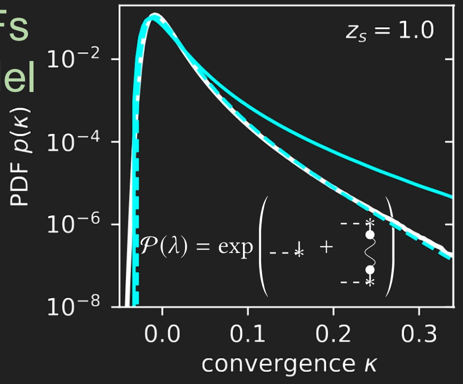
Advancing Cosmology with Robust ML

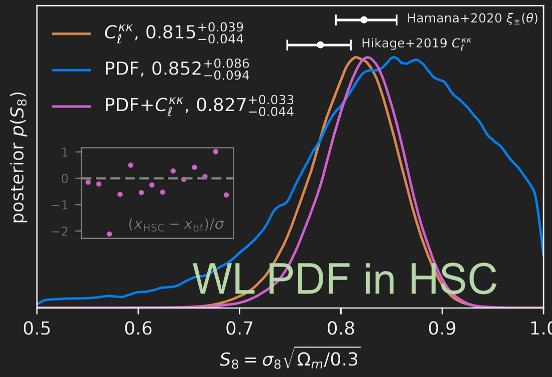
seminar, 10/3/23

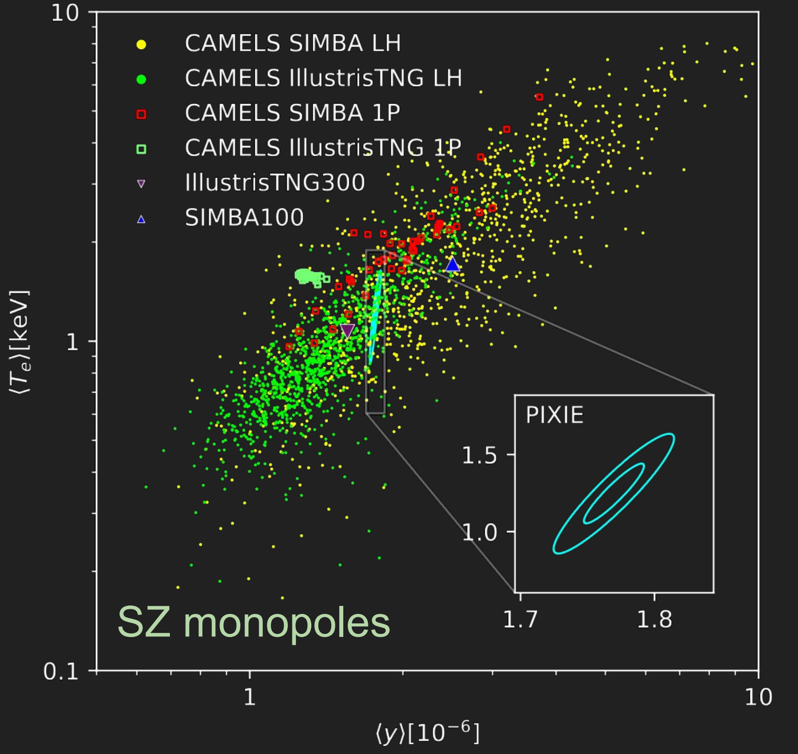
Leander Thiele (Princeton University)

Other things

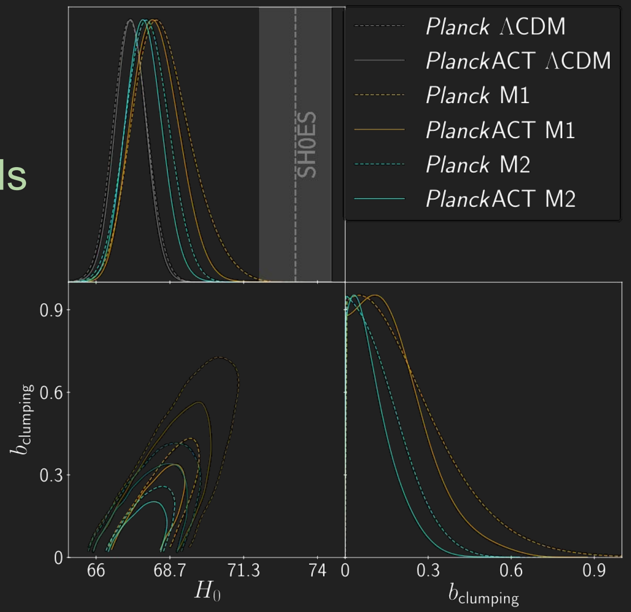
WL & tSZ PDFs from halo model





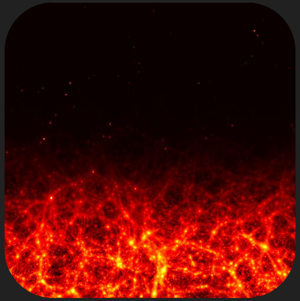


primordial magnetic fields & H_0 tension

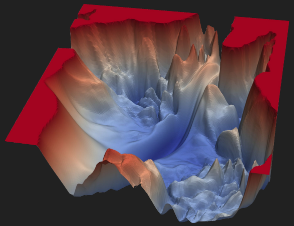



Generative AI is now
safe for the enterprise.
You're welcome.
glean.com | AI-POWERED ENTERPRISE SEARCH.

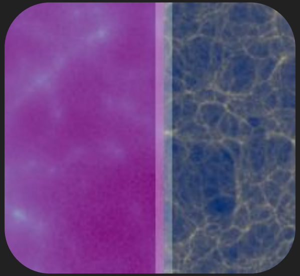

Deep Learning is now
safe for the sciences
You're welcome.
Cal
AI-POWERED ENTERPRISE SEARCH.



1) Quick example: Emulation



2) The past decade in deep learning

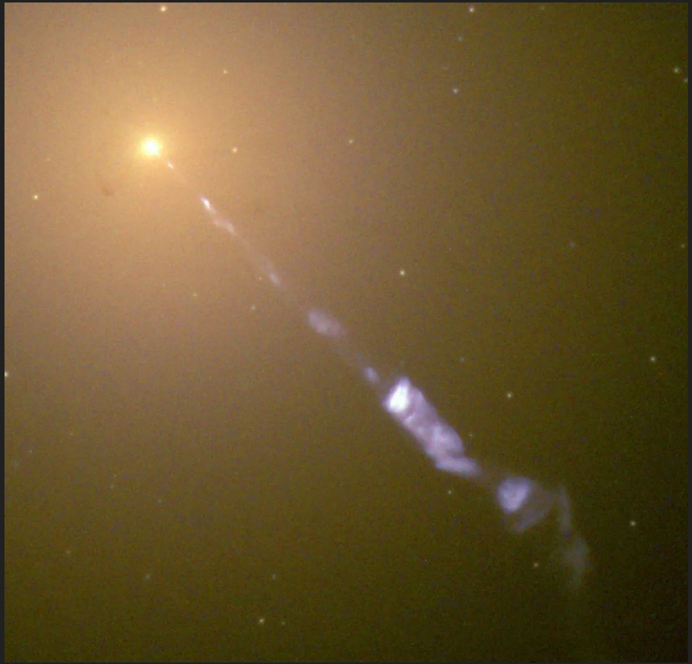


3) Inference

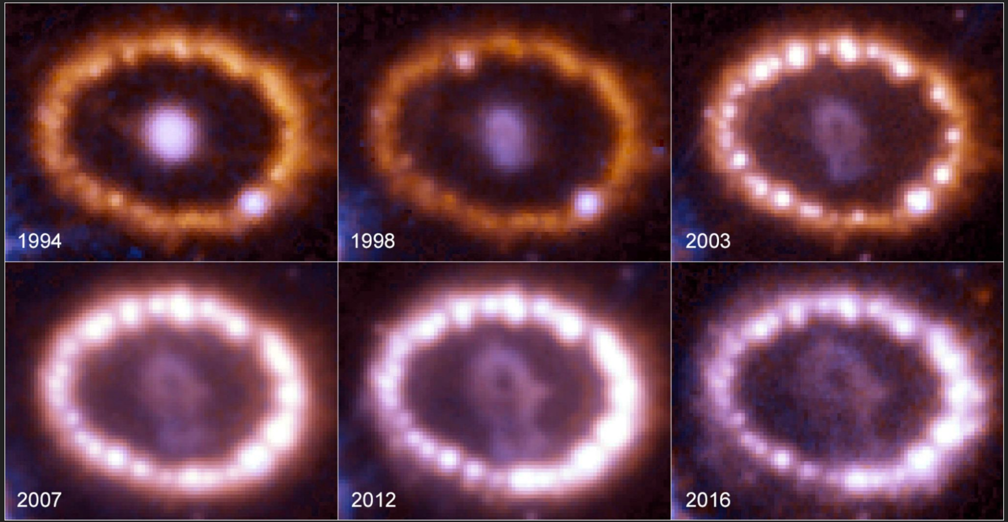
Quick example: Emulation

Deep learning as an approximator for complicated physics

Baryonic feedback

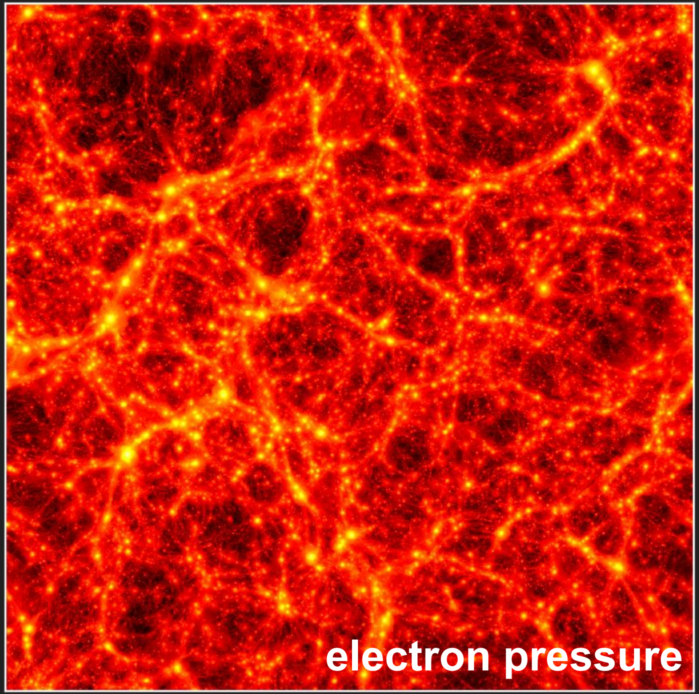


active galactic nucleus

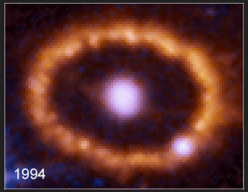


supernova

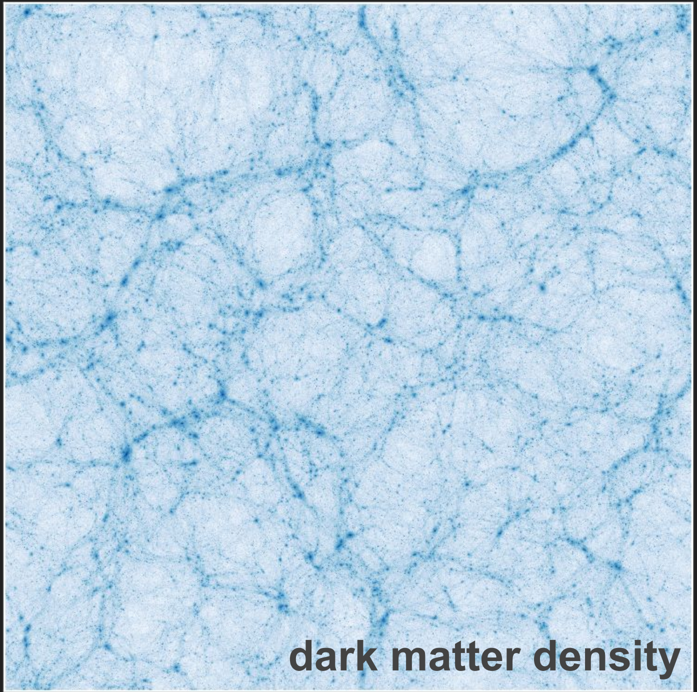
Mapping dark matter to baryons



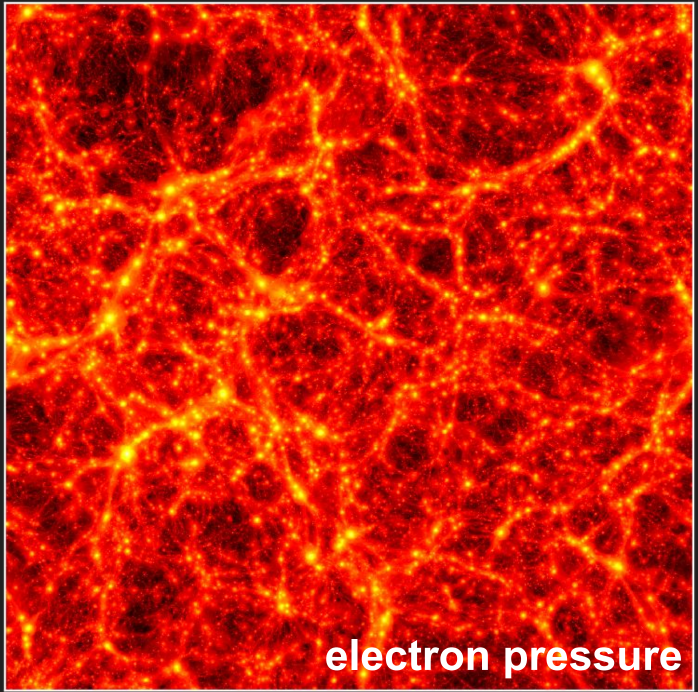
$\partial \mathbf{U}/\partial t+\mathrm{div}\mathbf{F}=0$



Mapping dark matter to baryons

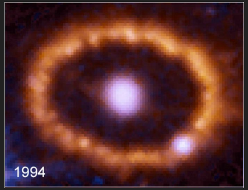


neural network



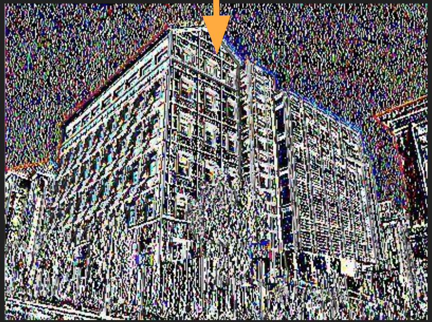
$F = G\, m_1\, m_2 / r^2$

$\partial \mathbf{U}/\partial t + \mathrm{div}\mathbf{F} = 0$

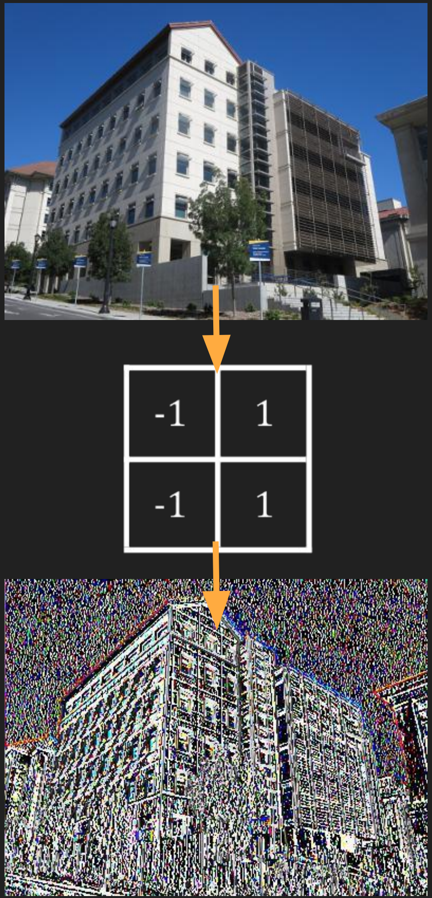


Convolutional approach

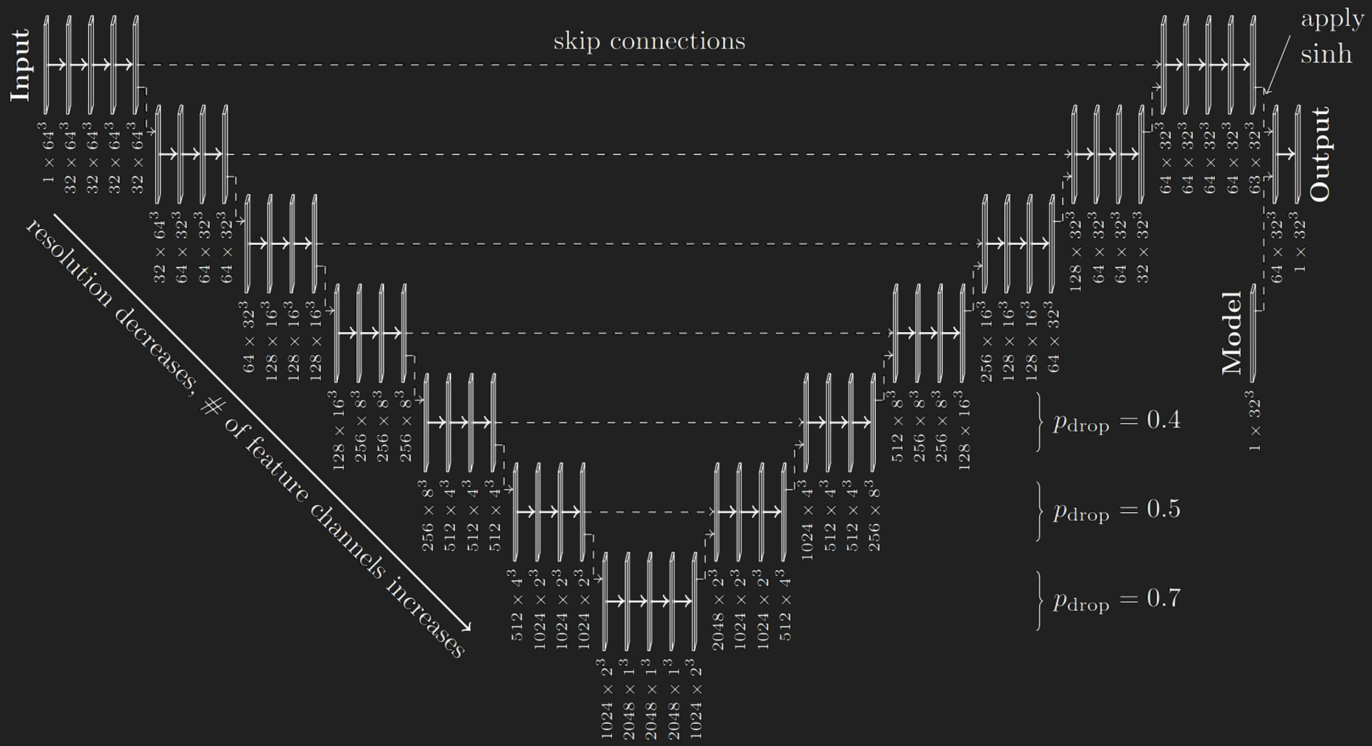
-1	1
-1	1



Convolutional approach

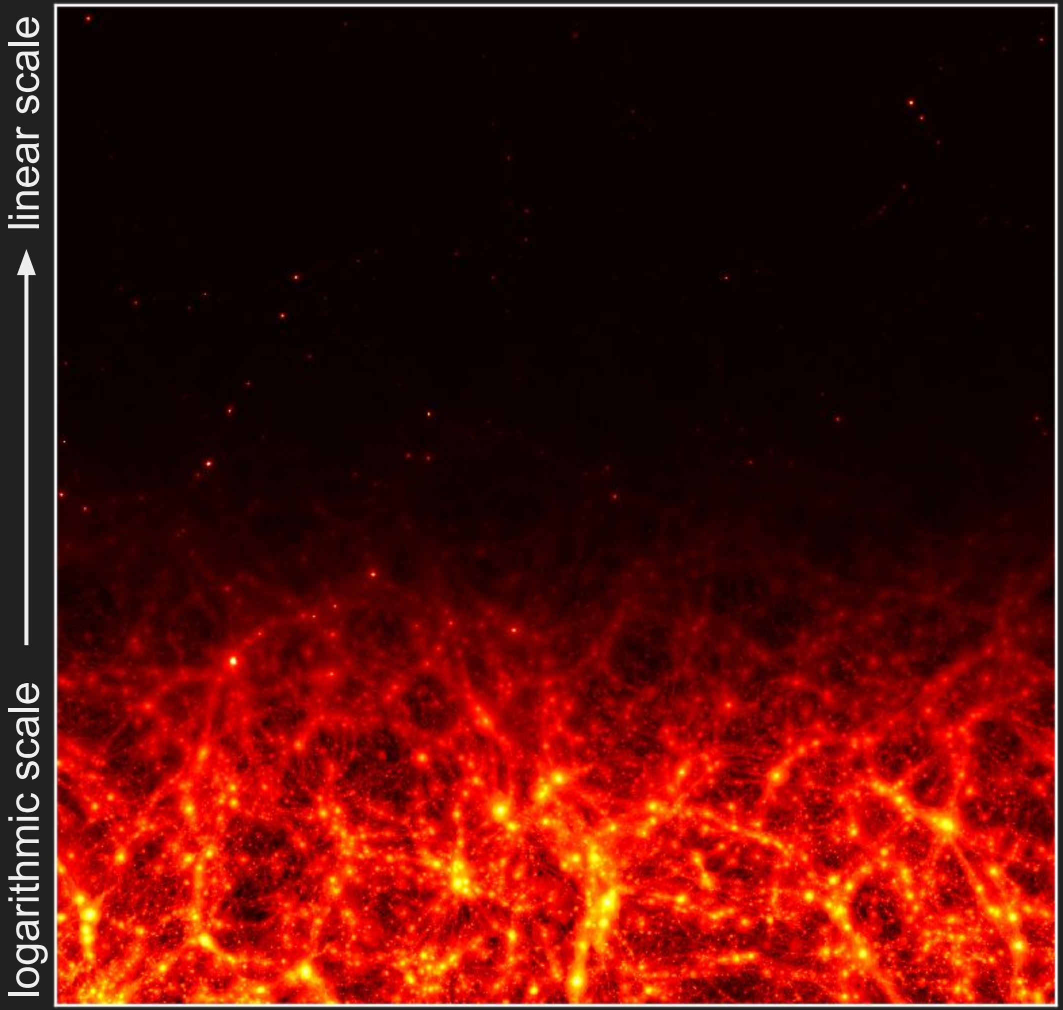


Convolutional Neural Network (CNN)



