Imperial College London

ICIC

European Research Council

erc

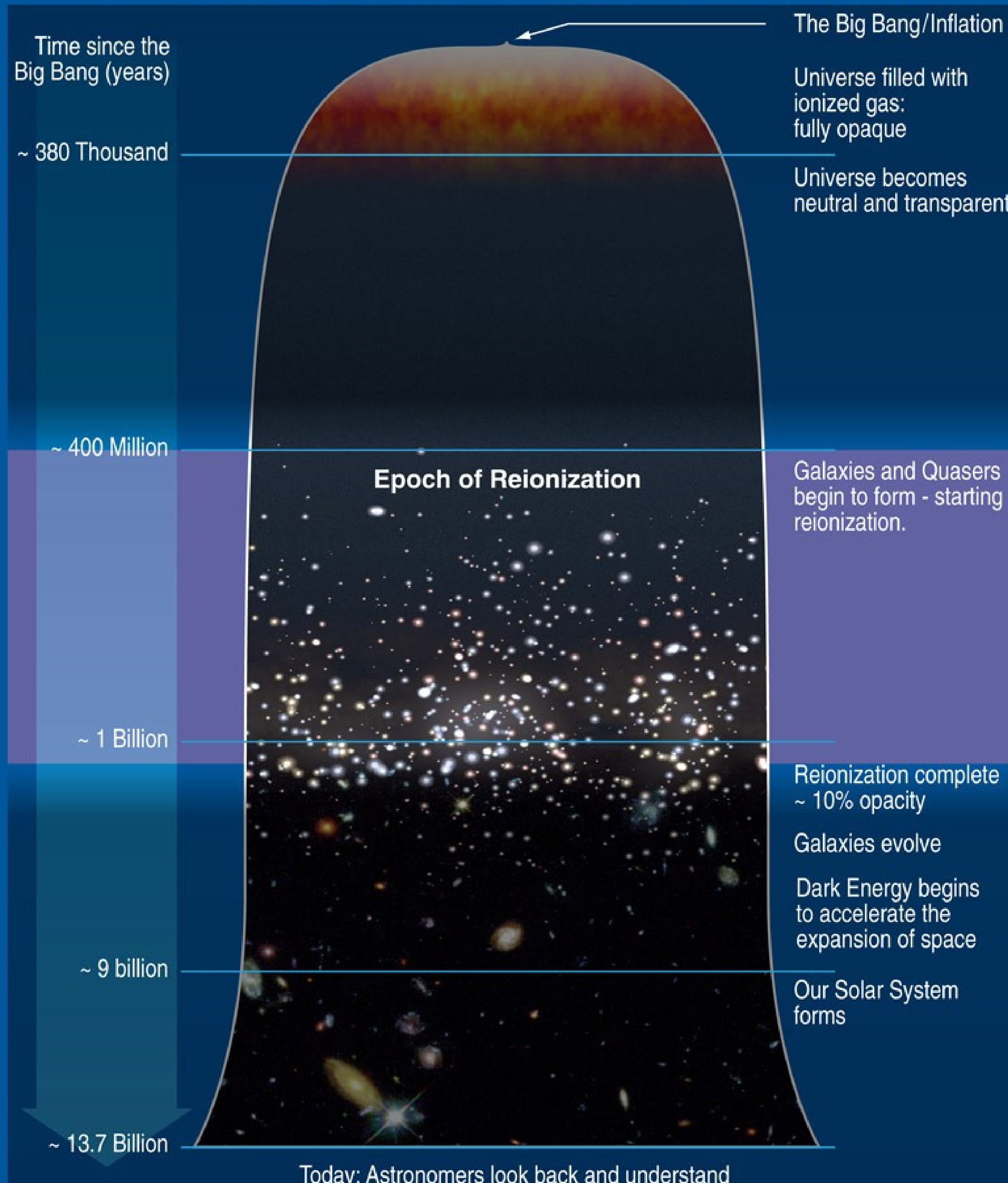
Emulating the Epoch of Cosmic Reionization

