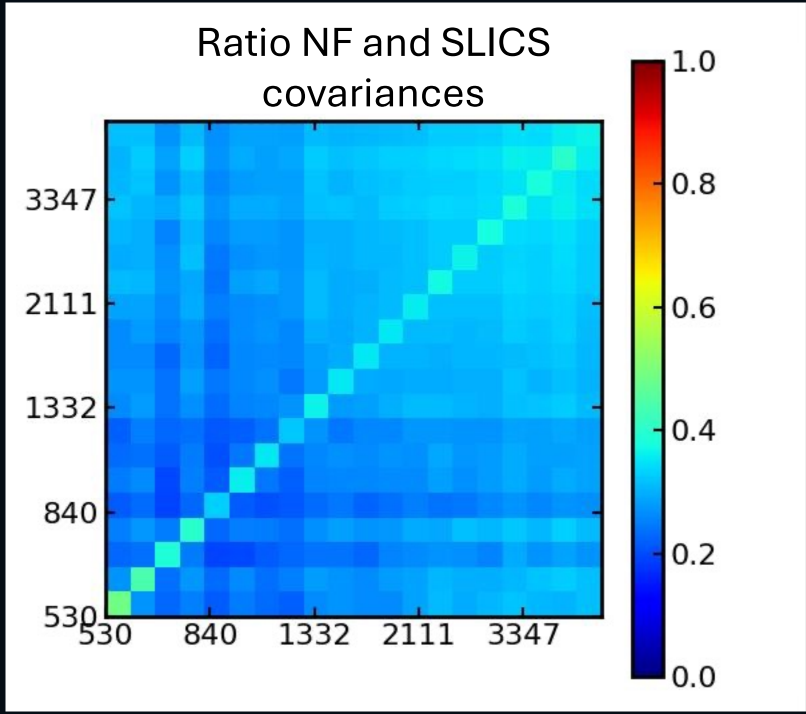


New NF network trained using only 100 SLICS maps to generate WL maps. These generated maps recover the statistics for PS and PDF. However, there is a limitation on recovering the variance and covariance.

What about covariance?

Is hard to recover the covariance using the pixel information only. SLICS simulations uses ~1000 pseudo-independent simulations to calculate the covariance matrix. We variate the number of simulations used to generate the training set.





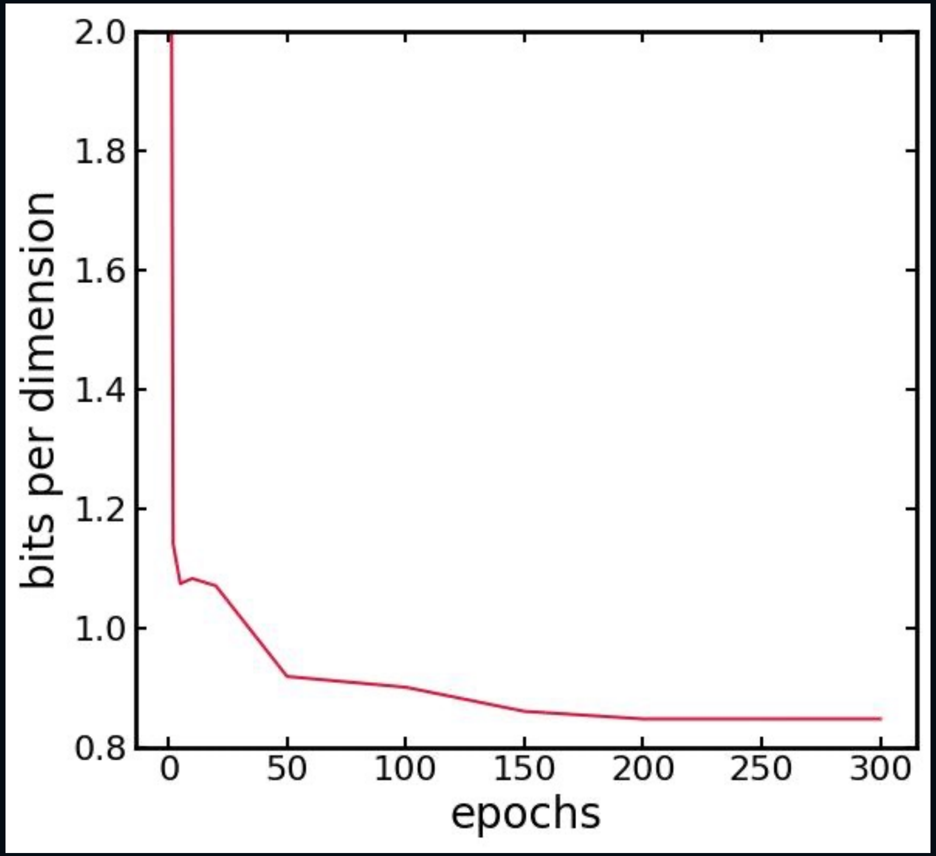
The power spectrum covariance is not purely statistical, it has physical information (including terms of the trispectrum).