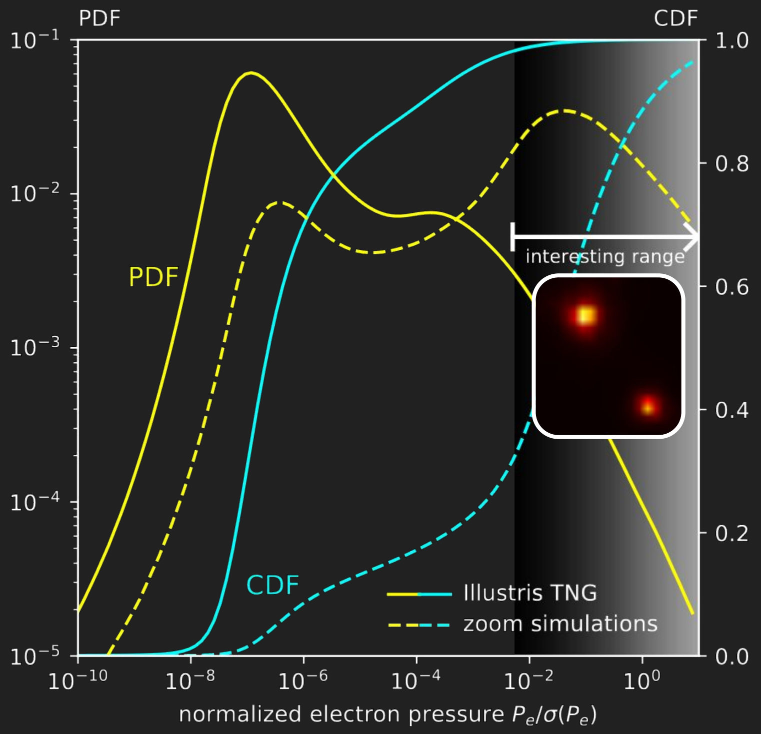
Sparsity problem

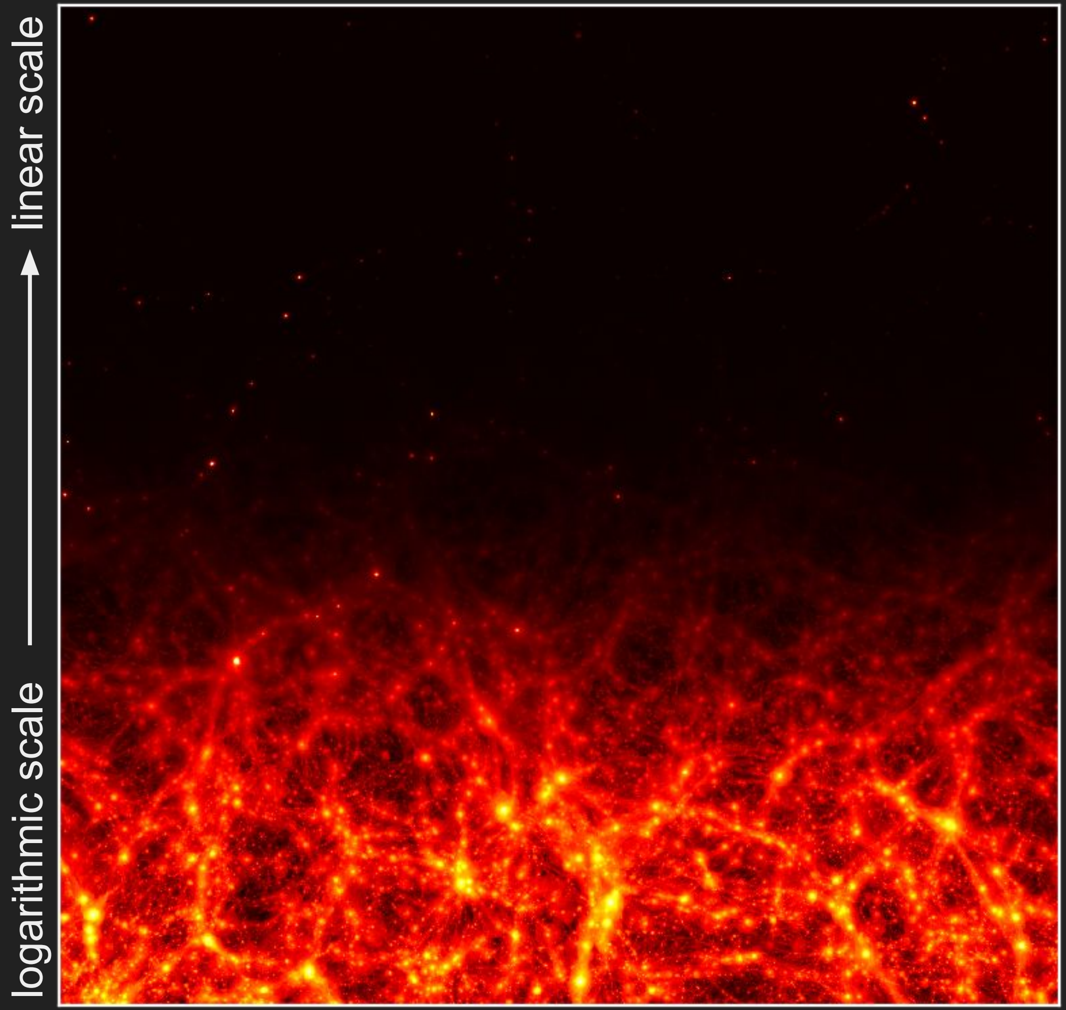
tSZ signal dominated by massive halos → rare objects!



Sparsity problem

tSZ signal dominated by
massive halos → rare objects!

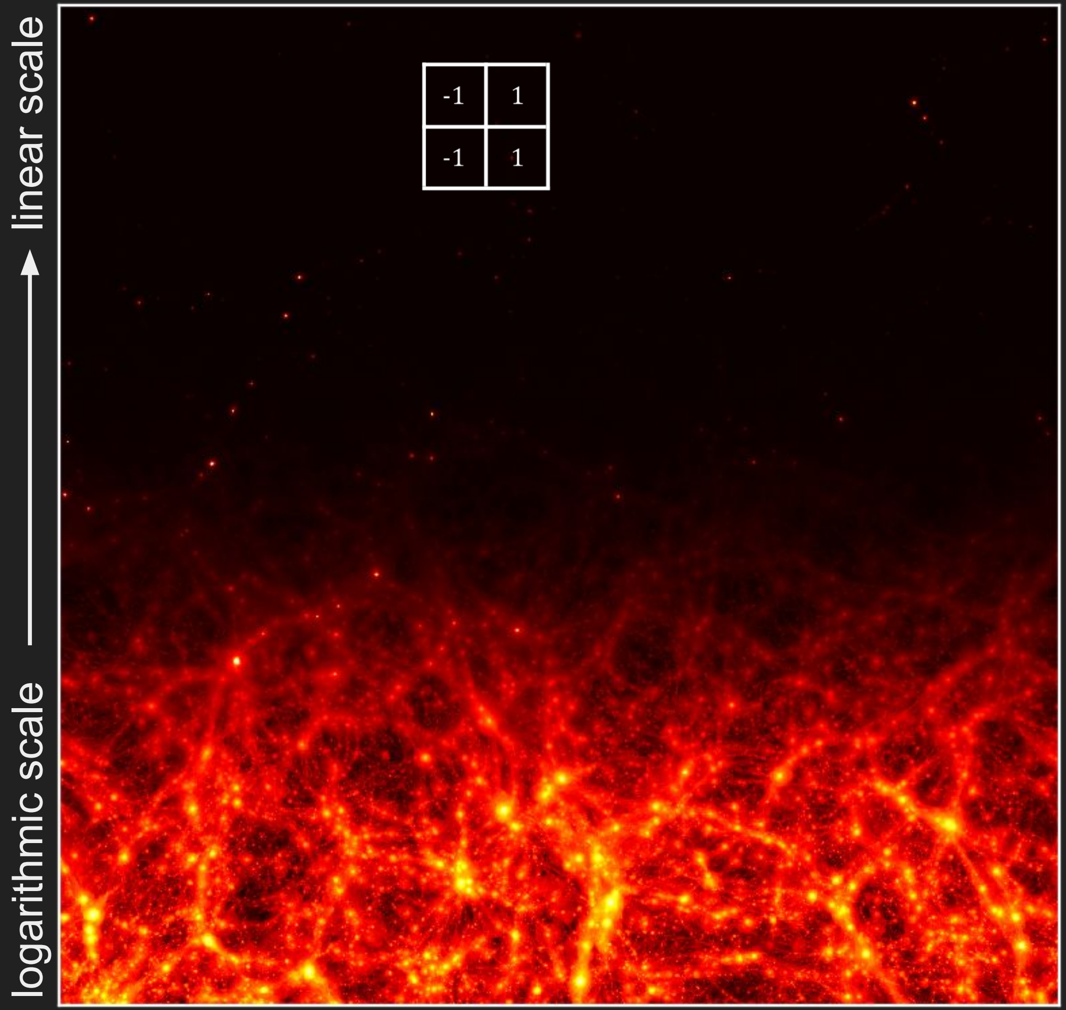




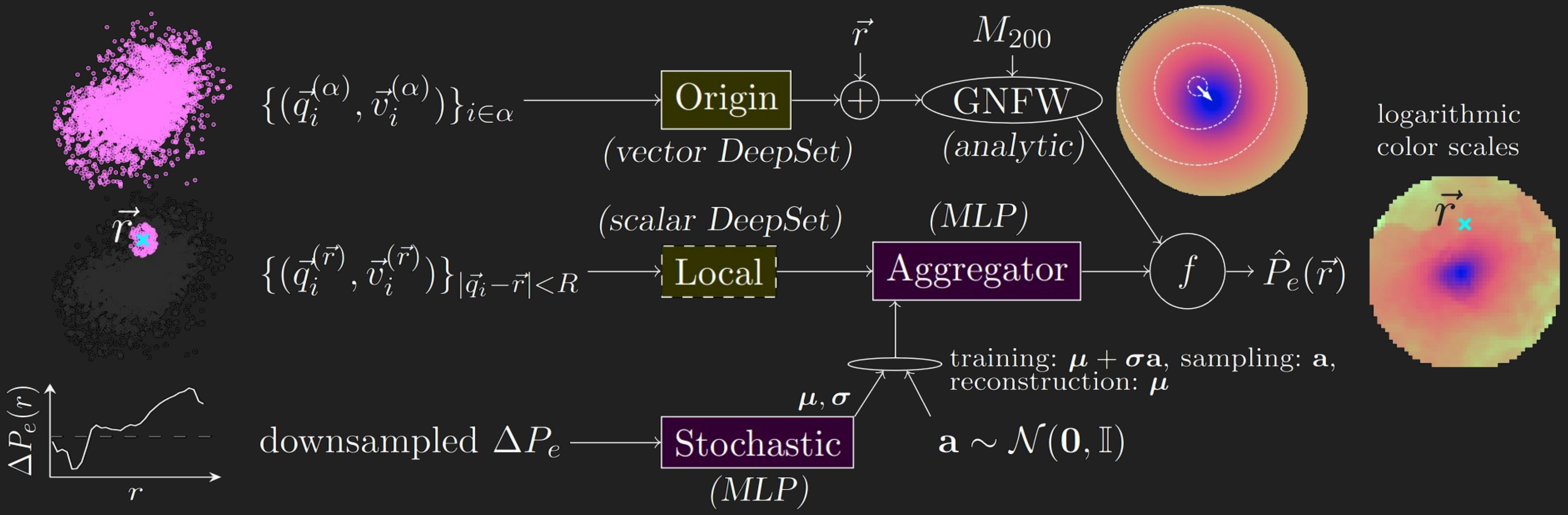
Sparsity problem

tSZ signal dominated by massive halos → rare objects!

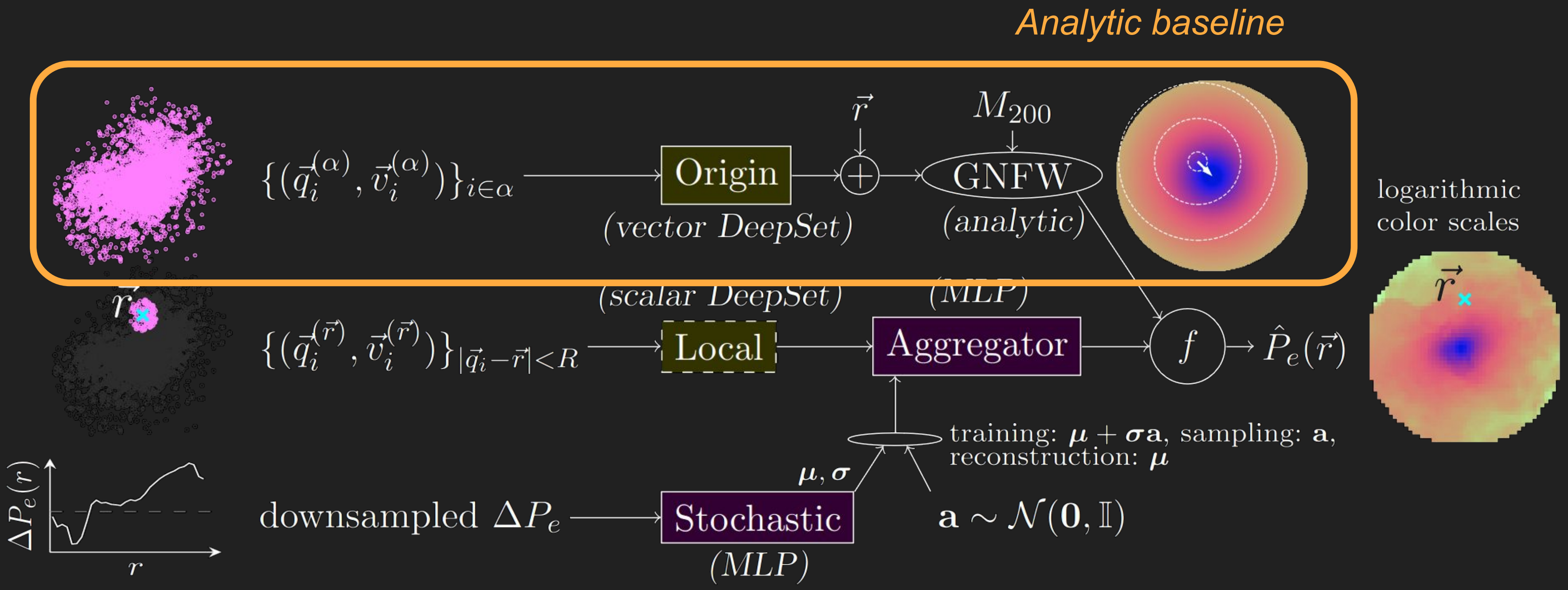
Convolutional approach not ideal.