arXiv: 1708.00011

Claude Schmit

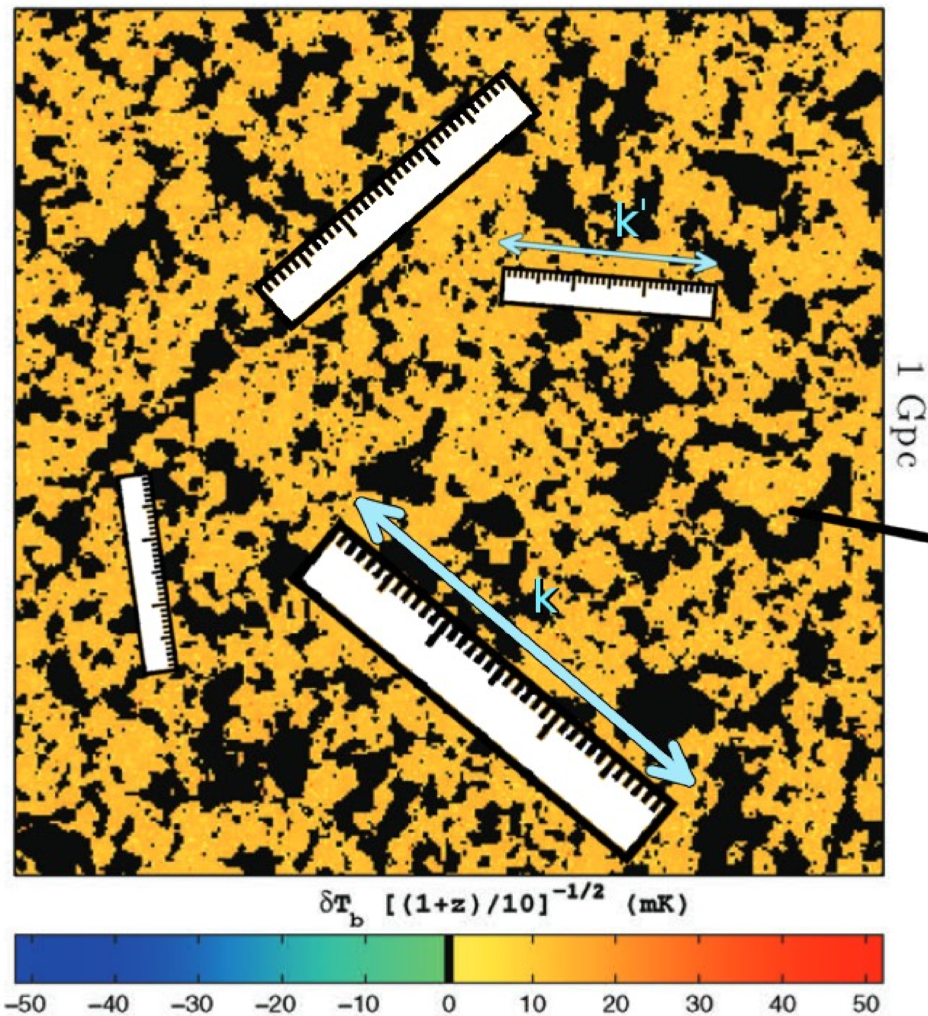
(supv. Jonathan Pritchard)