Metrics (Loss function)

Bits per dimension (from KL div. for images):

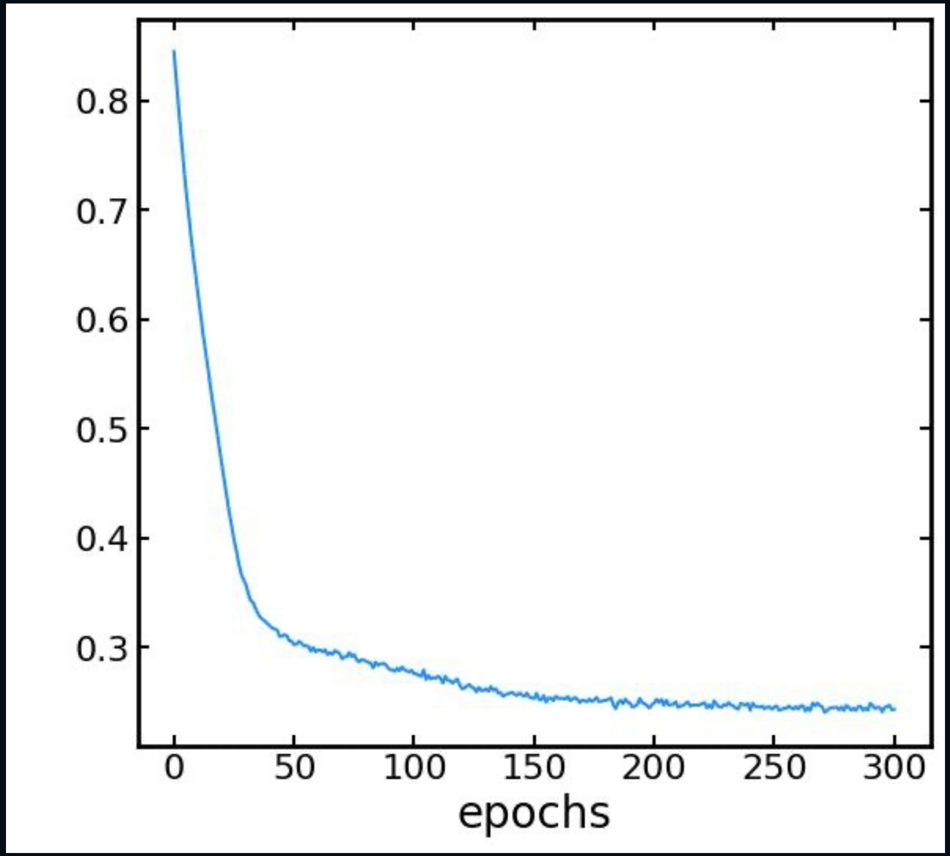
$$\mathrm{bpd} = \frac{\mathrm{nll}}{\log 2}\left(\prod_i d_i\right)^{-1};\ d_i\text{: dimensions (pixels)}$$

nll: negative log-likelihood.



Diffusion model: L1-loss (MAE)

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{pix}}}\sum_i |y_t - y_{\mathrm{pred}}|$$



Loss applied to Gaussian fields: Gaussian base (NF; prior), and Gaussian random noise (DM).

Limitations of NF/DM

They are extremely hard to train!

Normalizing flow

- Can be biased towards the mean (KL divergence).
- Limited to invertible transformations only.
- Accuracy limited to how informative the prior is.

Diffusion model:

- Several examples must be given during training.
- Noise is entangled to small scale information (more sensitive to shot noise).
- Sampling process can have large number of steps.

However, they still outperform VAEs and GANs for imaging production!

Some few test to try yet!

Summary and conclusions

- Machine learning models are valuable for encoding information of the LSS. Statistics for DM and WL fields using ML can help to constraint the cosmological model.

- Models can learn non-Gaussian features of the convergence field. Generative models aim to recover information of theory and simulations. **Information at the pixel level is recovered**.