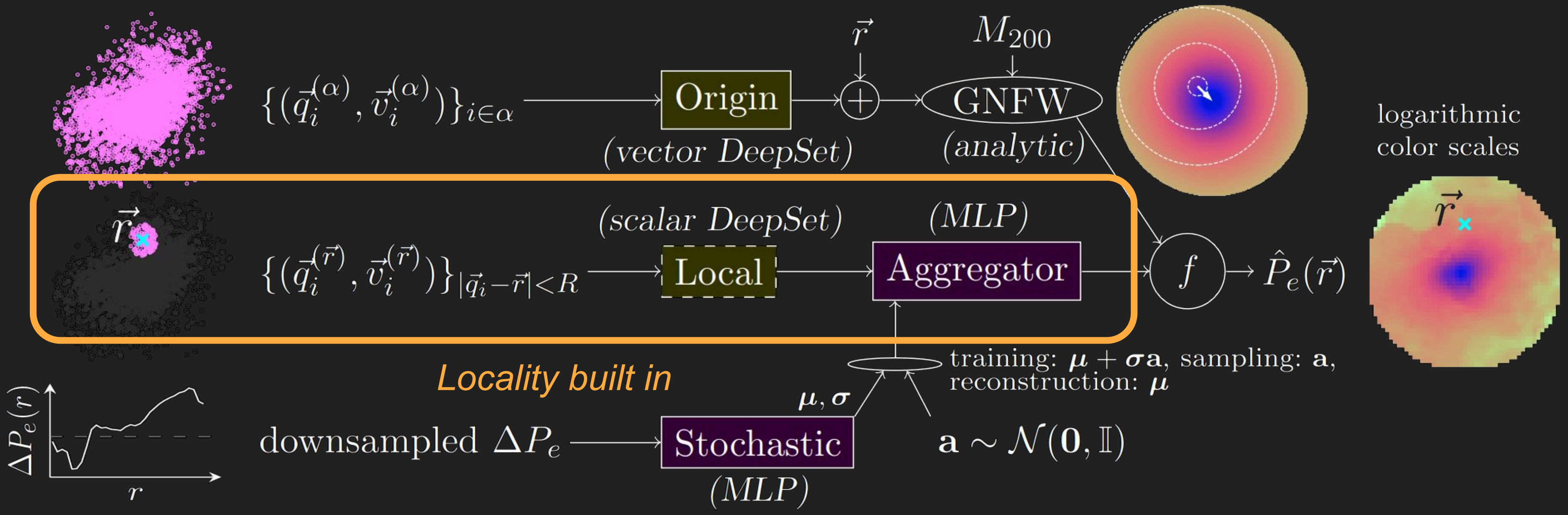
Set-based approach



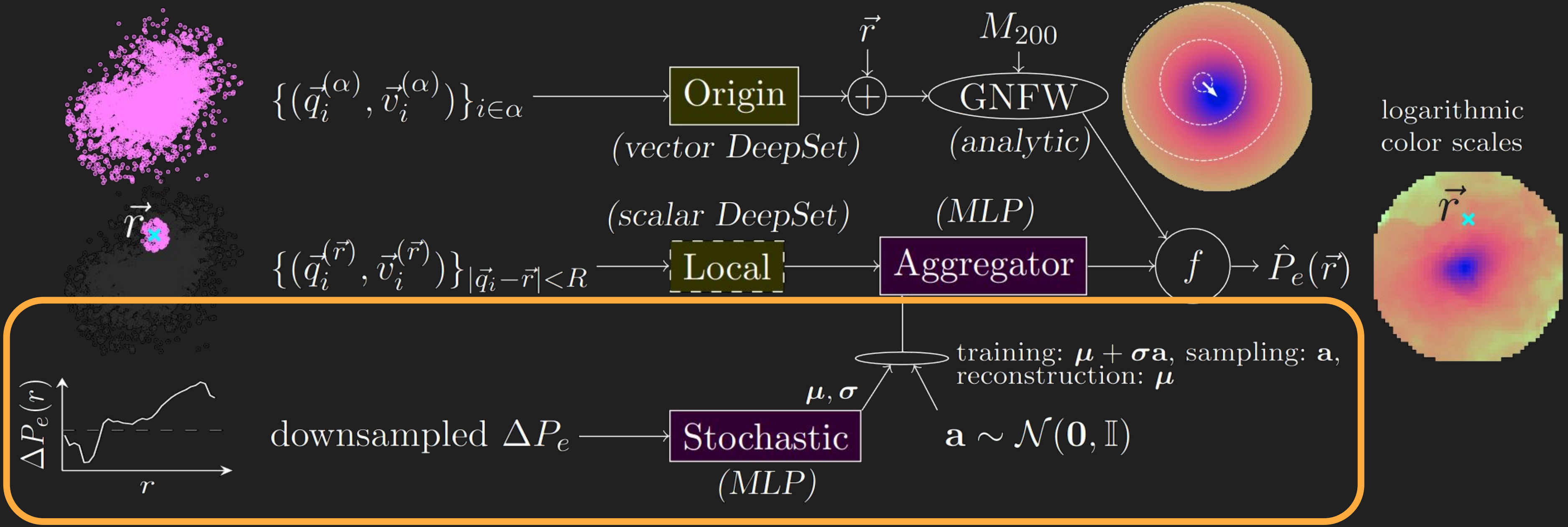
Set-based approach

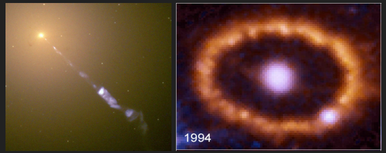


Set-based approach



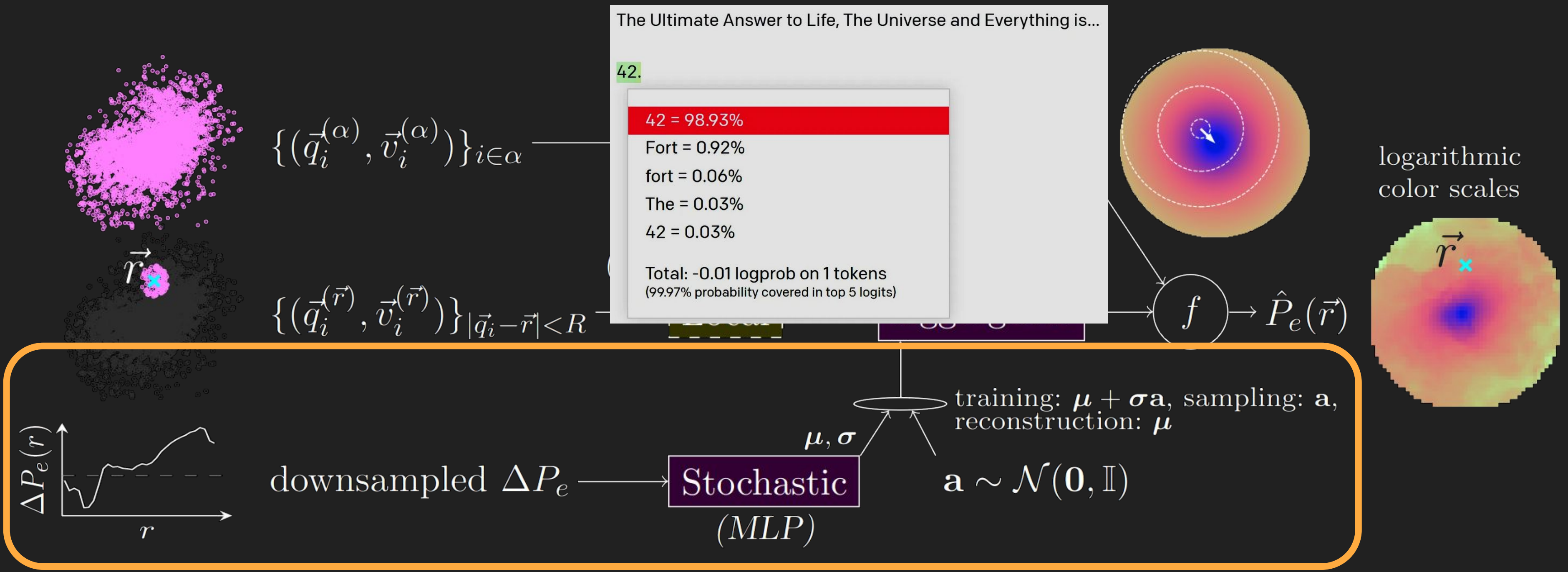
Set-based approach

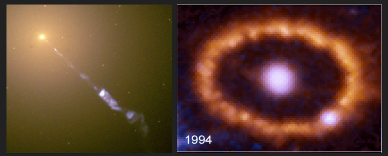




Probabilistic predictions

Set-based approach

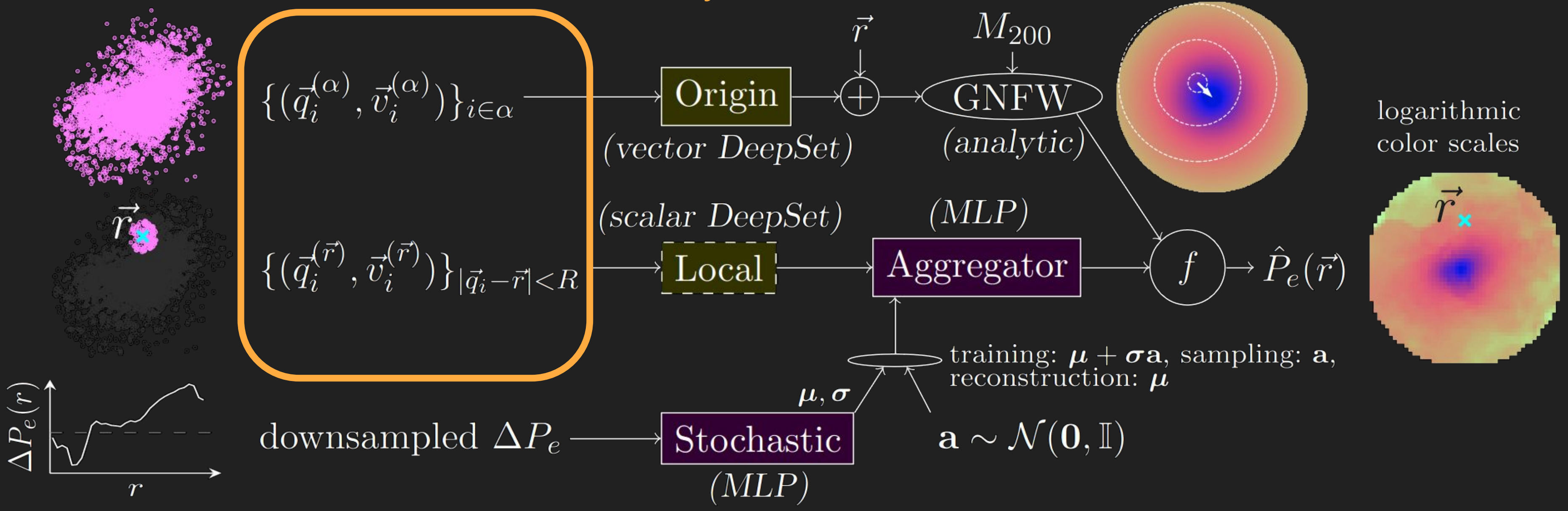




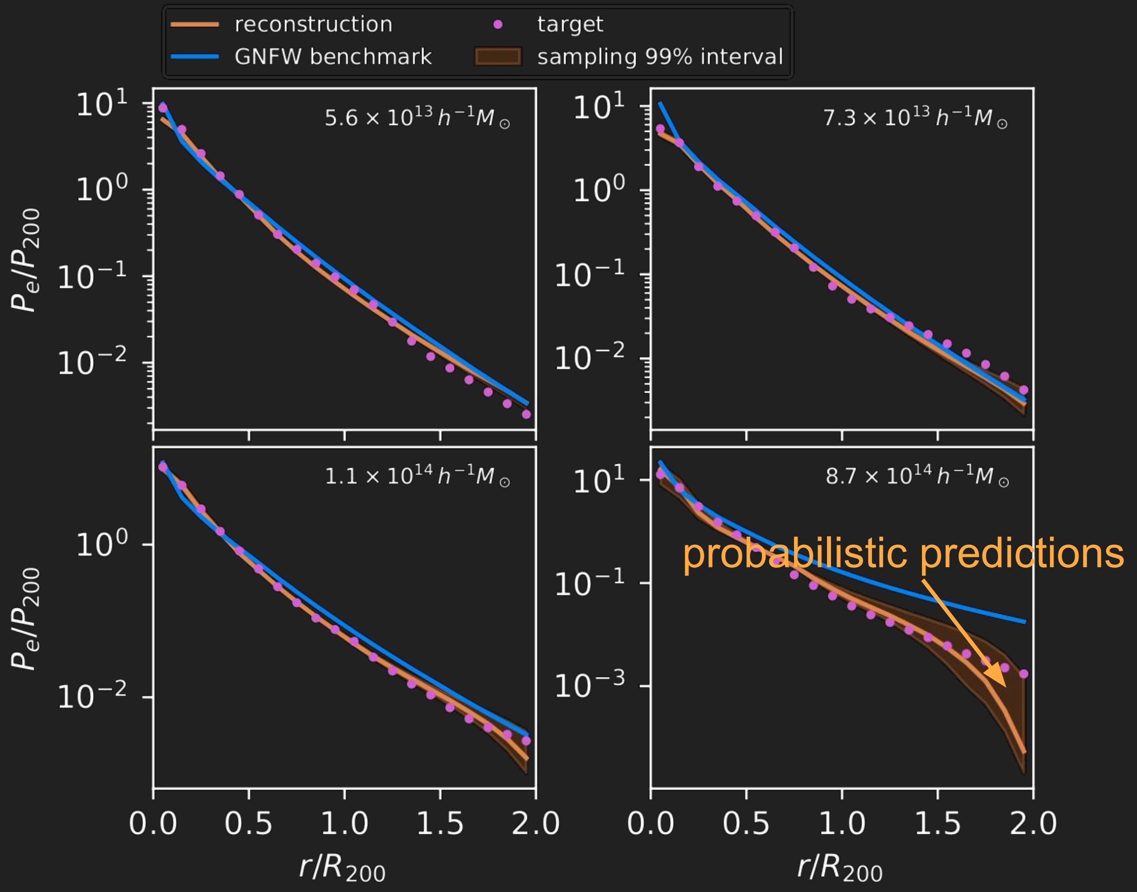
Probabilistic predictions

Set-based approach

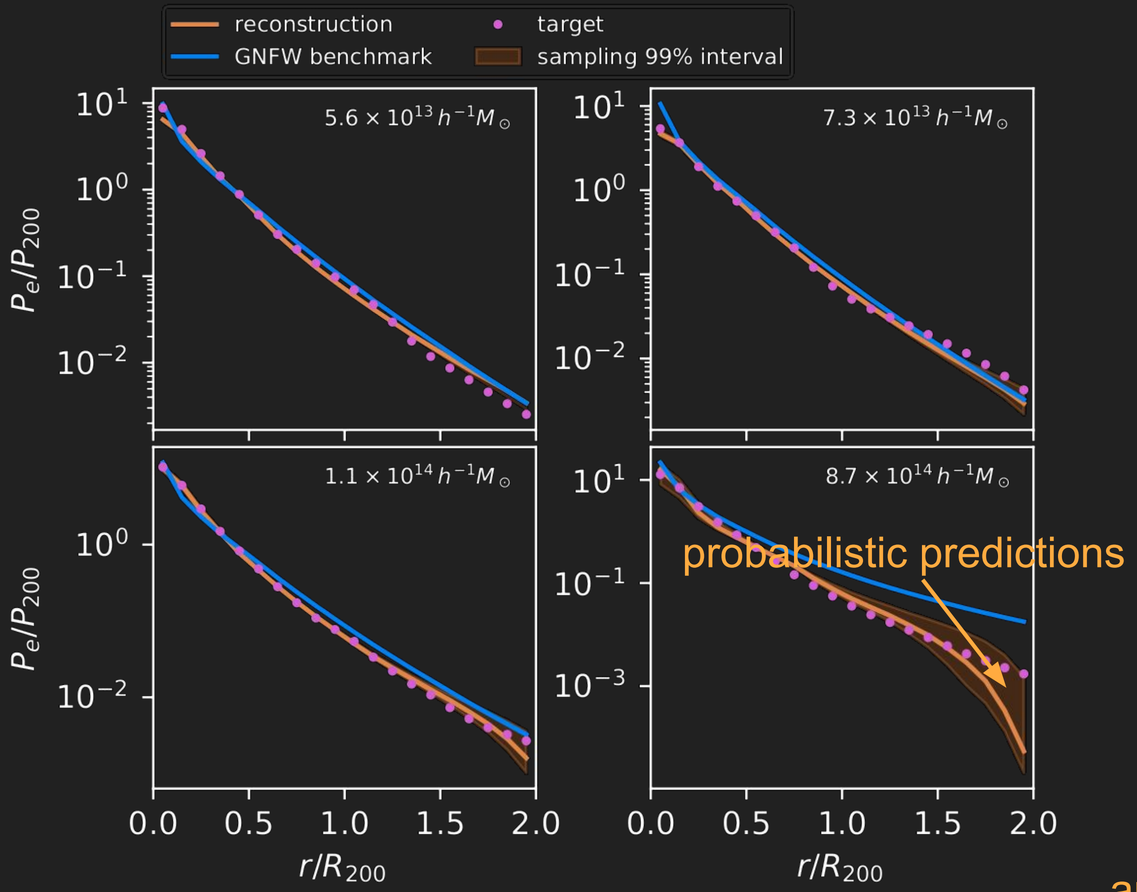
direct access to vectors → symmetries

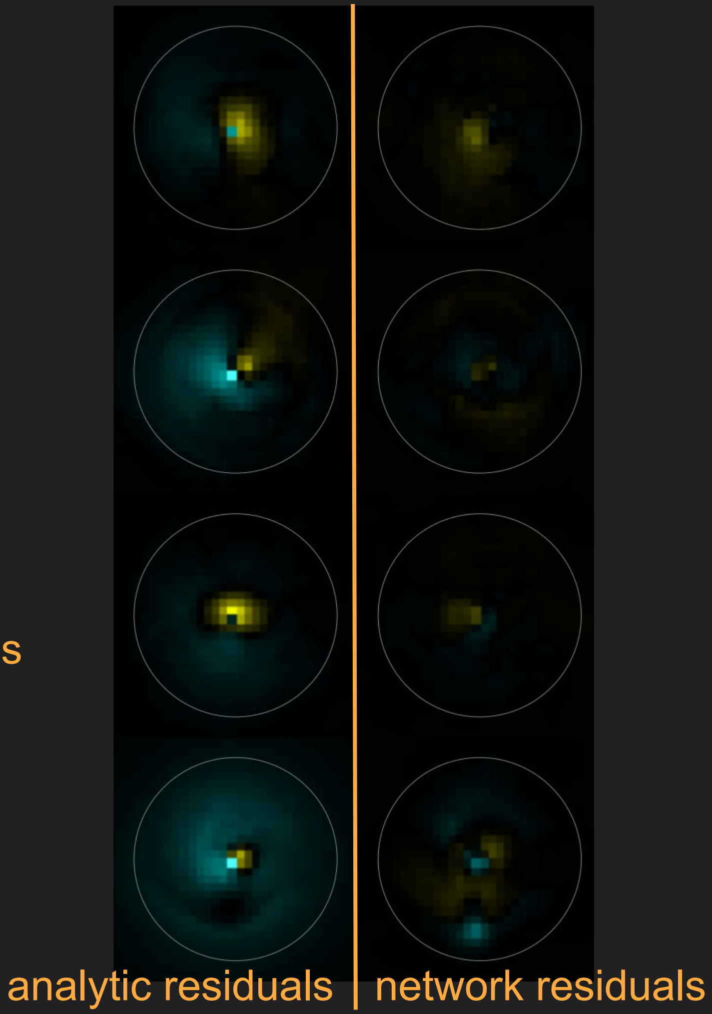


Model performance



Model performance





1) Understand the problem.

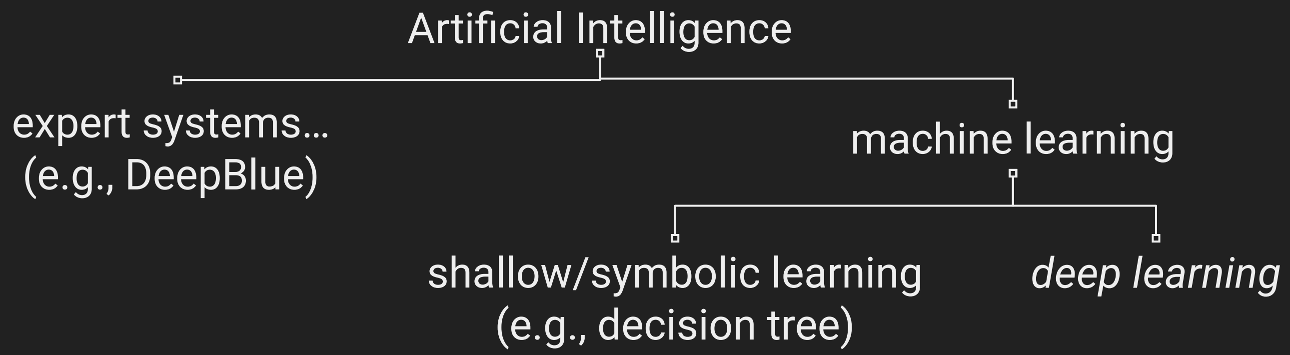
2) Have you tried linear regression / PCA / … ?

3) ML gives strictly what we ask it.

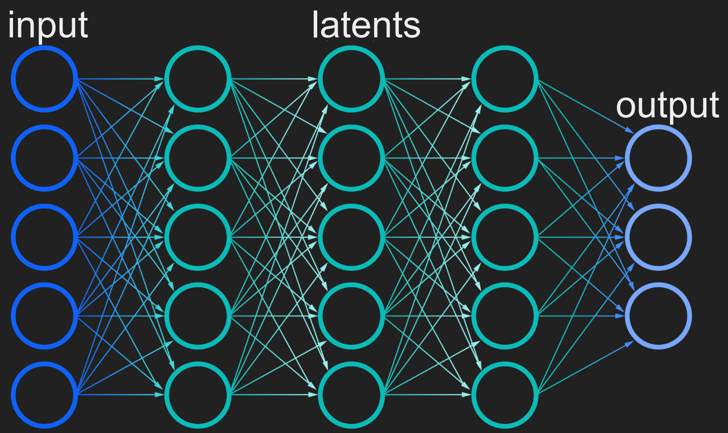
4) Decompose if possible.

5) First attempt rarely works. Rethink & rewrite.

The past decade in deep learning

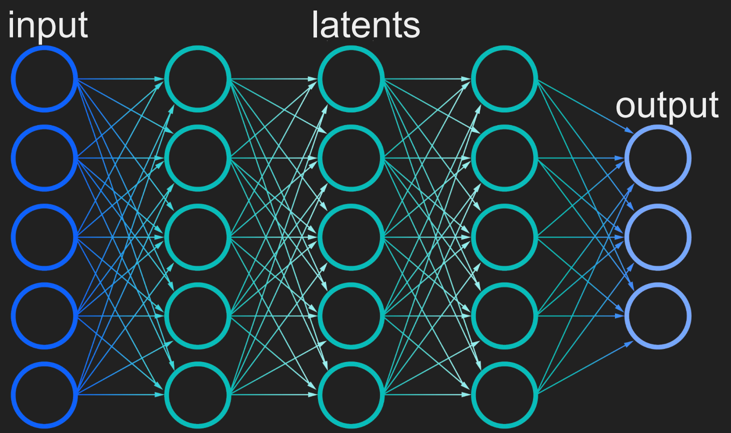

Artificial Intelligence
expert systems…
(e.g., DeepBlue)
machine learning
shallow/symbolic learning
(e.g., decision tree)
deep learning

Deep learning topography



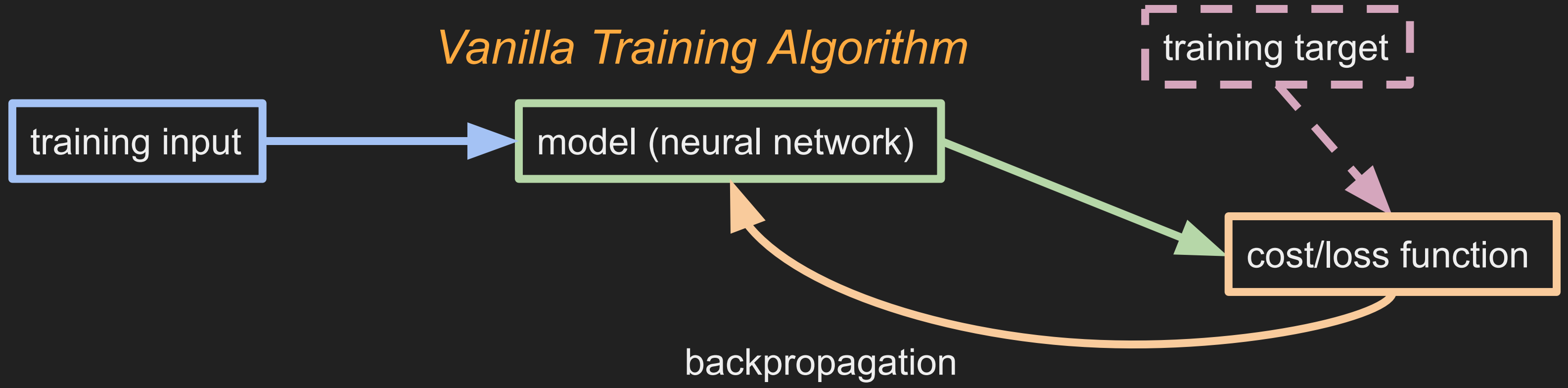
Universal Approximation Theorem:
Feedforward neural network can approximate “any” function given sufficient capacity.

Deep learning topography

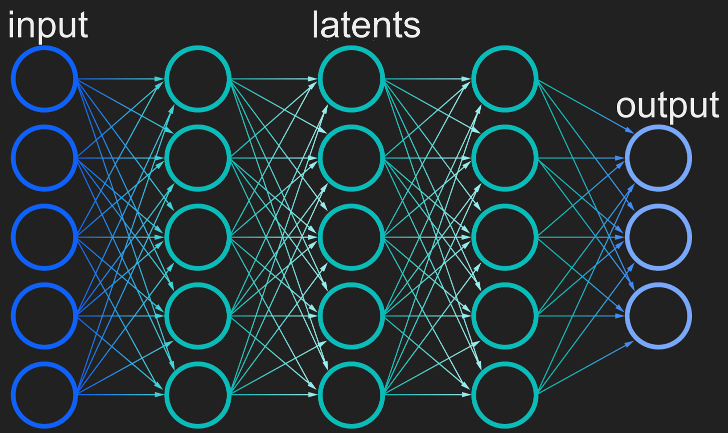


Universal Approximation Theorem:
Feedforward neural network can approximate “any” function given sufficient capacity.