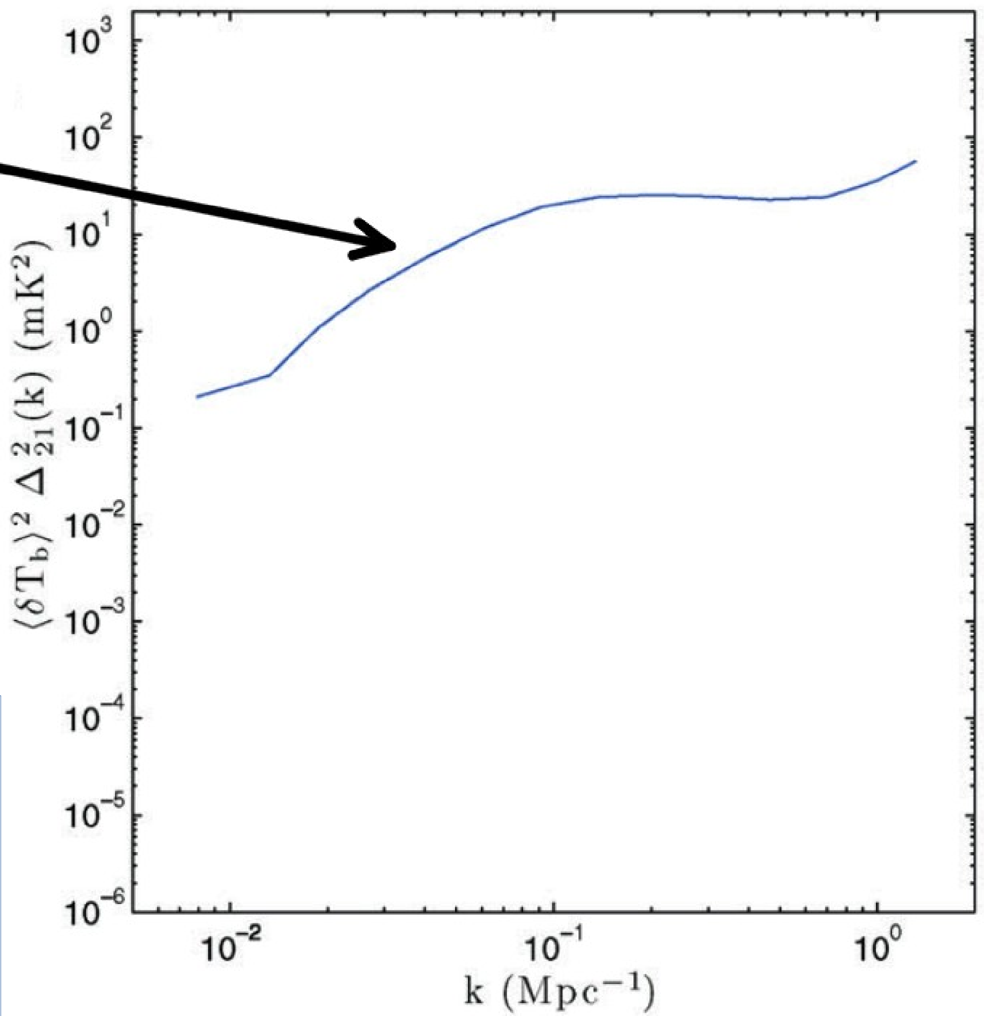
First Stars and Reionization Era



NASA/WMAP Science Team

We measure the correlation at different scales to obtain a power spectrum as a function of scale and redshift.