- Normalizing flow and diffusion model are broadly used as they keep the dimensionality of the data. They promise large variability of novel information. **However, some cosmology information is lost.**

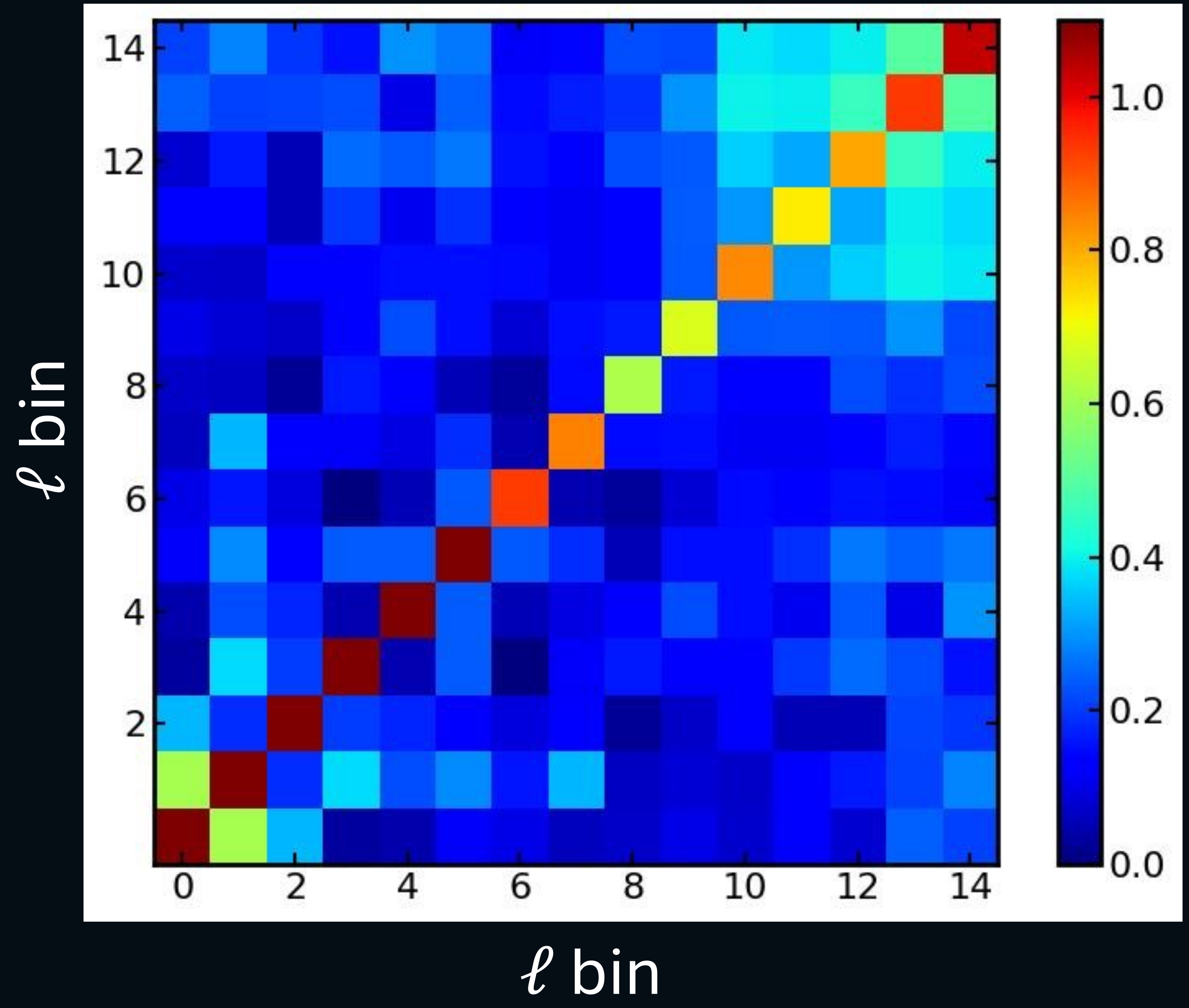
- While powerful to recover the mean values of the summary statistics, there is an underestimation of variance and covariance.

DeepLensing flow:

Thank you!

Covariance for DM models

Ratio between DM generated map covariance and ground truth



- Trained on CosmoGrid data (10x10 sq. deg patches).
- Recover the variance but not the off-diagonal terms.
- For DM the network (U-net denosier) can go deeper (more parameters) and be trained with more data.

Training and data augmentation

With SLICS simulations we can go from ~1k to ~10k (1OOM) , but we loose large scales.