Vanilla Training Algorithm



Deep learning topography

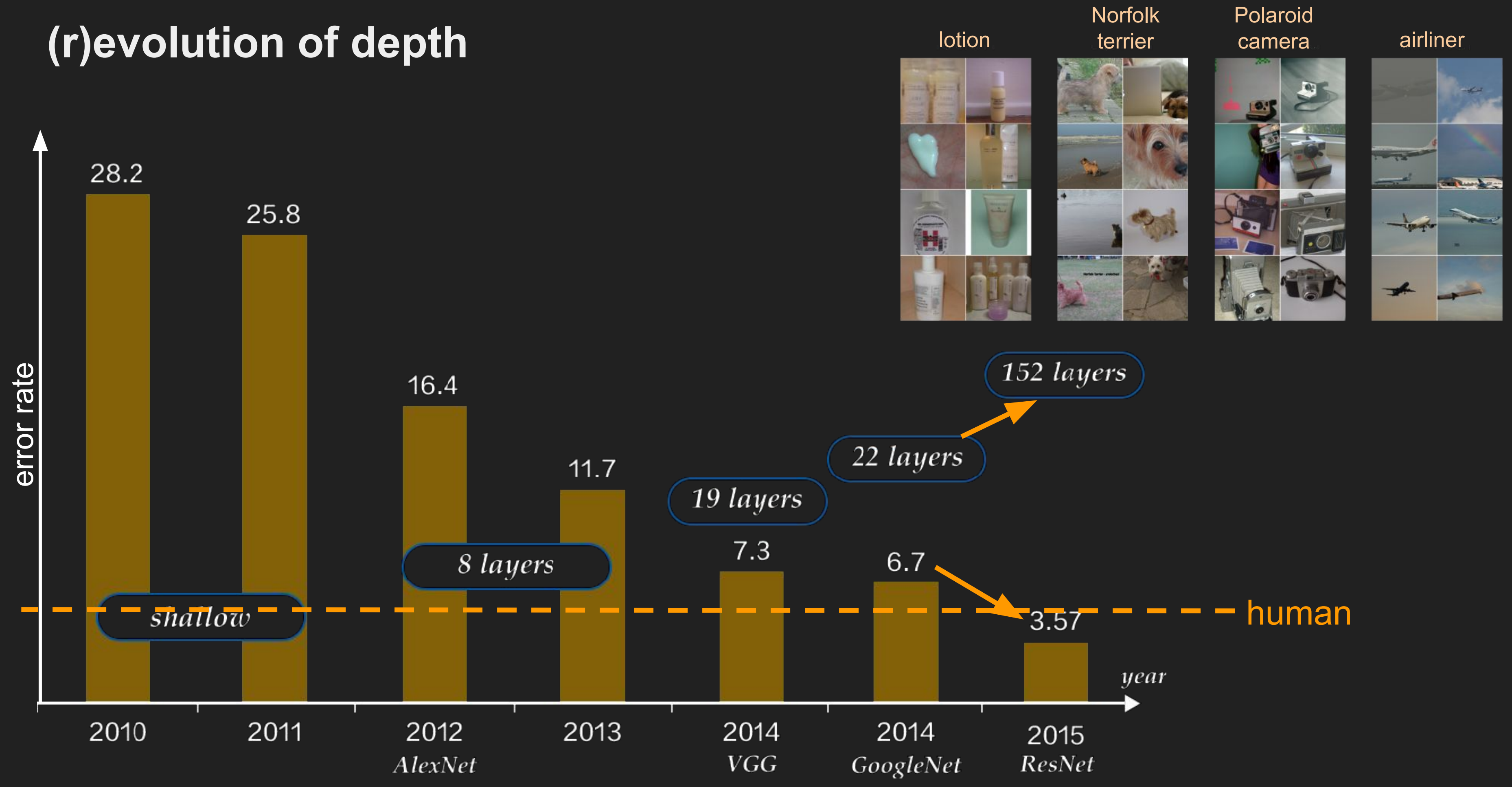


Universal Approximation Theorem:
Feedforward neural network can approximate "any" function given sufficient capacity.

"It is only slightly overstating the case to say that physics is the study of symmetry." [Anderson]

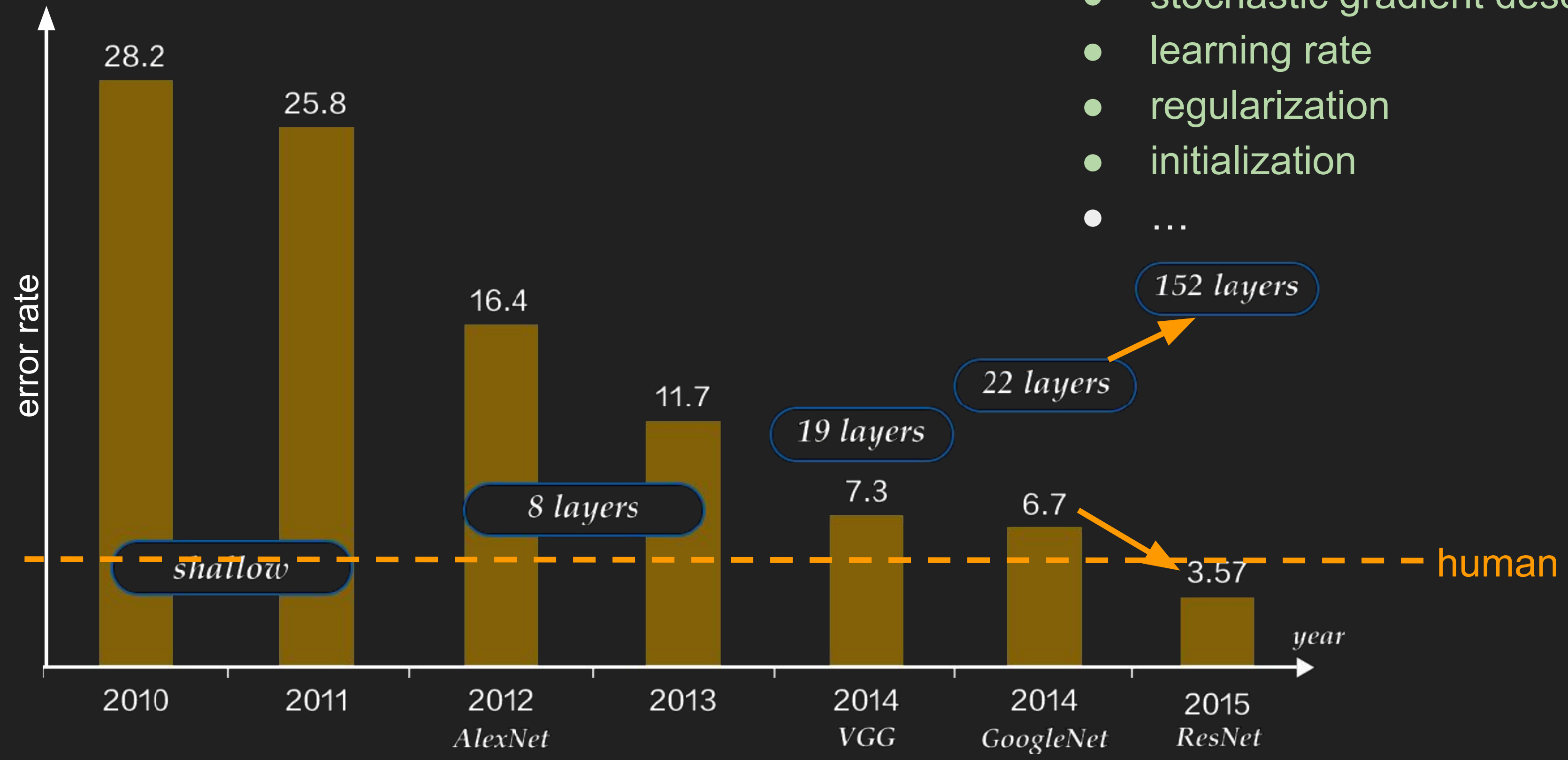
- translational sym: convolutional net
- time sym: recurrent net, transformer
- permutation sym: DeepSet, graph net
- …

(r)evolution of depth



(r)evolution of depth

- activation functions
- residual connections
- stochastic gradient descent
- learning rate
- regularization
- initialization
- …

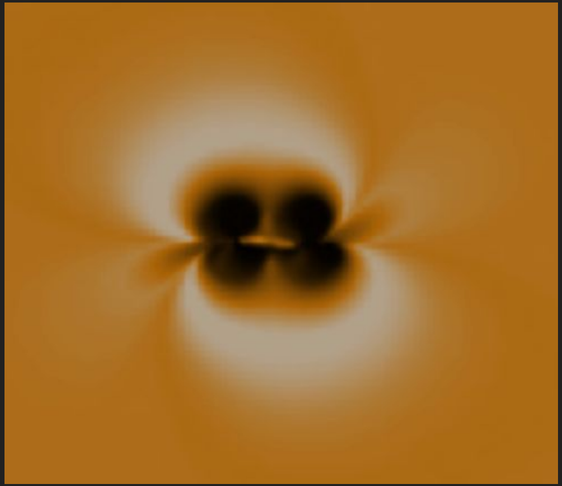


Complex problems take time

Brief (and incomplete) history of the binary black hole problem in numerical relativity

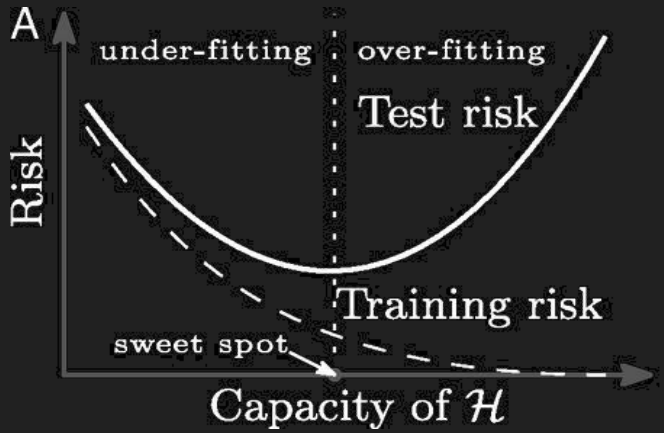
- L. Smarr, *PhD Thesis (1977)* : First head-on collision simulation
- P. Anninos, D. Hobill, E.Seidel, L. Smarr, W. Suen *PRL 71, 2851 (1993)* : Improved simulation of head-on collision
- B. Bruegmann *Int. J. Mod. Phys. D8, 85 (1999)* : First grazing collision of two black holes
- B. Bruegmann, W. Tichy, N. Jansen *PRL 92, 211101 (2004)* : First full orbit of a quasi-circular binary
- FP, *PRL 95, 121101 (2005)* : First "complete" simulation of a non head-on merger event: orbit, coalescence, ringdown and gravitational wave extraction
- M. Campanelli, C. O. Lousto, P. Marronetti, Y. Zlochower *(gr-qc/0511048)*; J. G. Baker, J. Centrella, D. Choi, M. Koppitz, J. van Meter *(gr-qc/0511103)* … several other groups have now repeated these results

- activation functions
- residual connections
- stochastic gradient descent
- learning rate
- regularization
- initialization
- …

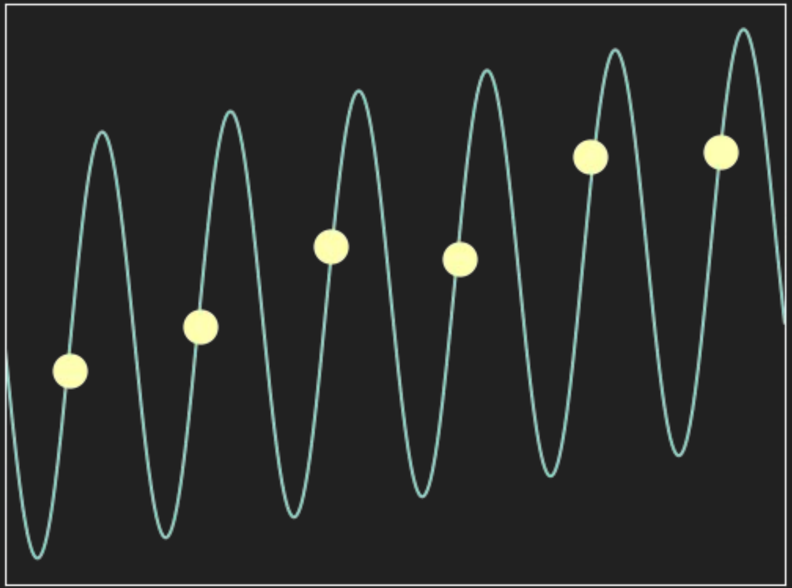


[F. Pretorius]

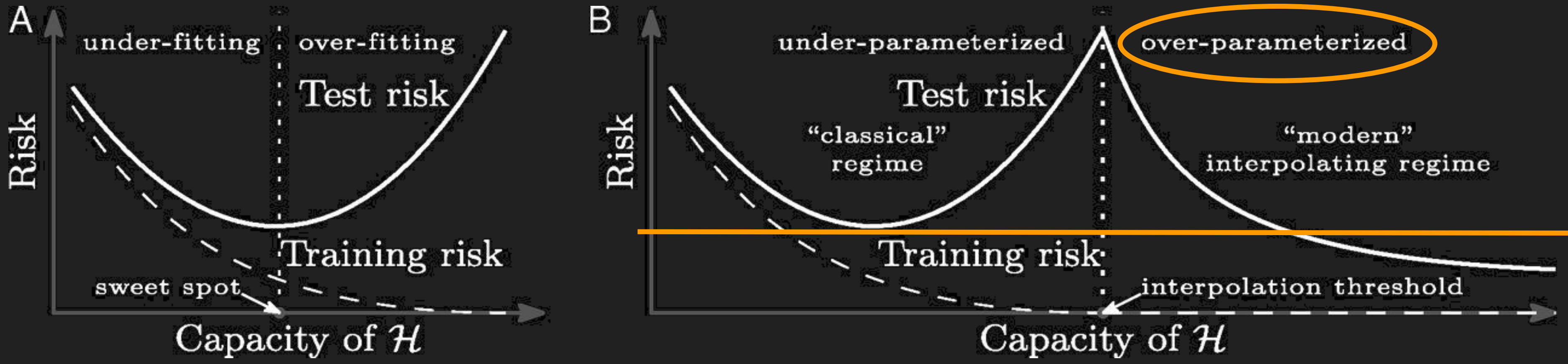
Large models overfit, right?



[Belkin+2019]

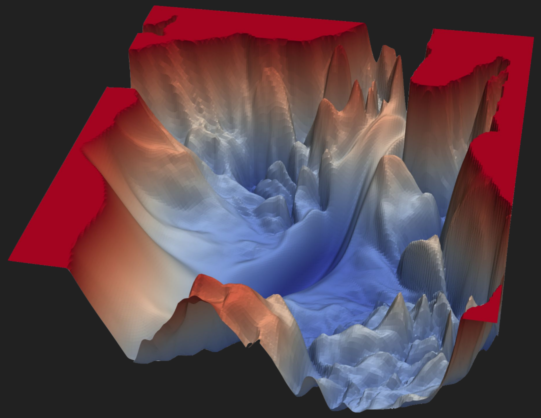


Over-parameterization miracle

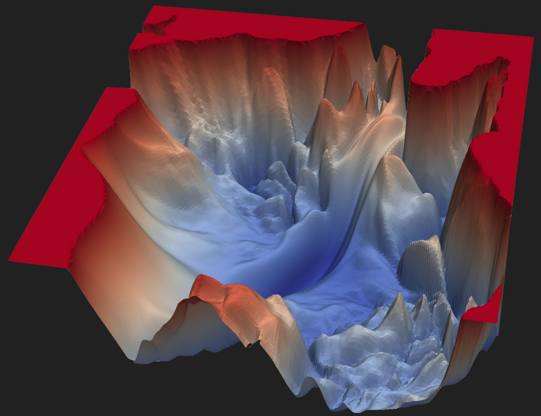


[Belkin+2019]

Why does it work?

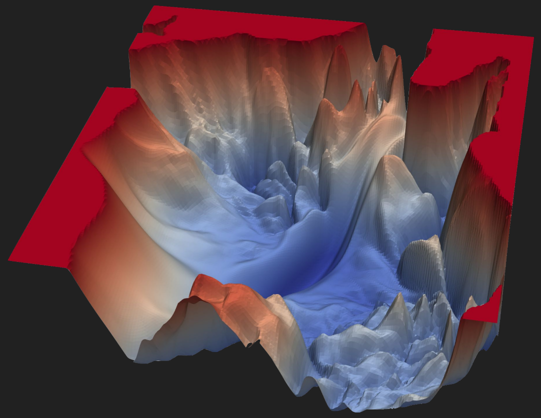


More Is Different



"The behavior of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity entirely new properties appear." [Anderson]

More Is Different

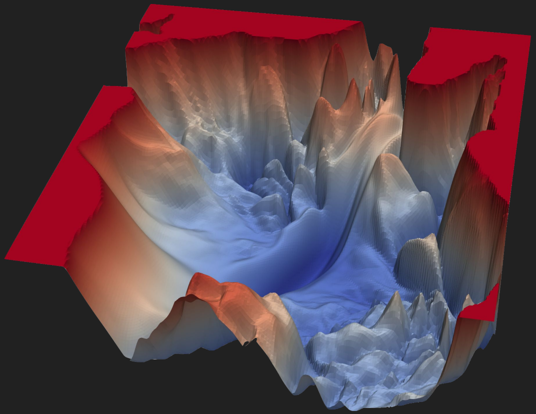


"The behavior of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity entirely new properties appear." [Anderson]

diversity

universality

More Is Different



Fundamental Physics, the Swampland of Effective Field Theory and Early Universe Cosmology

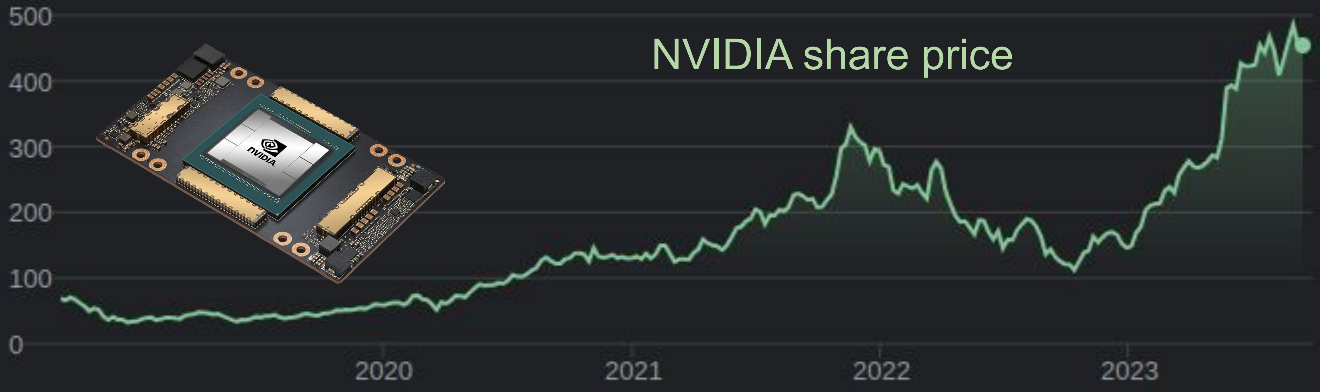
Robert Brandenberger

More Compute Is Different

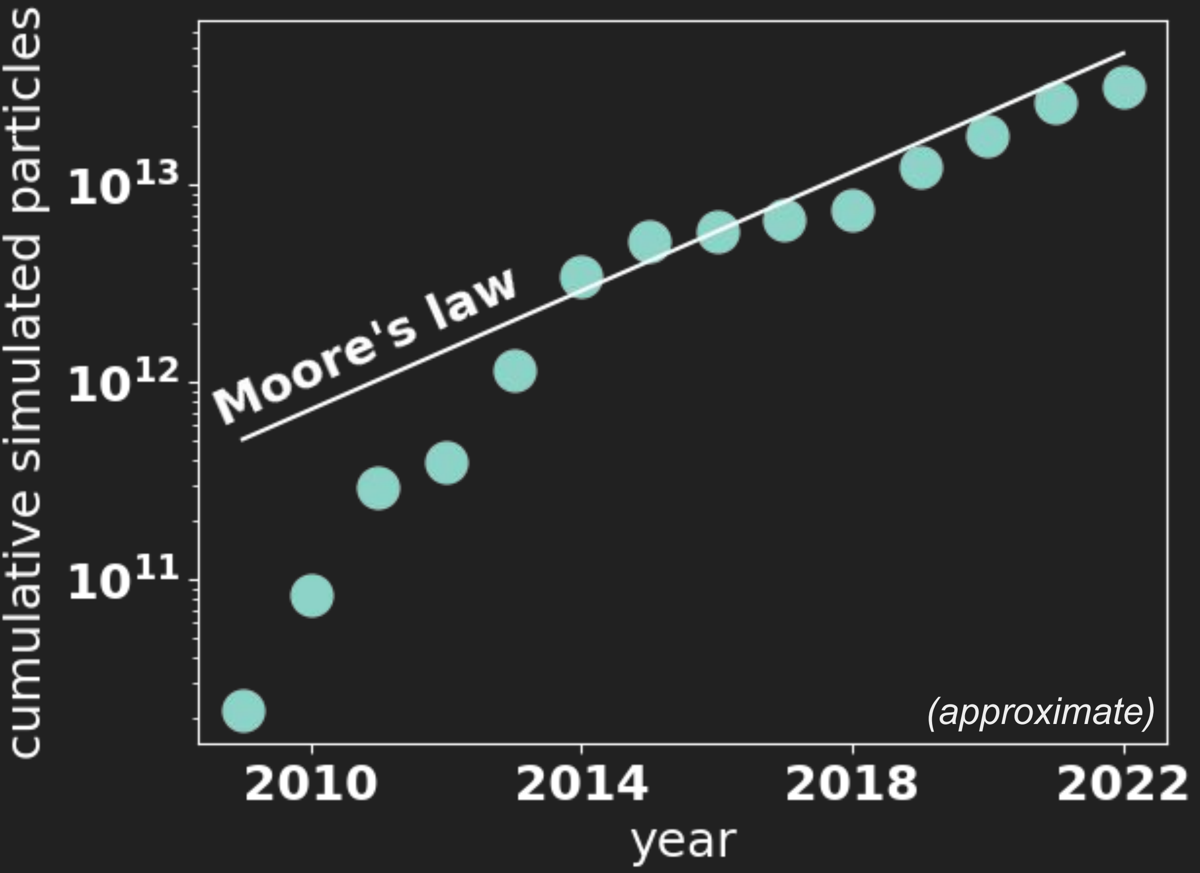
2012: graphical processing units (GPUs) are perfect for deep learning

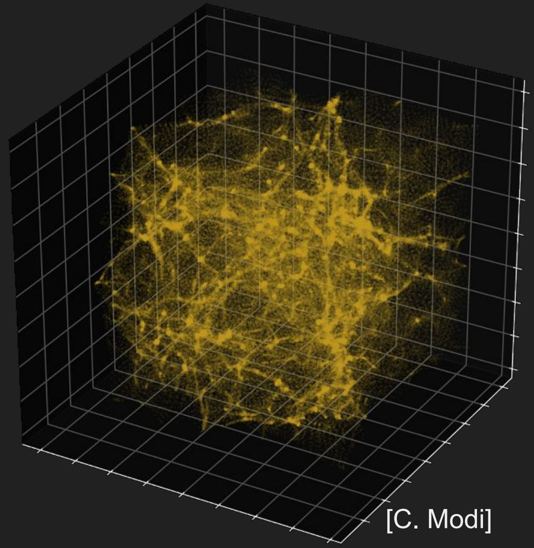
Krizhevsky+2012:

> In the end, the network's size is limited mainly by the amount of memory available on current GPUs and by the amount of training time that we are willing to tolerate. Our network takes between five and six days to train on two GTX 580 3GB GPUs. All of our experiments suggest that our results can be improved simply by waiting for faster GPUs and bigger datasets to become available.