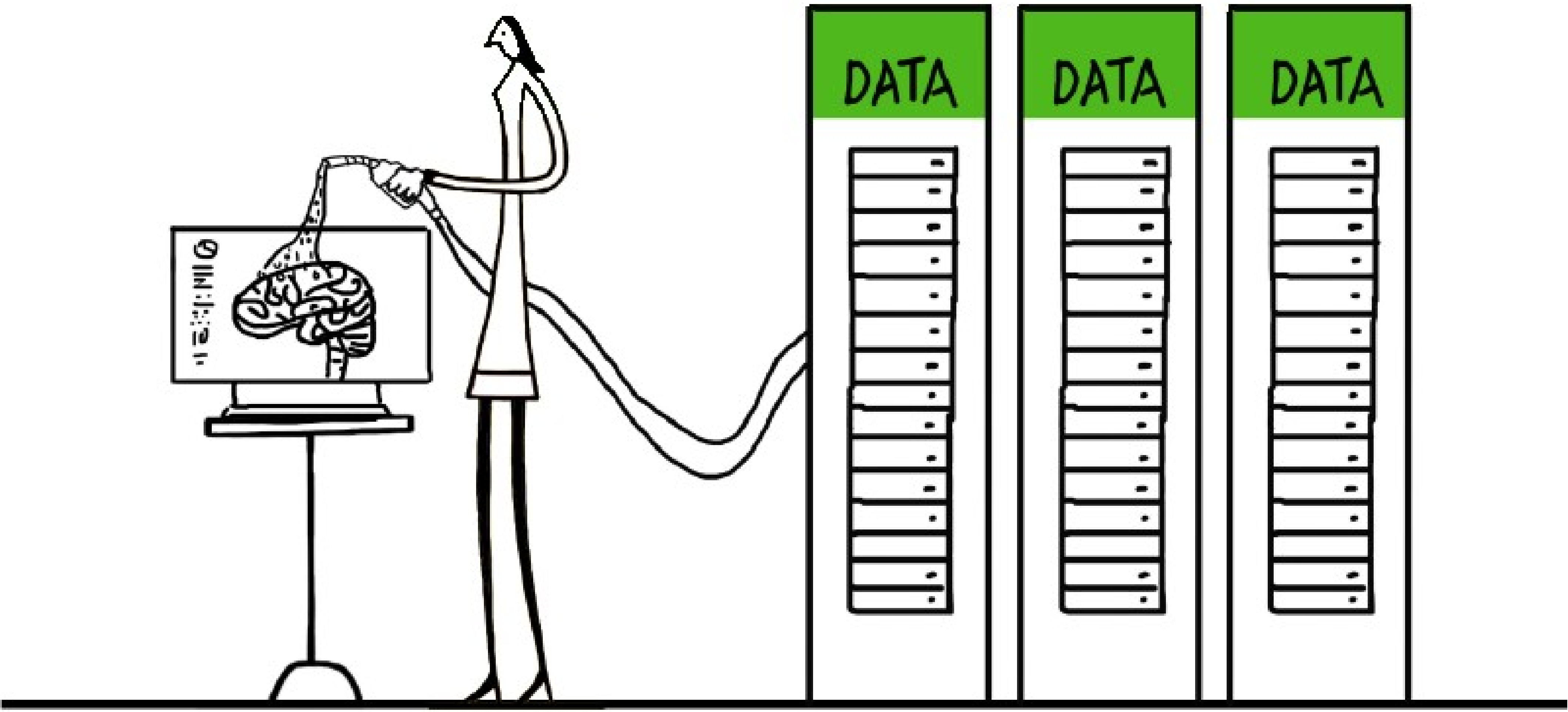




$$\delta_{21}(\boldsymbol{x}, z) = \delta T_b(\boldsymbol{x}) - \langle \delta T_b \rangle$$

$$\left\langle \tilde{\delta}_{21}(\boldsymbol{k}) \tilde{\delta}_{21}(\boldsymbol{k'}) \right\rangle = (2\pi)^3 \delta^D(\boldsymbol{k} - \boldsymbol{k'}) P_{21}(k)$$

Machine Learning



Machine Learning

Concepts:

- *Machine learning* techniques allow algorithms to learn patterns from data and make predictions according to those patterns.
- *Supervised learning*: The algorithm receives input and desired output pairs and creates a map connecting inputs and outputs.
- *Emulation*: Given a simulation model, an emulator is used to predict unknown simulation model outputs

Applications in Cosmology:

- Heitmann et al. (arXiv: 0902.0429): Precision Emulation of the Nonlinear Matter Power Spectrum using gaussian processes.
- Agarwal et al. (arXiv: 1203.1695): Non-linear matter power spectrum interpolation through artificial neural networks.

Machine Learning

Application in 21cm Cosmology:

Shimabukuro & Semelin (arXiv: 1701.07026):

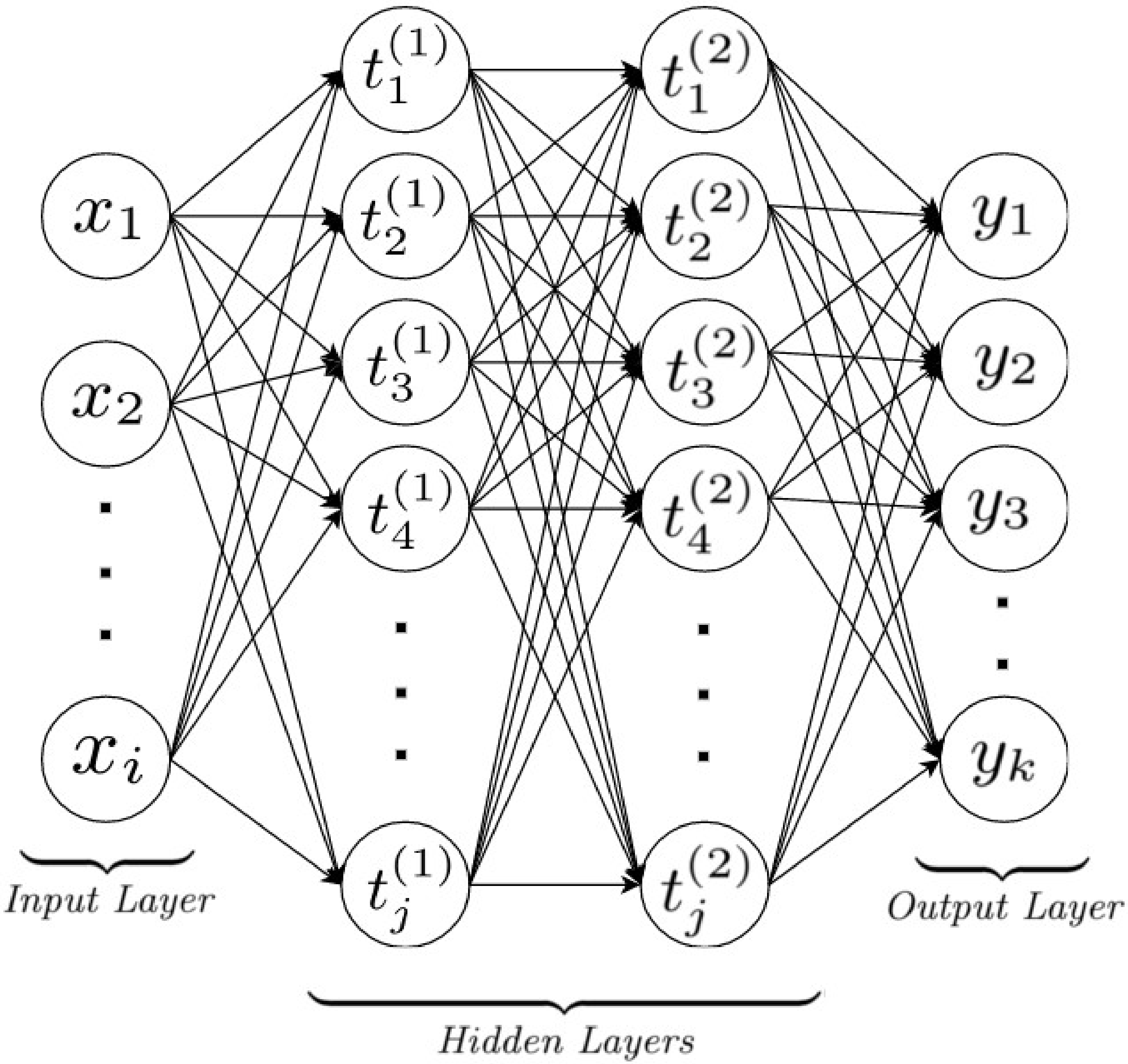
- Artificial Neural Network (ANN) trained on 21cmFast simulations.
- ANN used to predict the model parameters from the Power Spectrum.