Moore Is Different: cosmological simulations





[C. Modi]

Inference

Parameter estimation faced with intractable likelihoods

Bayesian inference

$$\underbrace{P(\text{parameters} \mid \text{data})}_{\text{[posterior]}} = \frac{\overbrace{P(\text{data} \mid \text{parameters})}^{\text{[likelihood]}}\ \overbrace{P(\text{parameters})}^{\text{[prior]}}}{\underbrace{P(\text{data})}_{\text{[evidence]}}}$$

Bayesian inference

$$\underbrace{P(\text{parameters} \mid \text{data})}_{\text{[posterior]}} = \frac{\overbrace{P(\text{data} \mid \text{parameters})}^{\text{[likelihood]}} \; \overbrace{P(\text{parameters})}^{\text{[prior]}}}{\underbrace{P(\text{data})}_{\text{[evidence]}}}$$

More concretely:

θ=interesting parameters, η=nuisance parameters, ζ=initial conditions,
x=data, m=model

$$P(x|\theta) = \int \mathcal{D}\eta \, \mathcal{D}\zeta \, \delta[\, x - m(\theta, \eta, \zeta) \,]$$

[observational errors modify integrand]

The need for Implicit Likelihood Inference

$$P(x|\theta) = \int \mathcal{D}\eta\, \mathcal{D}\zeta\, \delta[\, x - m(\theta, \eta, \zeta)\,]$$

In the traditional case:

$$P(x|\theta) = \int \mathcal{D}\eta\, \text{Gaussian}[\, x - \mu(\theta, \eta), \text{Covariance}\,]$$

[evaluate remaining low-dimensional η integral with Monte Carlo]

The need for Implicit Likelihood Inference

$$P(x|\theta) = \int \mathcal{D}\eta\, \mathcal{D}\zeta\, \delta[\, x - m(\theta, \eta, \zeta)\,]$$

In the traditional case:

$$P(x|\theta) = \int \mathcal{D}\eta\, \text{Gaussian}[\, x - \mu(\theta, \eta), \text{Covariance}\,]$$

But what do we do if:

1) integration over initial conditions ζ is impossible analytically,

and/or

2) $\mu(\theta, \eta)$ is difficult to approximate?

The need for Implicit Likelihood Inference

$$P(x|\theta) = \int \mathcal{D}\eta\, \mathcal{D}\zeta\, \delta[\, x - m(\theta, \eta, \zeta)\,]$$

In the traditional case:

$$P(x|\theta) = \int \mathcal{D}\eta\, \mathrm{Gaussian}[\, x - \mu(\theta, \eta), \mathrm{Covariance}\,]$$

But what do we do if:

1) integration over initial conditions ζ is impossible analytically,

and/or

2) $\mu(\theta, \eta)$ is difficult to approximate?

Assume that we have a simulator that can evaluate the model $m(\theta, \eta, \zeta)$ *to required accuracy*.

Implicit Likelihood Inference

Assume that we have a simulator that can evaluate the model m(θ, η, ζ) *to required accuracy*.

Use simulated samples to train neural approximator:

neural likelihood estimation

neural posterior estimation

$$P(\text{parameters} \mid \text{data}) = \frac{\boxed{P(\text{data} \mid \text{parameters})}\; P(\text{parameters})}{P(\text{data})}$$

neural ratio estimation

Neural ratio estimation (NRE)

$$P(\text{parameters} \mid \text{data}) = \frac{\boxed{\frac{P(\text{data} \mid \text{parameters})}{P(\text{data})}}\, P(\text{parameters})}{}$$

- draw θ, θ' ~ P(parameters)
- simulate x_{sim} ~ P(data | θ) [drawing η, ζ]
- evaluate neural net: $y = f(x_{sim}, \theta)$; $y' = f(x_{sim}, \theta')$
- classification loss, e.g., $\mathscr{L} = -\log(y) - \log(1-y')$
- $f^* = \text{argmin} \int D\theta\, D\theta'\, Dx_{sim}\, \mathscr{L}$
- $\rightarrow P(x \mid \theta) / P(x) = f^*(x, \theta) / [\, 1 - f^*(x, \theta)\,]$

Neural ratio estimation (NRE)

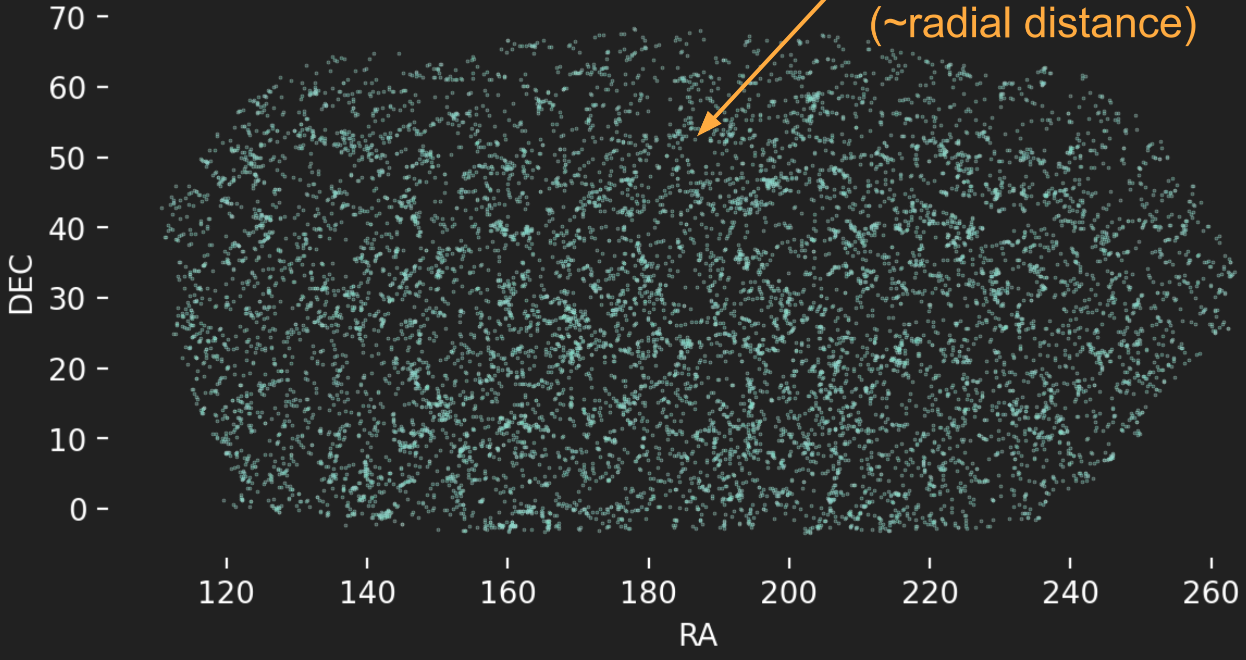
$$P(\text{parameters} \mid \text{data}) = \frac{\boxed{\dfrac{P(\text{data} \mid \text{parameters})}{P(\text{data})}}\; P(\text{parameters})}{}$$

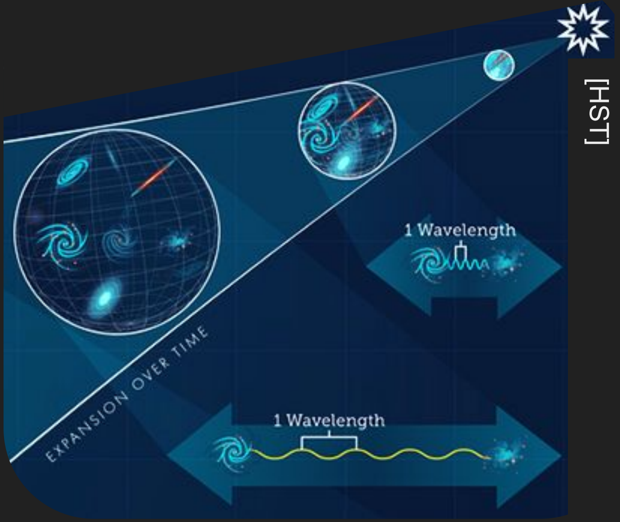
- draw $\theta, \theta' \sim P(\text{parameters})$
- simulate $x_{sim} \sim P(\text{data} \mid \theta)$ [drawing η, ζ]
- evaluate neural net: $y = f(x_{sim}, \theta)$; $y' = f(x_{sim}, \theta')$
- classification loss, e.g., $\mathscr{L} = -\log(y) - \log(1-y')$
- $f^* = \text{argmin} \int D\theta\, D\theta'\, Dx_{sim}\, \mathscr{L}$
- $\rightarrow P(x \mid \theta) / P(x) = f^*(x, \theta) / [\, 1 - f^*(x, \theta)\,]$

- any simulatable effect can be incorporated
- rephrase as classification problem → sophisticated machinery exists
- no formal difference between nuisance parameters and initial conditions

A 3-D map of the universe

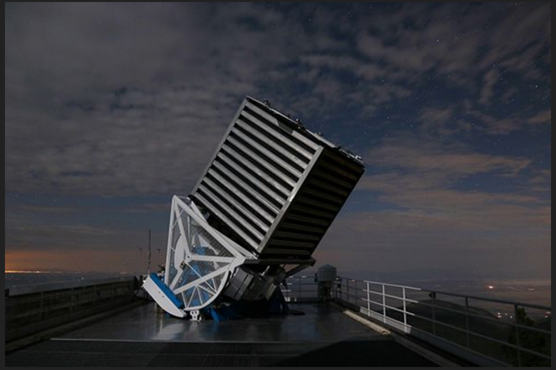
each dot a galaxy
with redshift
(~radial distance)





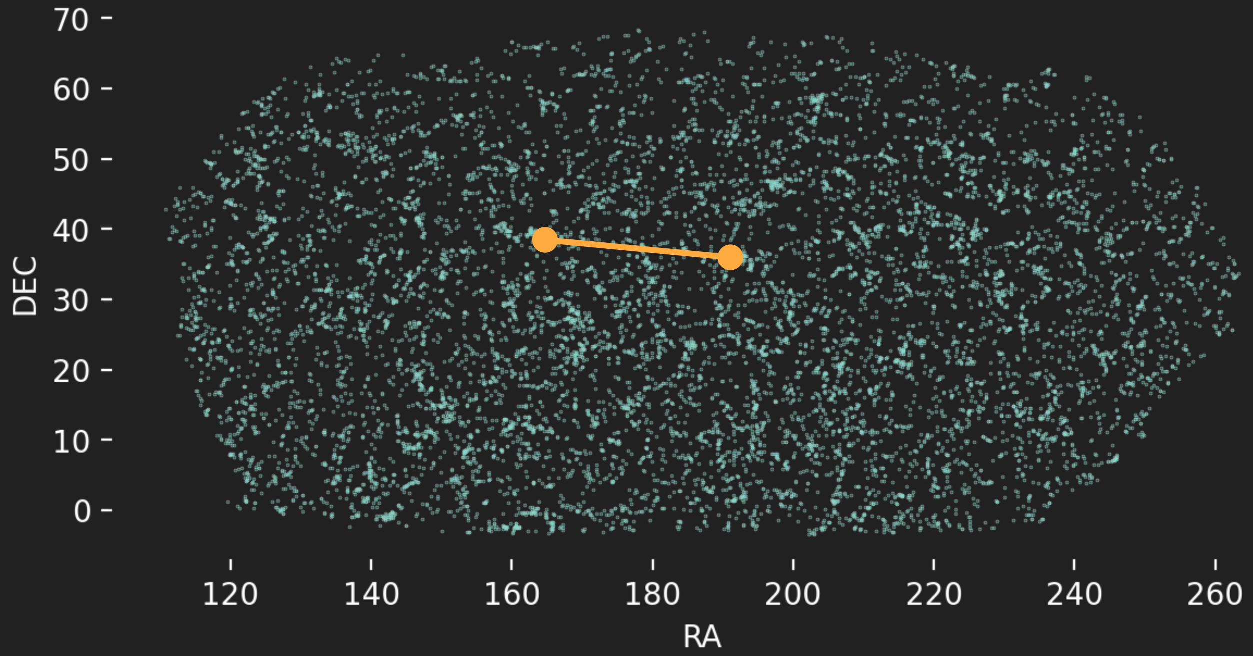
[HST]

SDSS/BOSS survey

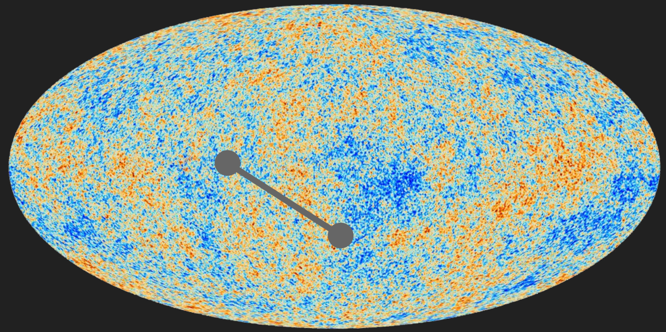


How to summarize this map?

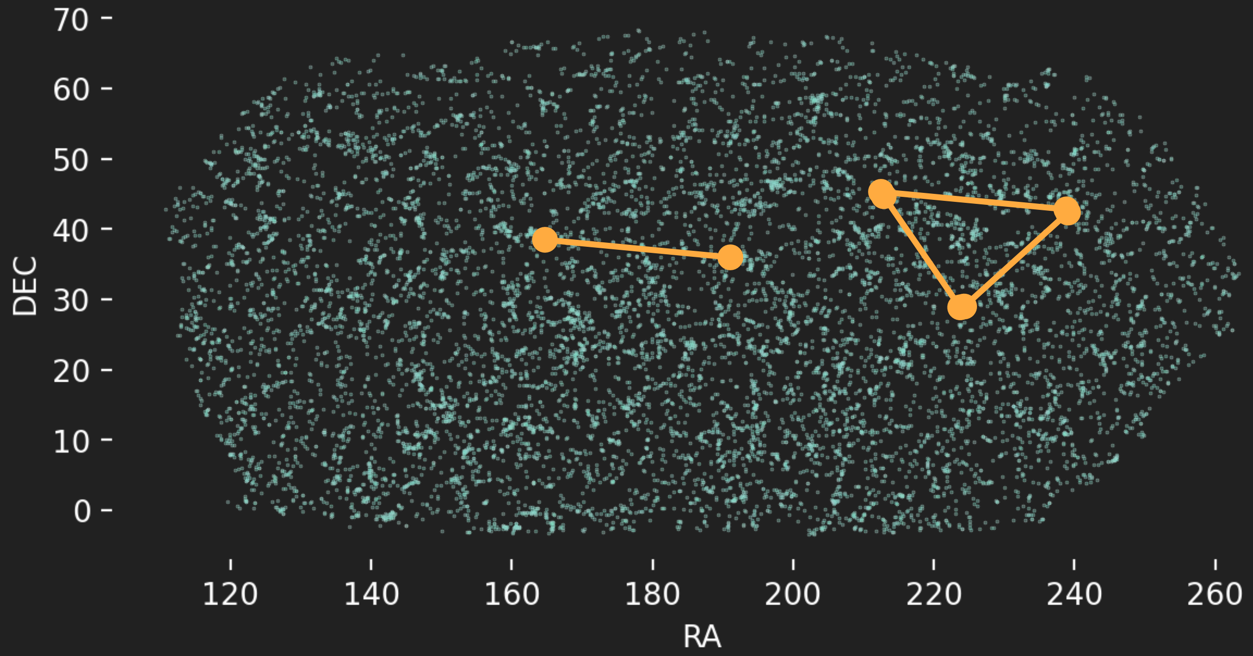
1) pairs of galaxies
   (power spectrum)



How to summarize this map?



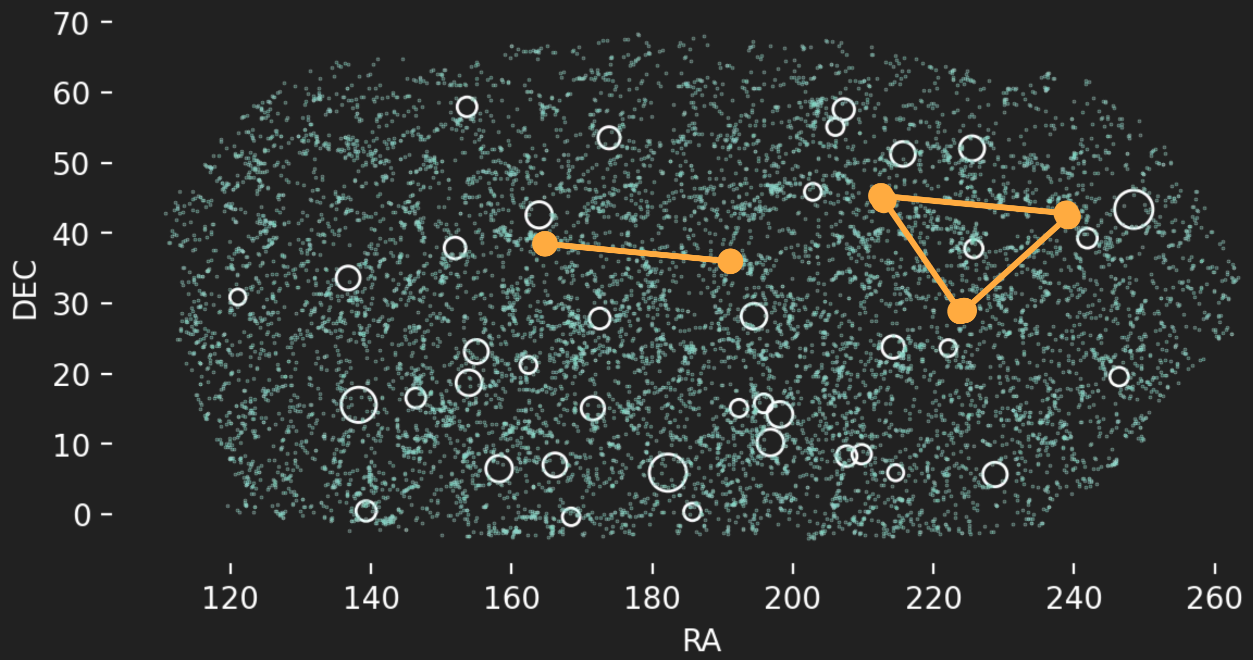
1) pairs of galaxies
   (power spectrum)
2) triangles of galaxies
   (bispectrum)
3) …



How to summarize this map?