The inverse Problem:

Kern et al. (arXiv: 1705.04688):

- Emulation of 21cmFast simulations using Gaussian Processes.
- Inference of the 21cmFast model parameters.

Our Neural Network Design



- Data is loaded into the input layer.
- The input nodes are combined into the hidden layer.

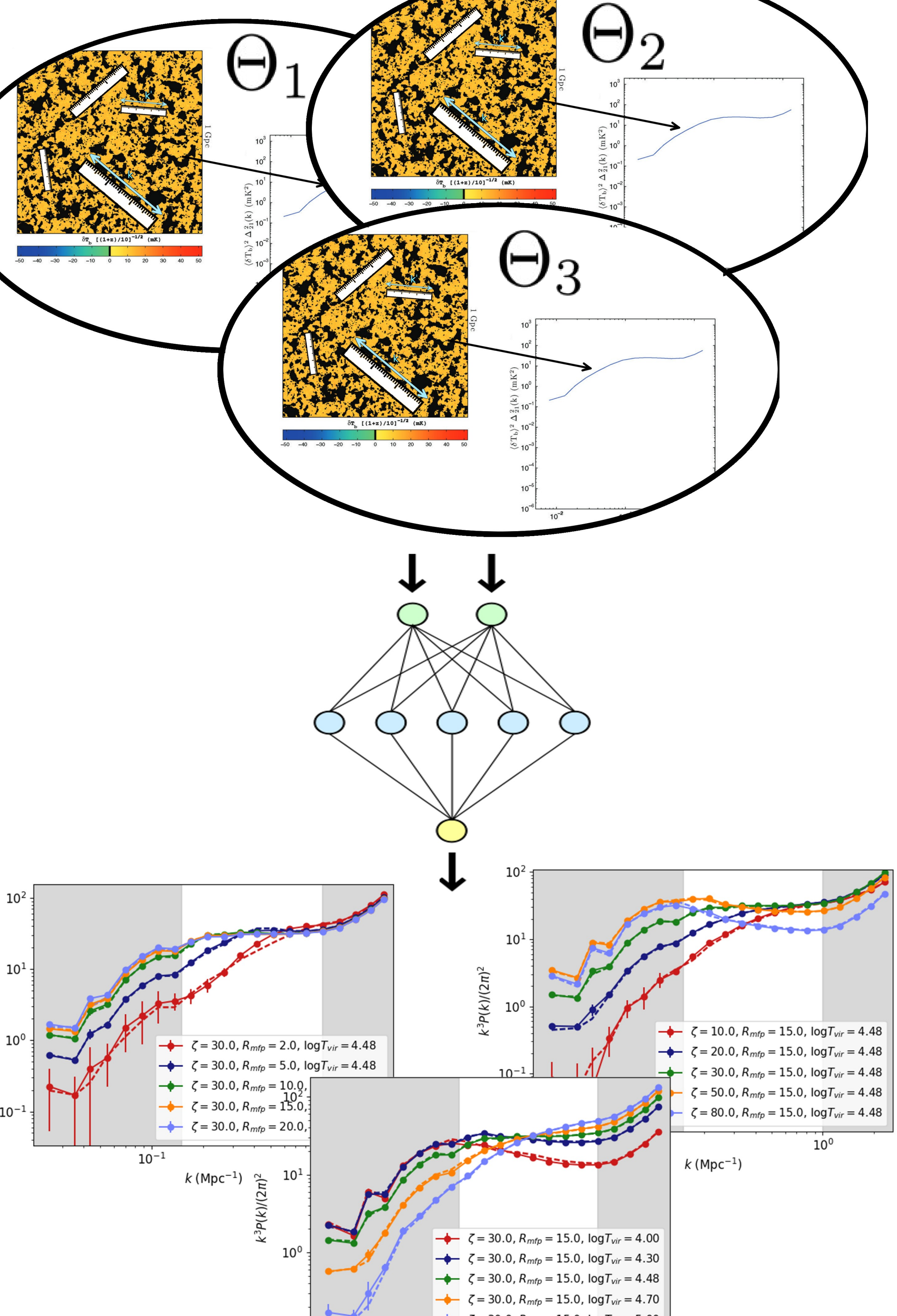
$$s_j^{(1)} = \sum_{i=1}^{N_i} x_i w_{ij}^{(1)}$$