1) pairs of galaxies (power spectrum)
2) triangles of galaxies (bispectrum)
3) …

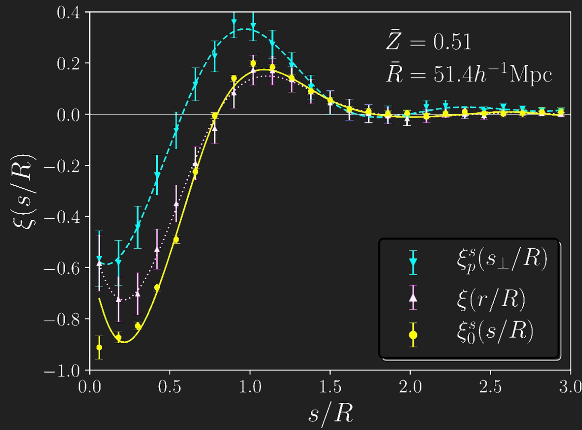
4) "empty regions" (*cosmic voids*)



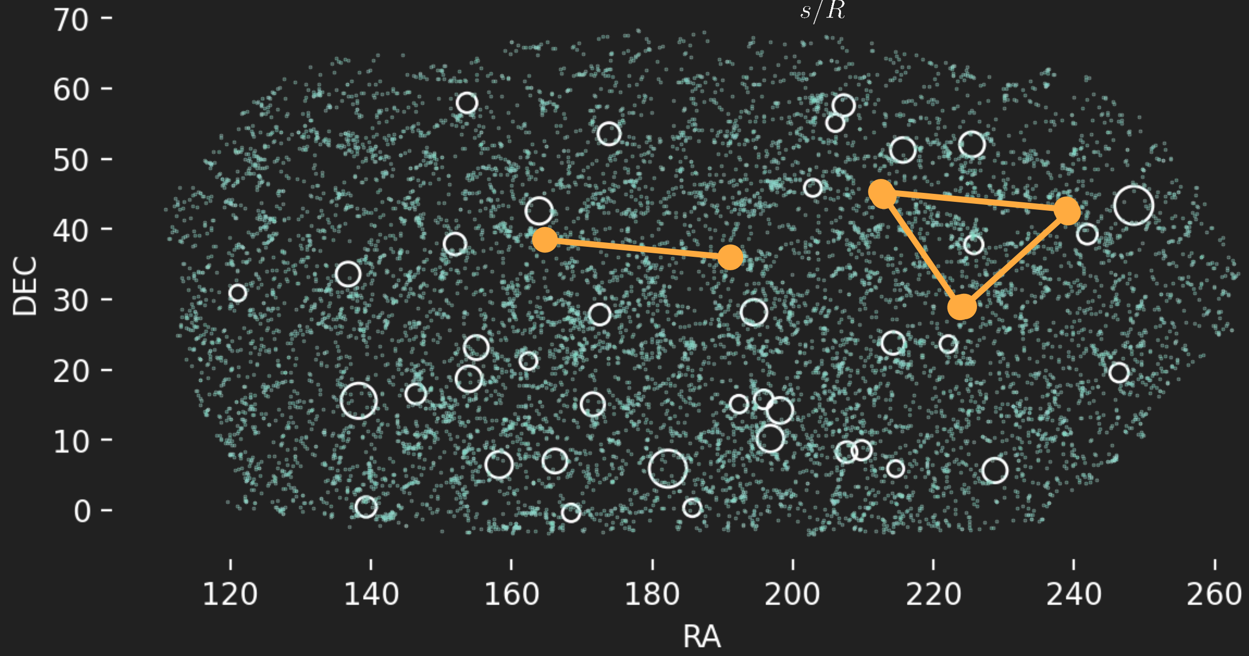
How to summarize this map?

1) pairs of galaxies (power spectrum)
2) triangles of galaxies (bispectrum)
3) …

4) “empty regions” (*cosmic voids*)



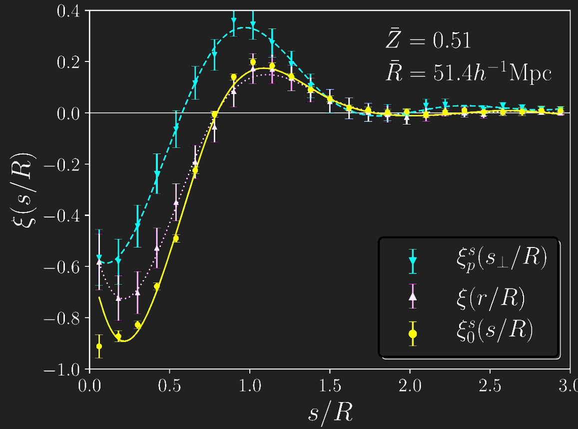
[Hamaus+2020]



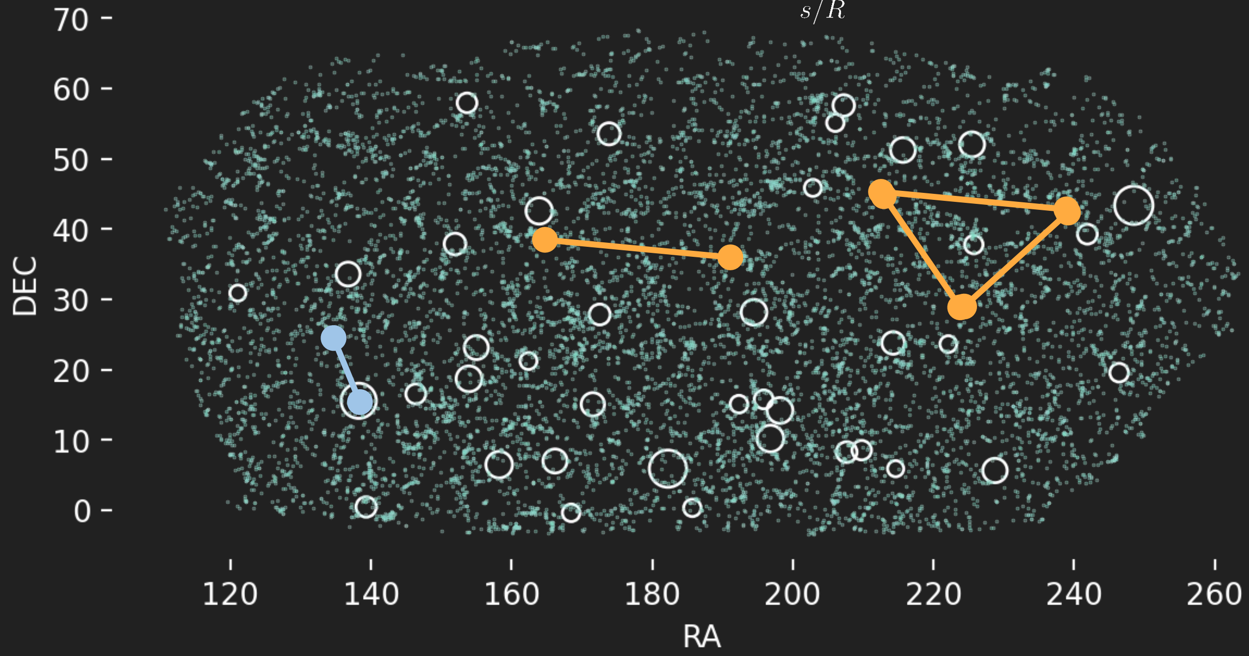
How to summarize this map?

1) pairs of galaxies
   (power spectrum)
2) triangles of galaxies
   (bispectrum)
3) …

4) "empty regions"
   (*cosmic voids*)
   – size distribution
   – void-galaxy pairs
   – …



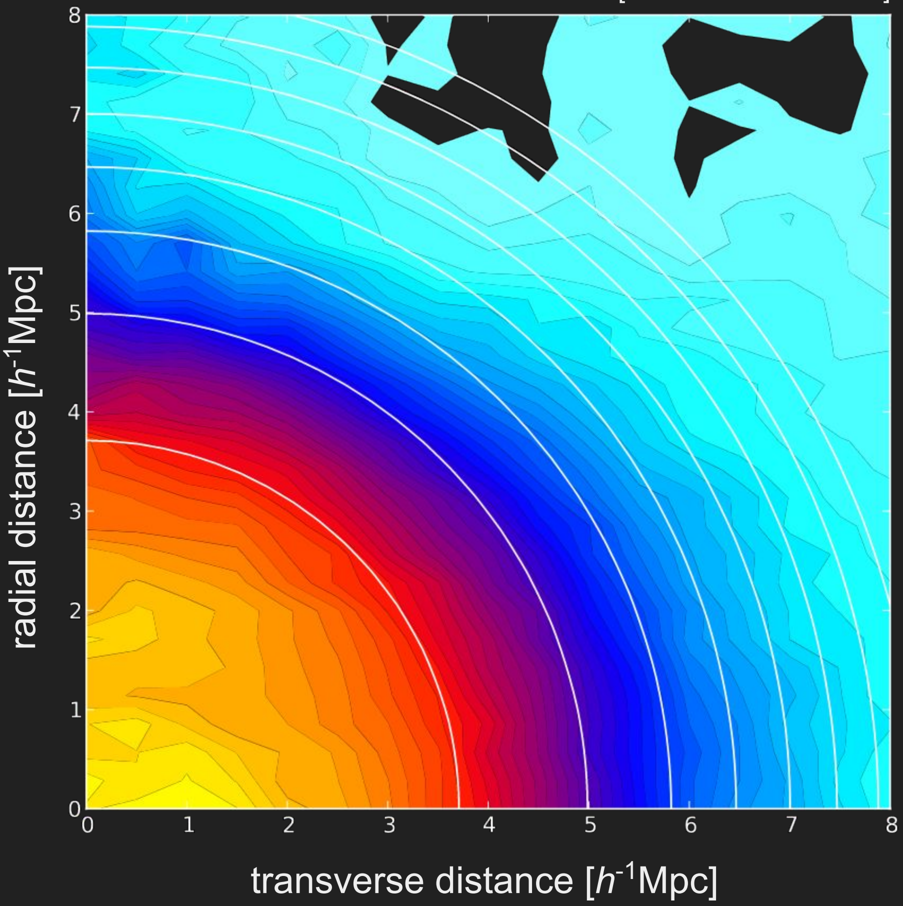
[Hamaus+2020]



What can voids do for us?

- not virialized → ideal for redshift-space distortions, Alcock-Paczynski
  - matter density Ω_m
  - growth of structure *f/b*

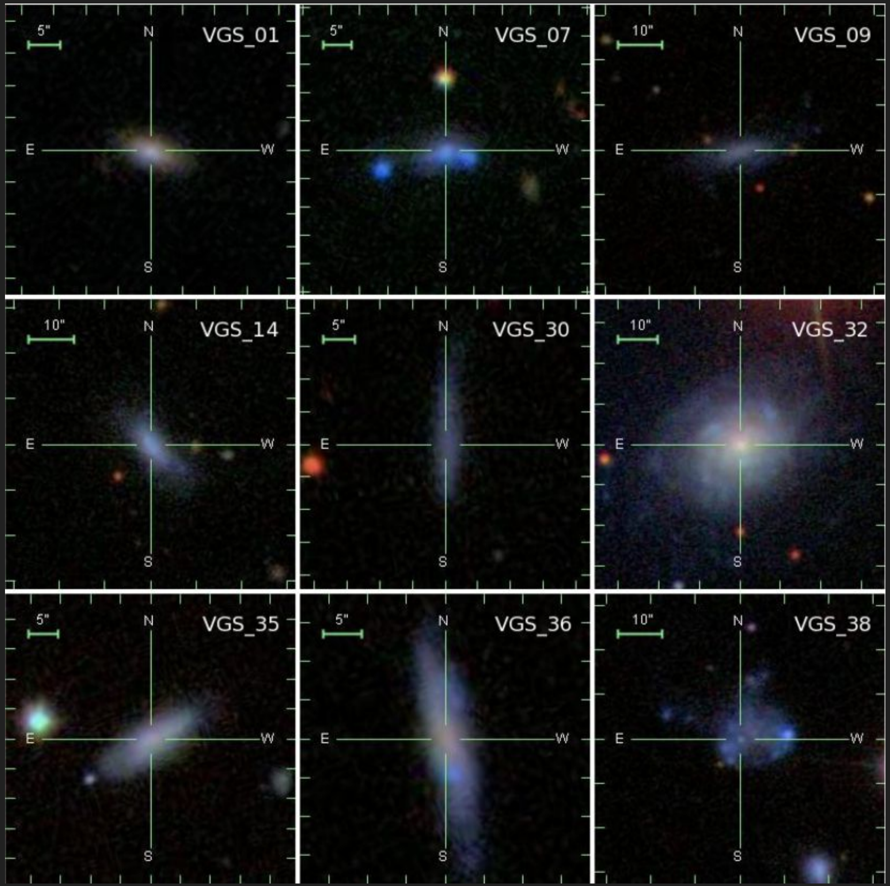
[Lavaux&Wandelt2011]



What can voids do for us?

[Kreckel+2011]

- not virialized → ideal for redshift-space distortions, Alcock-Paczynski
  - matter density Ω_m
  - growth of structure *f/b*
- isolated galaxies → galaxy evolution in simple environment

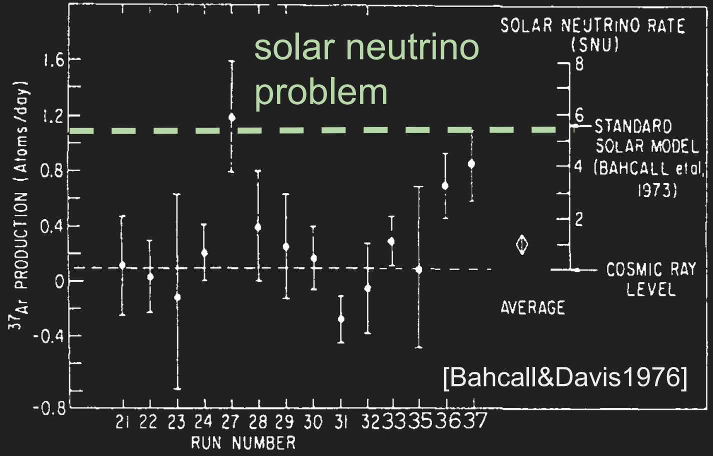


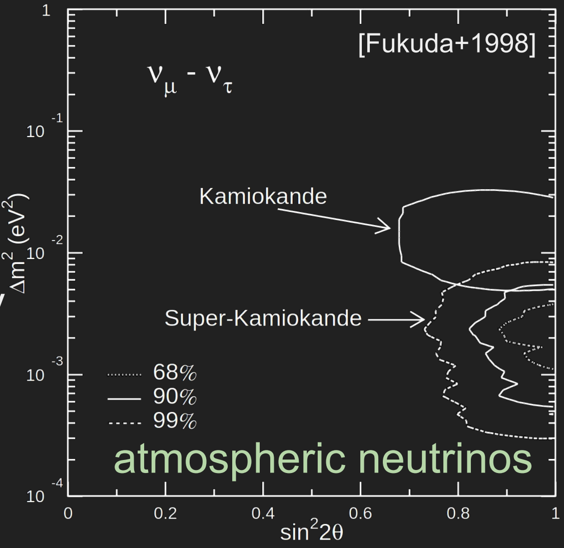
What can voids do for us?

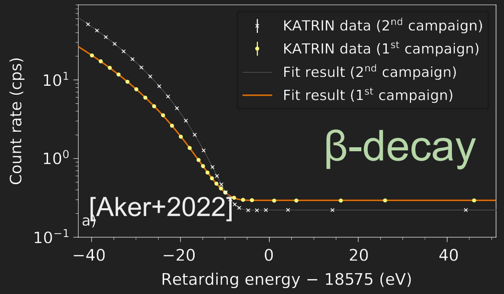
- not virialized → ideal for redshift-space distortions, Alcock-Paczynski
    - matter density Ω_m
    - growth of structure *f/b*
- isolated galaxies → galaxy evolution in simple environment
- complementary to correlation functions since voids upweight *underdense* regions
    - corrections to general relativity
    - dark energy
    - neutrino mass

What can voids do for us?

- not virialized → ideal for redshift-space distortions, Alcock-Paczynski
  - matter density Ω_m
  - growth of structure *f/b*
- isolated galaxies → galaxy evolution in simple environment
- complementary to correlation functions since voids upweight *underdense* regions
  - corrections to general relativity
  - dark energy
  - neutrino mass

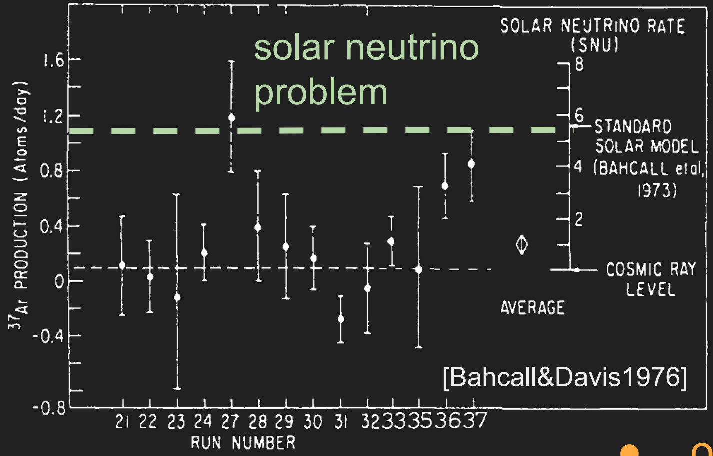


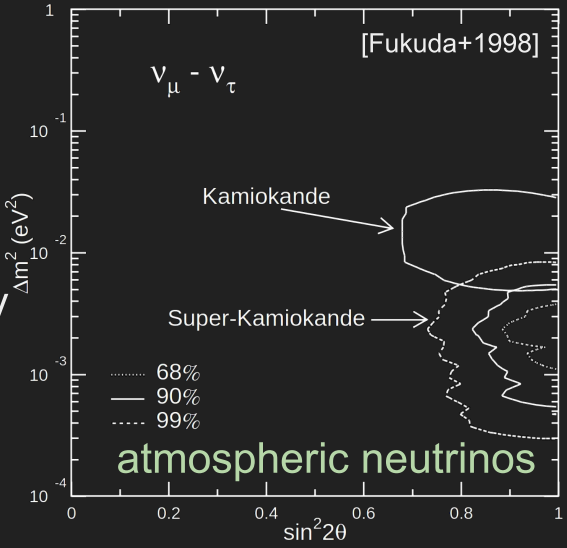




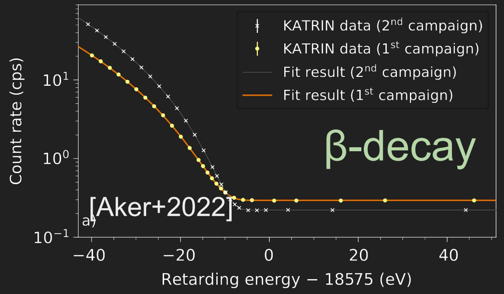
What can voids do for us?

- not virialized → ideal for redshift-space distortions, Alcock-Paczynski
  - matter density Ω_m
  - growth of structure *f/b*
- isolated galaxies → galaxy evolution in simple environment
- complementary to correlation functions since voids upweight *underdense* regions
  - corrections to general relativity
  - dark energy
  - neutrino mass





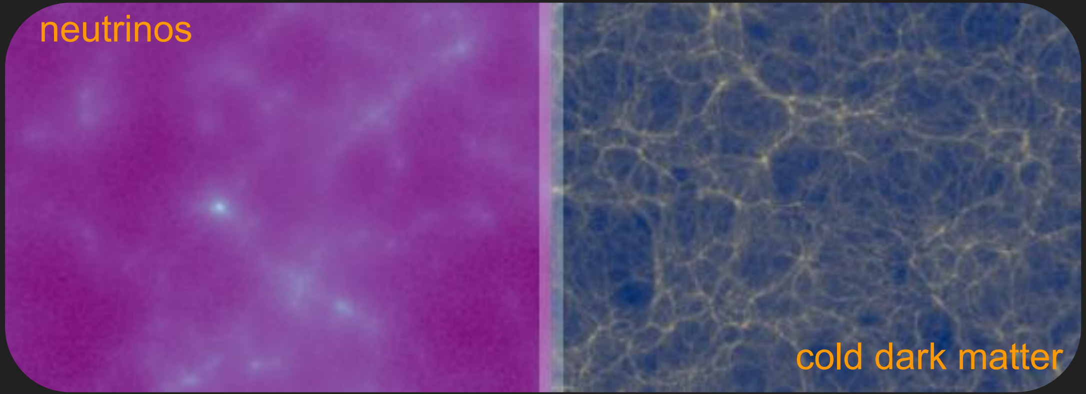
- origin of mass (seesaw?)
- neutrino nature?
- dark energy?
- surprises?

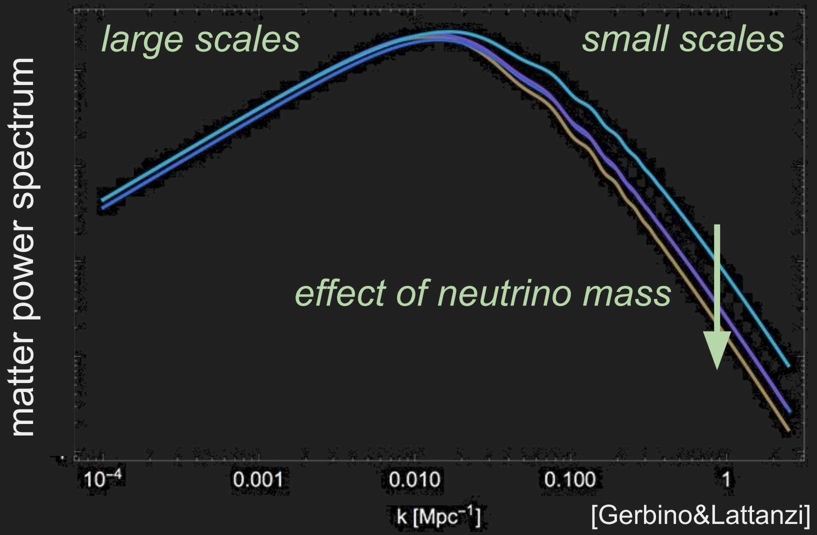


What can voids do for us?

- not virialized → ideal for redshift-space distortions, Alcock-Paczynski
  - matter density Ω_m
  - growth of structure *f/b*
- isolated galaxies → galaxy evolution in simple environment
- complementary to correlation functions since voids upweight *underdense* regions
  - corrections to general relativity
  - dark energy
  - neutrino mass

[FLAMINGO]

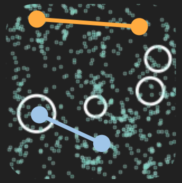




[Gerbino&Lattanzi]

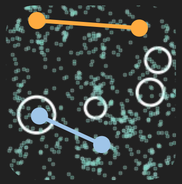
Problem formulation

- obtain constraints on neutrino mass sum, Σm_ν
- using 3-D galaxy map from BOSS survey
- summary statistics:
  - galaxy auto-power spectrum
  - void size function
  - void-galaxy cross power spectrum



Problem formulation

- obtain constraints on neutrino mass sum, Σm_ν
- using 3-D galaxy map from BOSS survey
- summary statistics:
  - galaxy auto-power spectrum
  - void size function
  - void-galaxy cross power spectrum



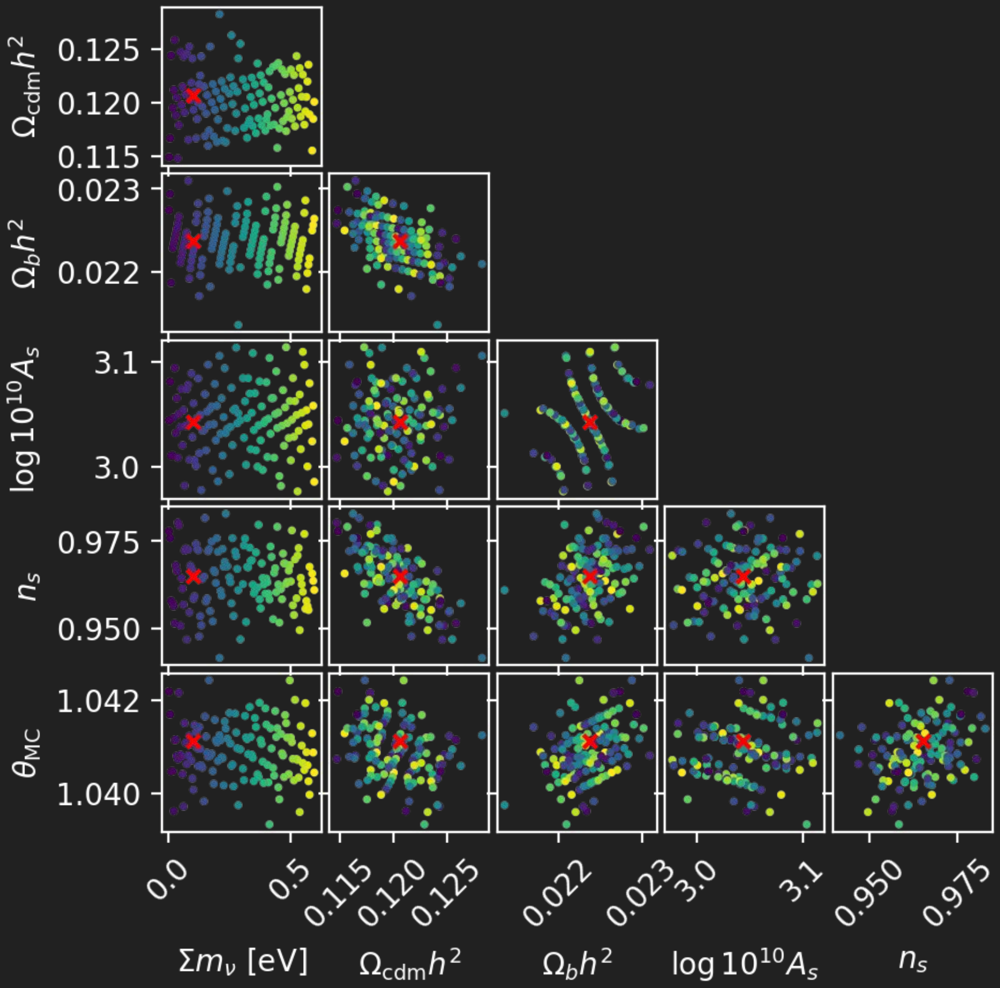
- model based on simulations
- perform implicit-likelihood inference

Simulations

Recall: sum over simulation samples needs to approximate

$$\int D\theta\ D\theta'\ Dx_{sim}$$

→ simulate according to prior $P(\theta)\ P(\eta)\ P(\zeta)$



Simulations

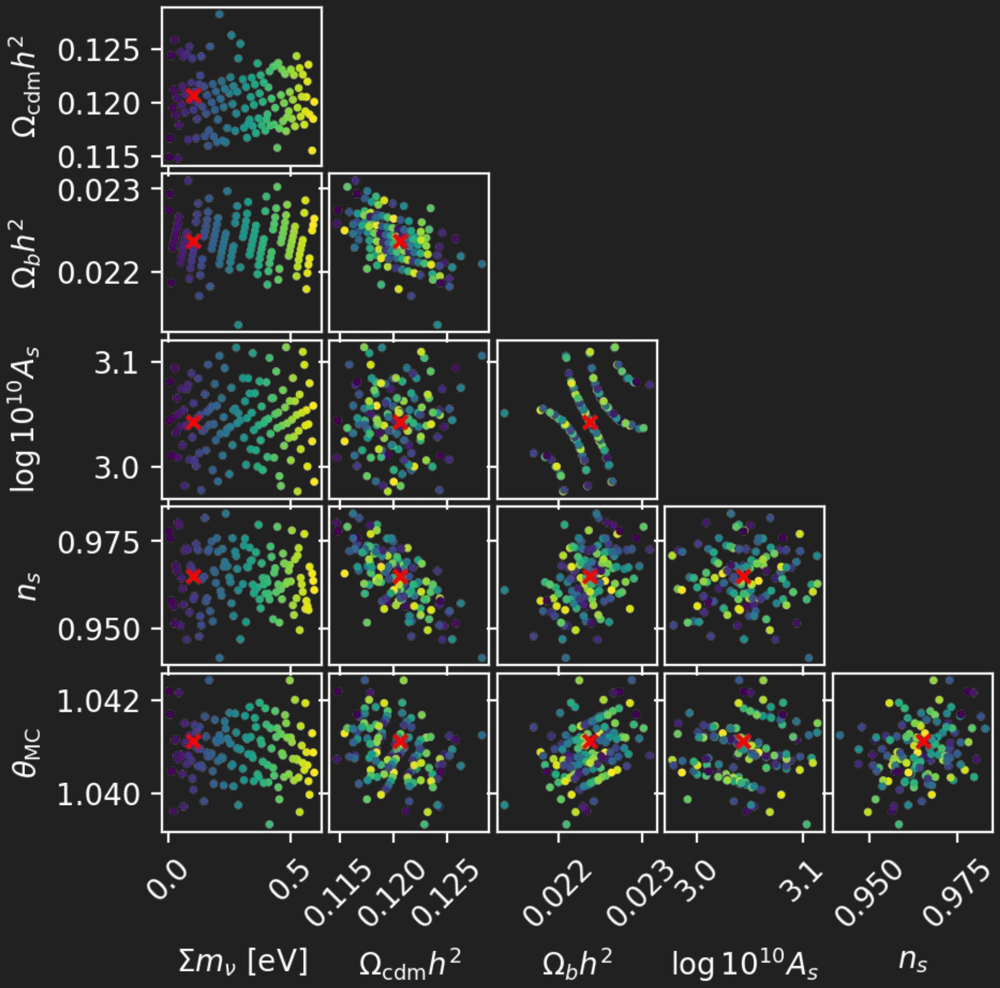
Recall: sum over simulation samples needs to approximate

$$\int D\theta \, D\theta' \, Dx_{\text{sim}}$$

→ simulate according to prior $P(\theta)\,P(\eta)\,P(\zeta)$

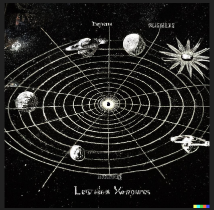
Place galaxies into gravity-only simulations using halo-occupation distribution model.

Implement survey realism.

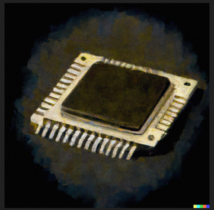


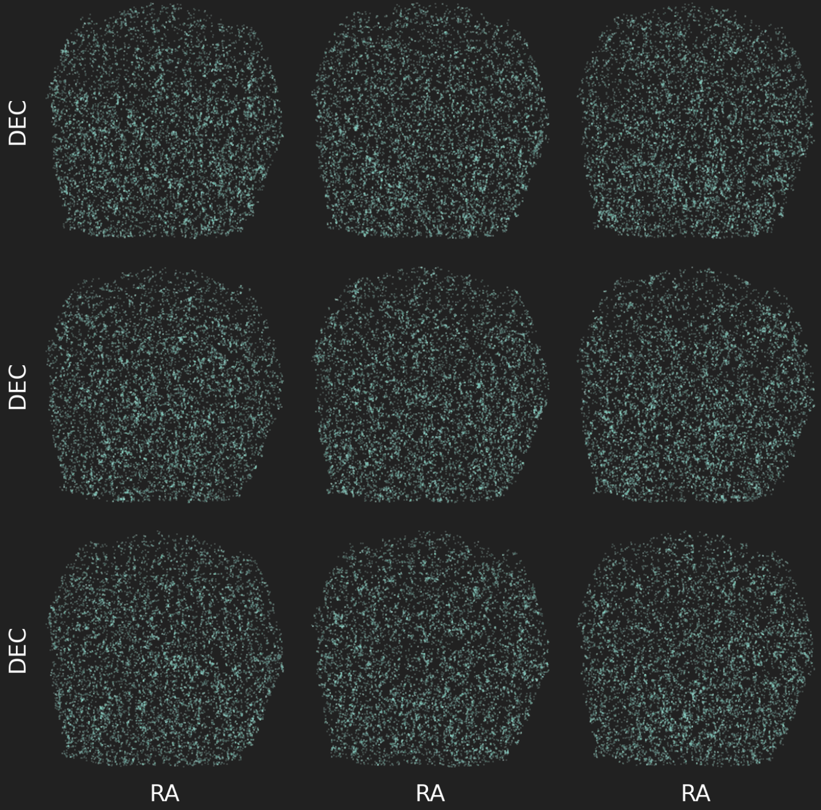
Spot our universe!

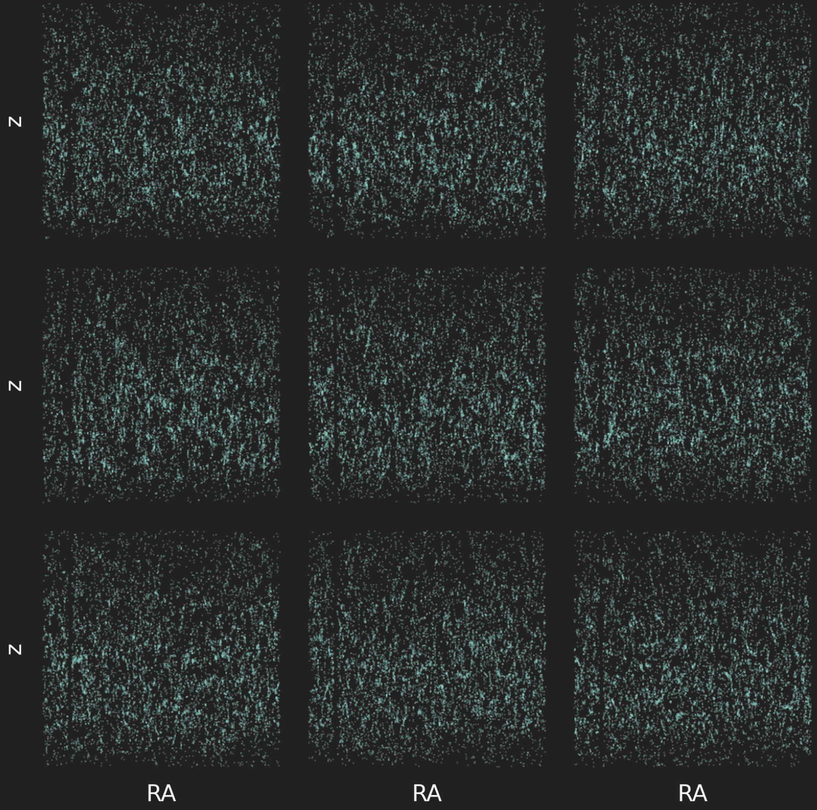
1x



8x



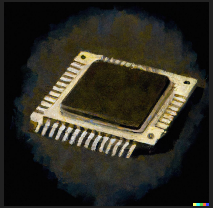


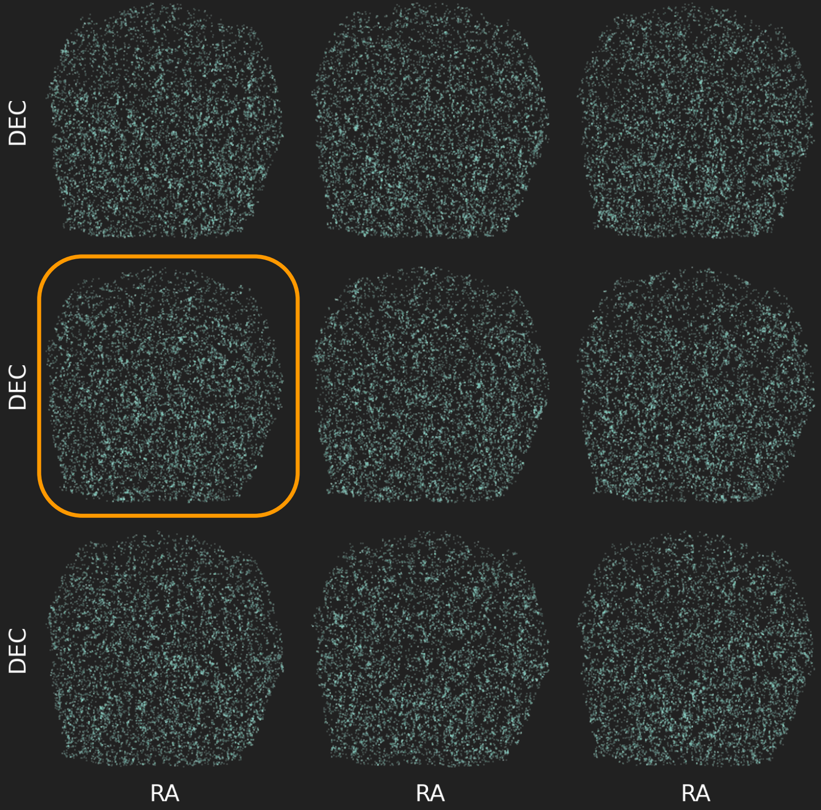


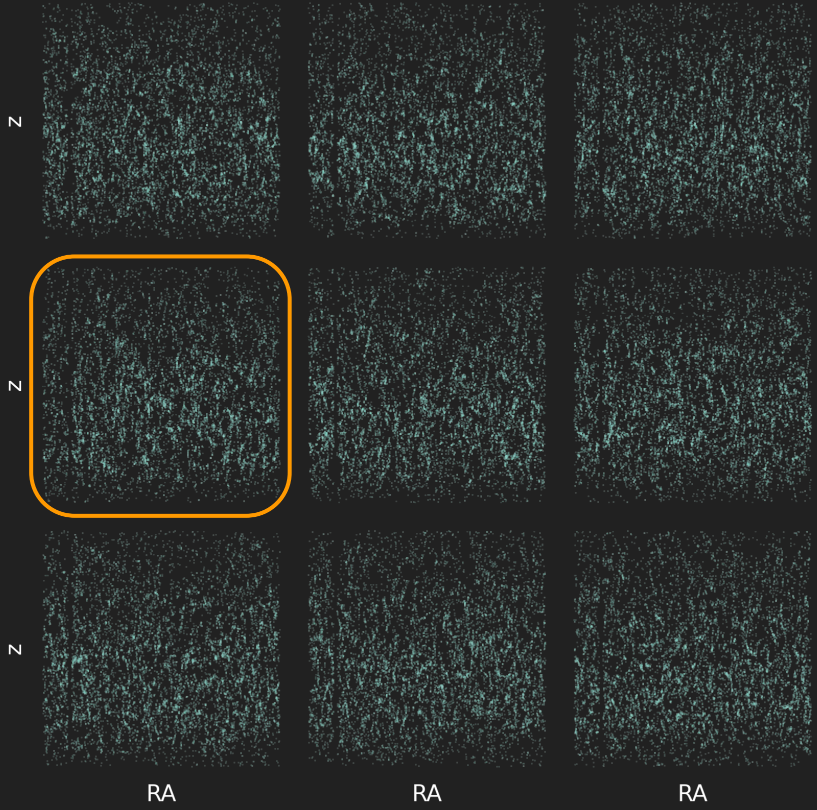
Spot our universe!

1x

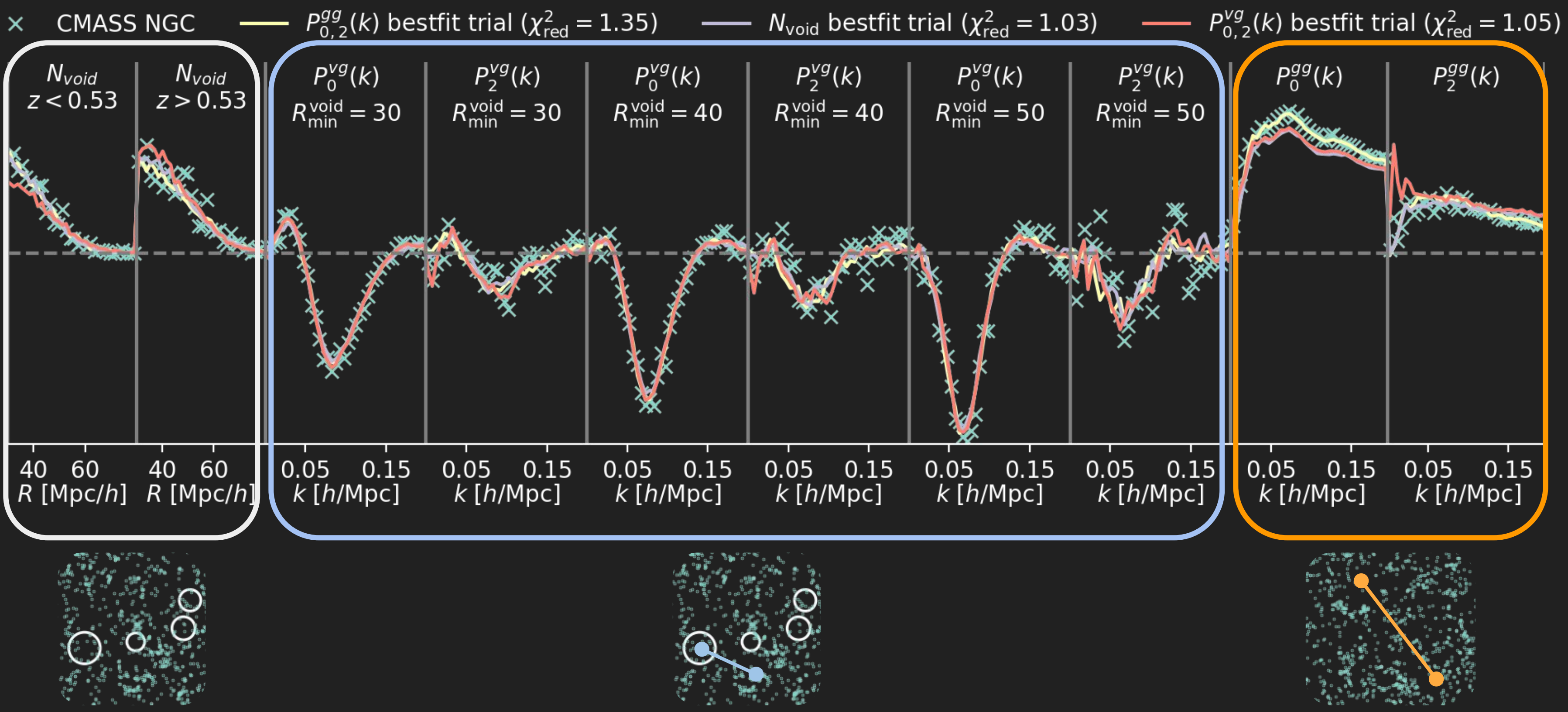
8x



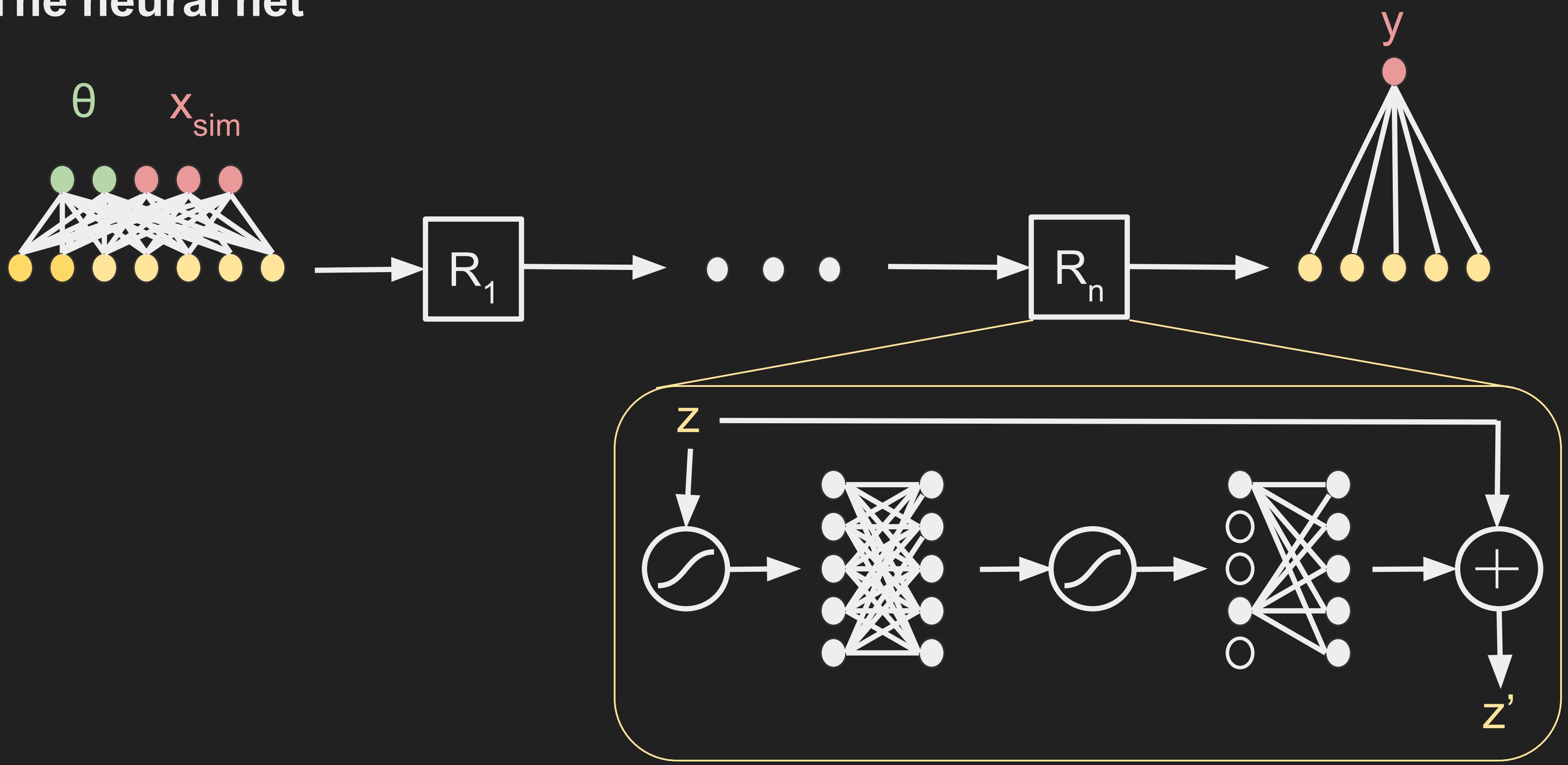




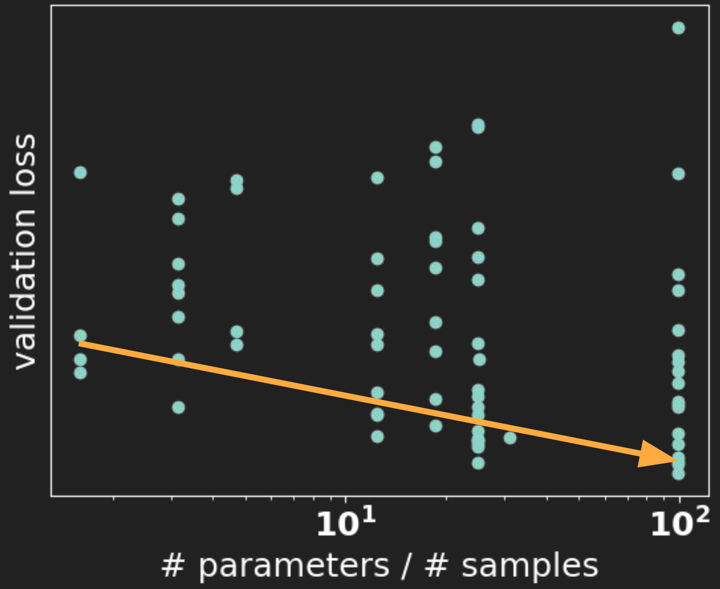
Simulations agree with data

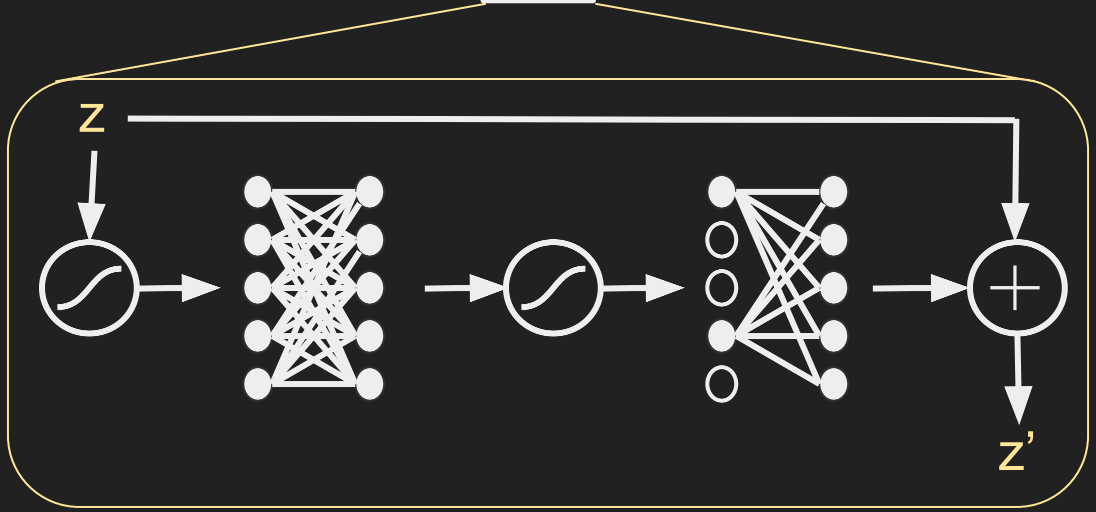


The neural net



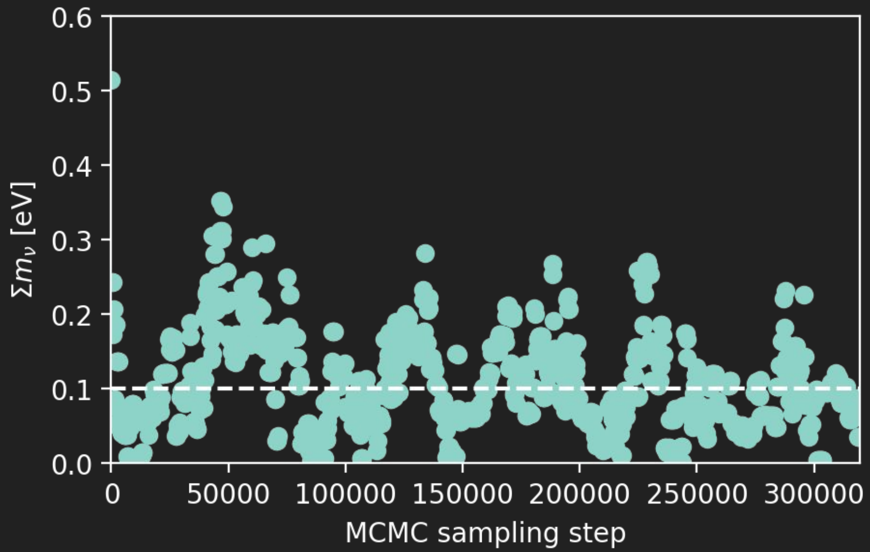
The neural net





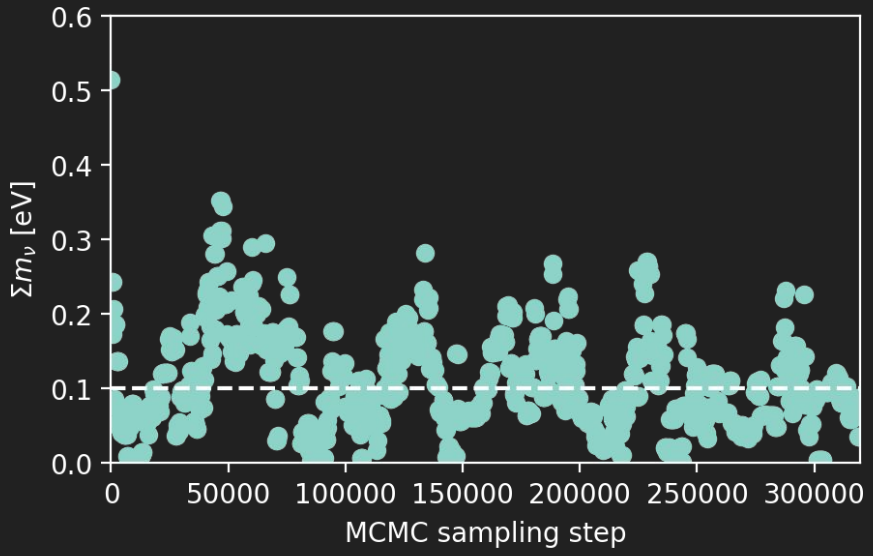
Testing trained neural network's quality

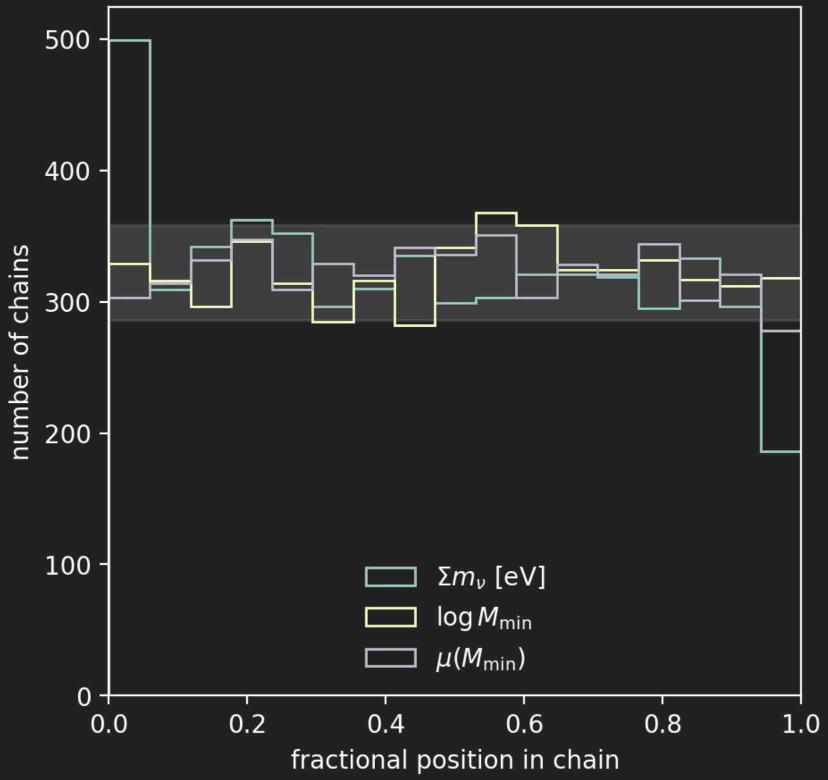
Randomly picking a posterior sample should be statistically indistinguishable from the true parameter value.



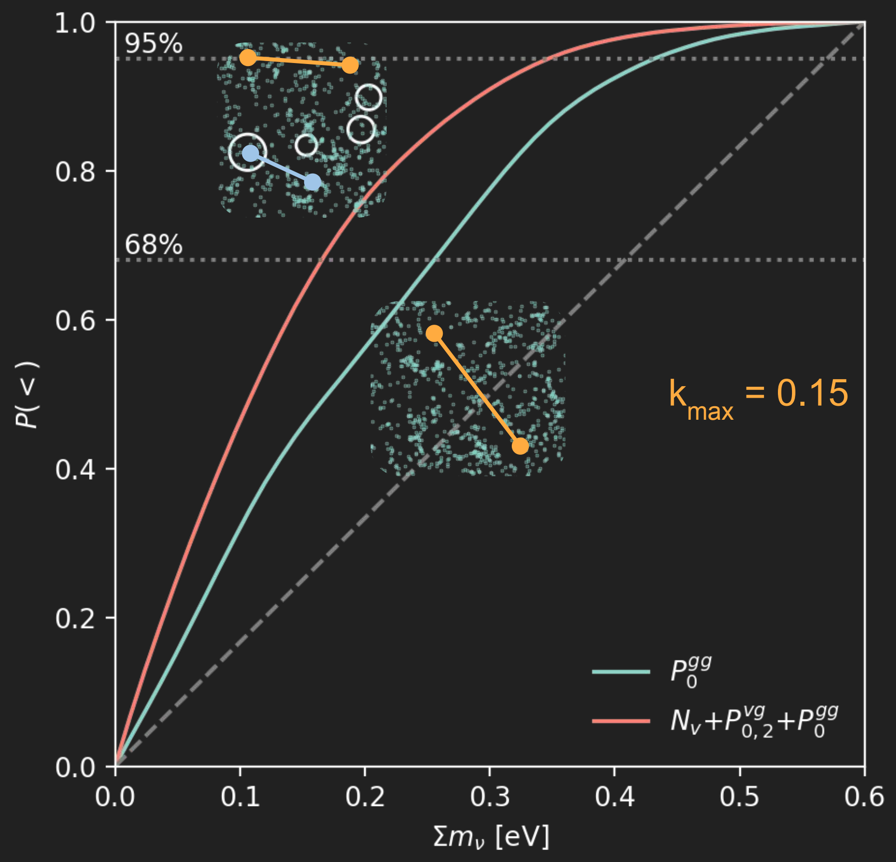
Testing trained neural network's quality

Randomly picking a posterior sample should be statistically indistinguishable from the true parameter value.

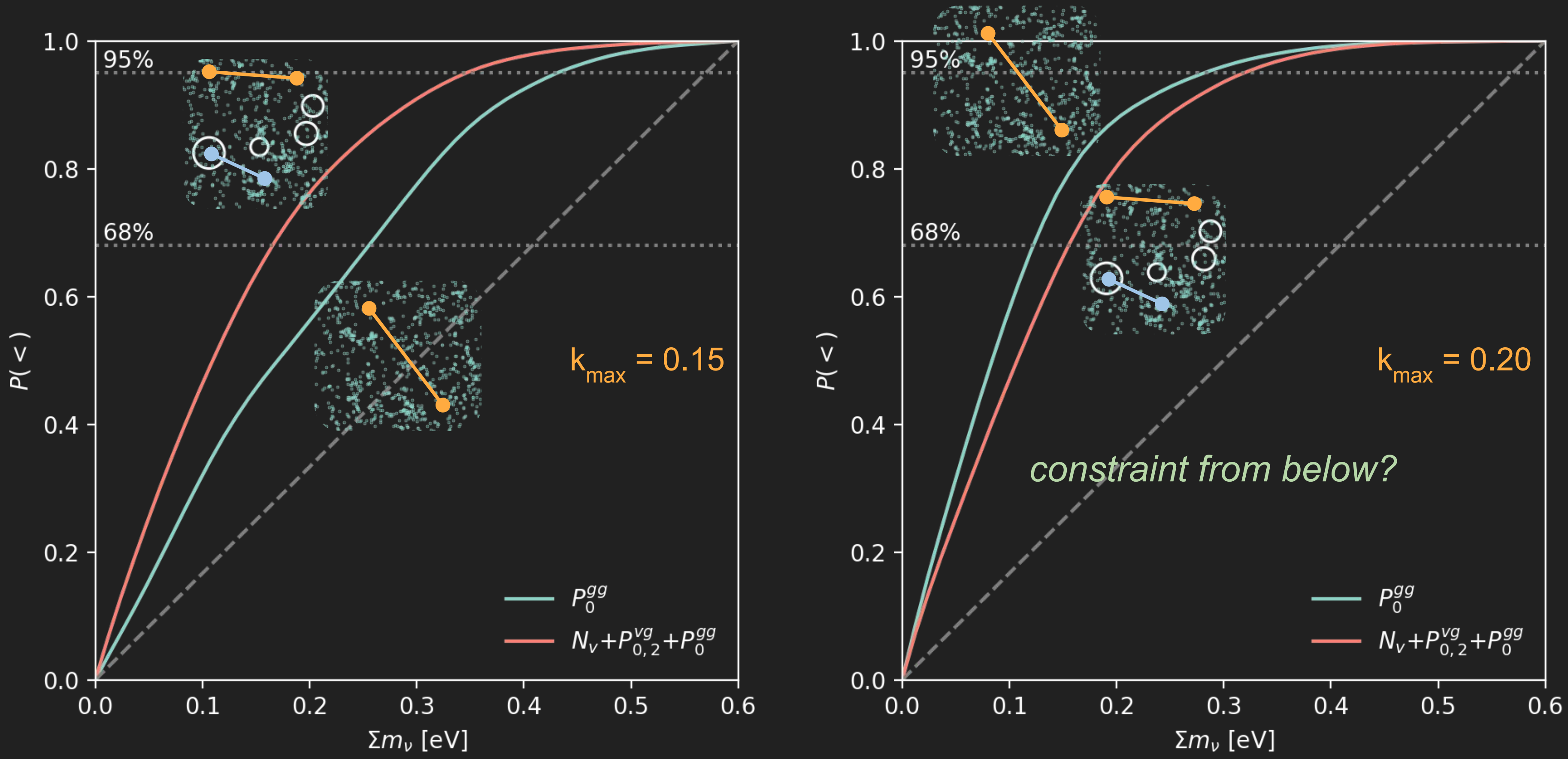




Neutrino mass constraints



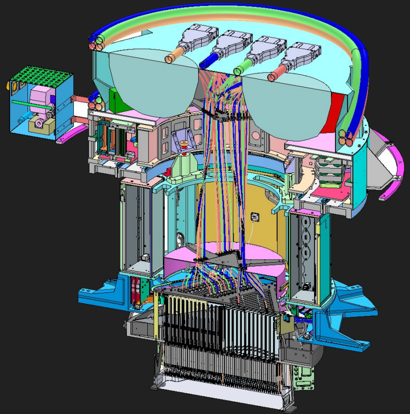
Neutrino mass constraints



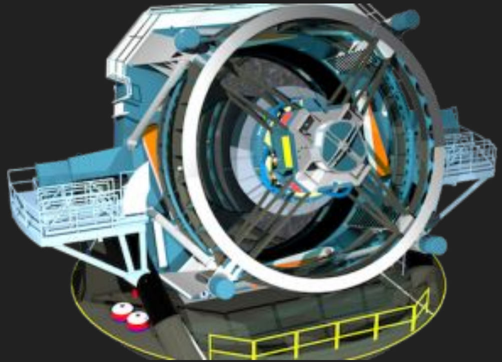
The future for voids

interplay between density contrast and number

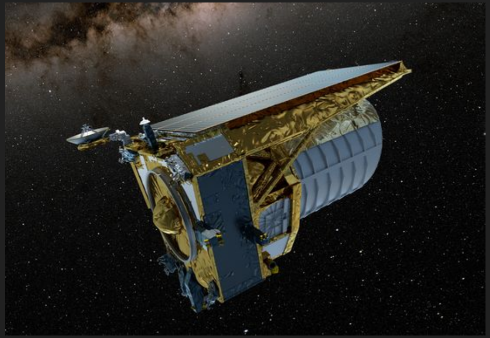
DESI: better numbers



PFS Lyman-α: clean small voids at higher redshifts

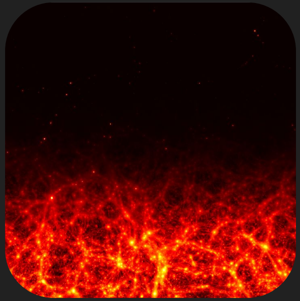


Rubin/LSST: large photo-z voids

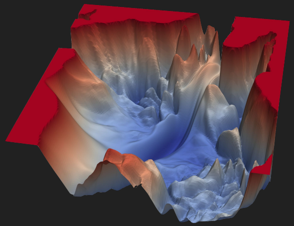


Euclid: everything?

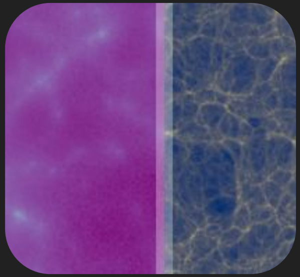
Summary



1) Emulating complicated physics can be done with deep learning. Precise problem formulation and domain knowledge help.

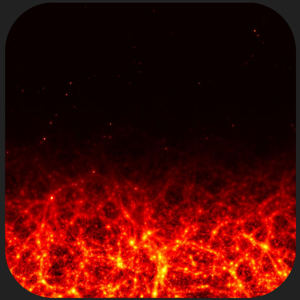


2) Deep learning has made great progress in the past decade. Over-parameterized neural nets have become trainable & state-of-the-art.

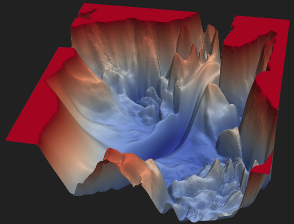


3) Implicit-likelihood inference enables learning a likelihood from simulated samples. Useful both if likelihood form unknown and as alternative to high-dimensional emulators.

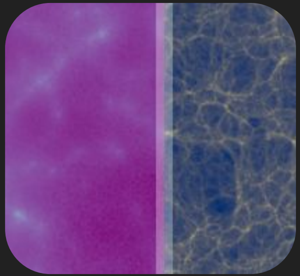
Summary



1) Emulating complicated physics can be done with deep learning. Precise problem formulation and domain knowledge help.



2) Deep learning has made great progress in the past decade. Over-parameterized neural nets have become trainable & state-of-the-art.



3) Implicit-likelihood inference enables learning a likelihood from simulated samples. Useful both if likelihood form unknown and as alternative to high-dimensional emulators.

Thank you for your attention!