- Each neuron is activated when certain features are present.

$$t_j^{(l)} = g\left(s_j^{(l)}\right)$$

- Activated neurons are combined into additional hidden layers or the output layer.

$$y_k = \sum_{i=1}^{N_j} t_i^{(L)} w_{ik}^{(L+1)}$$



- Evaluate the power spectrum at multiple points in parameter space to generate training sets using **21cmFast** optimised for MCMCs.
  - → Mesinger et al. (arXiv: 1003.3878)
  - → Greig and Mesinger (arXiv: 1501.06576)

- Create training sets using various amounts of samples for various training set sizes.

- Use these training sets to train a Neural Network.

- The Neural Network tells us how the power spectrum behaves at points we **did not** evaluate.

Training

ζ - Ionizing efficiency:

- Combines the star formation rate, the number of ionizing photons produced and the proportion of photons released into inter-galactic space

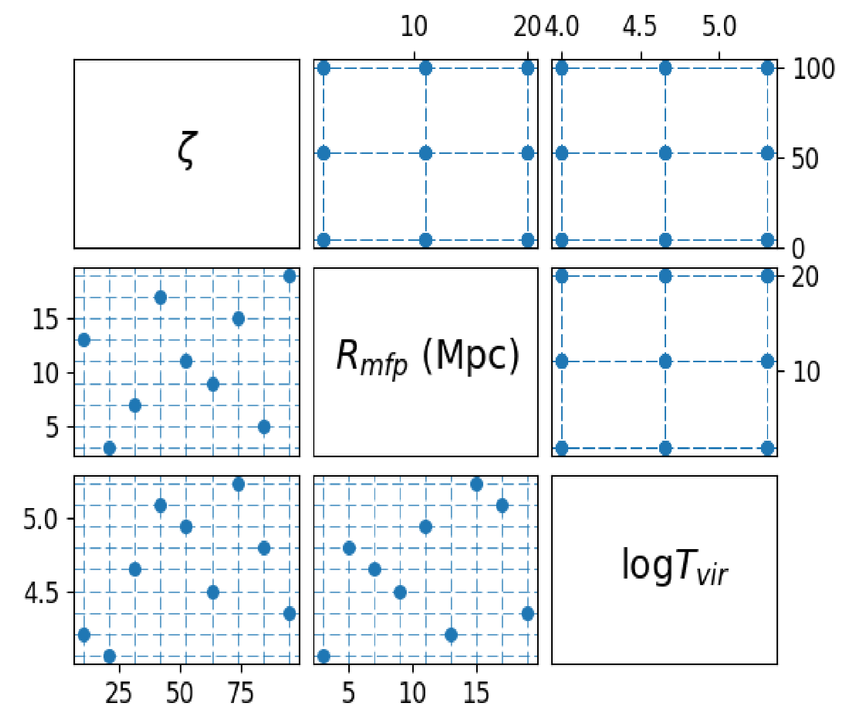
R_{mfp} - Mean free path of ionizing photons:

- Measures the average size of ionized bubbles around galaxies

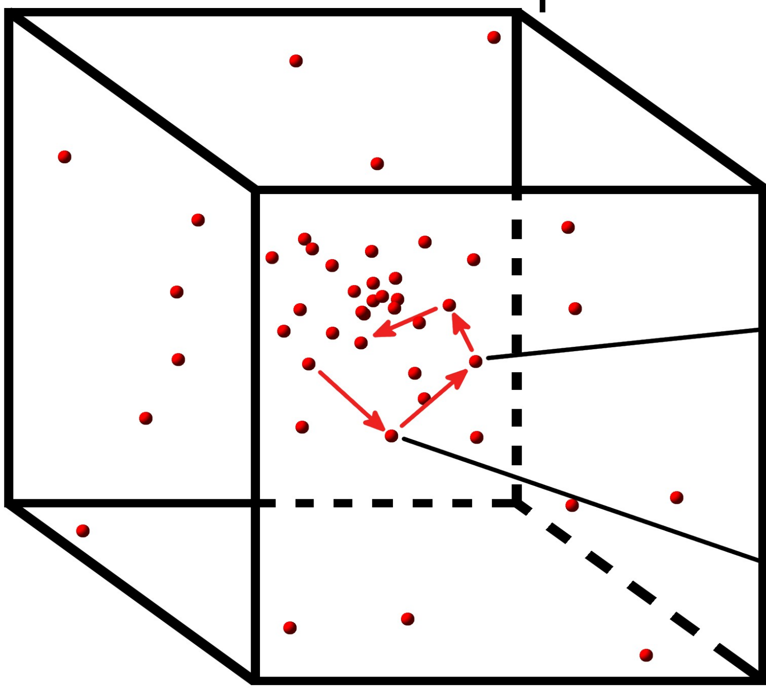
T_{vir} - Minimum virial temperature:

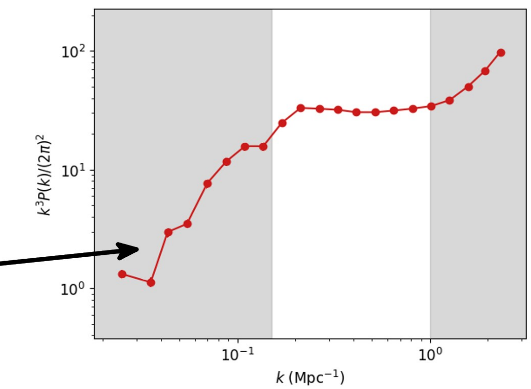
- Minimum virial temperature of gravitationally bound dark matter halos at which star formation occurs

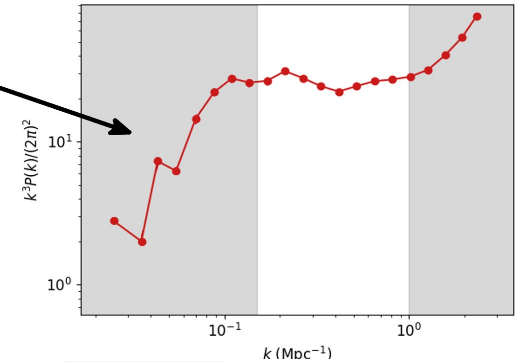
- Training set parameters drawn from **Latin hypercube**.
- Superior scaling properties to grid-based approach, for high dimensional parameter space.
- No parameter value is computed twice.
- Training sets with 100, 1000, 10000 LH samples generated.



- Use Neural Network to predict the power spectrum at any point in parameter space



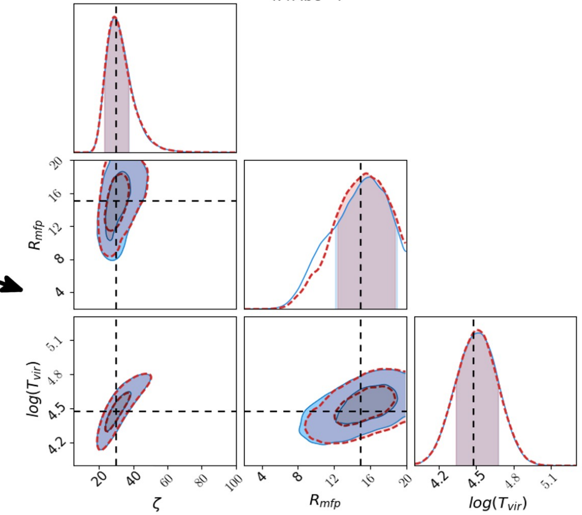




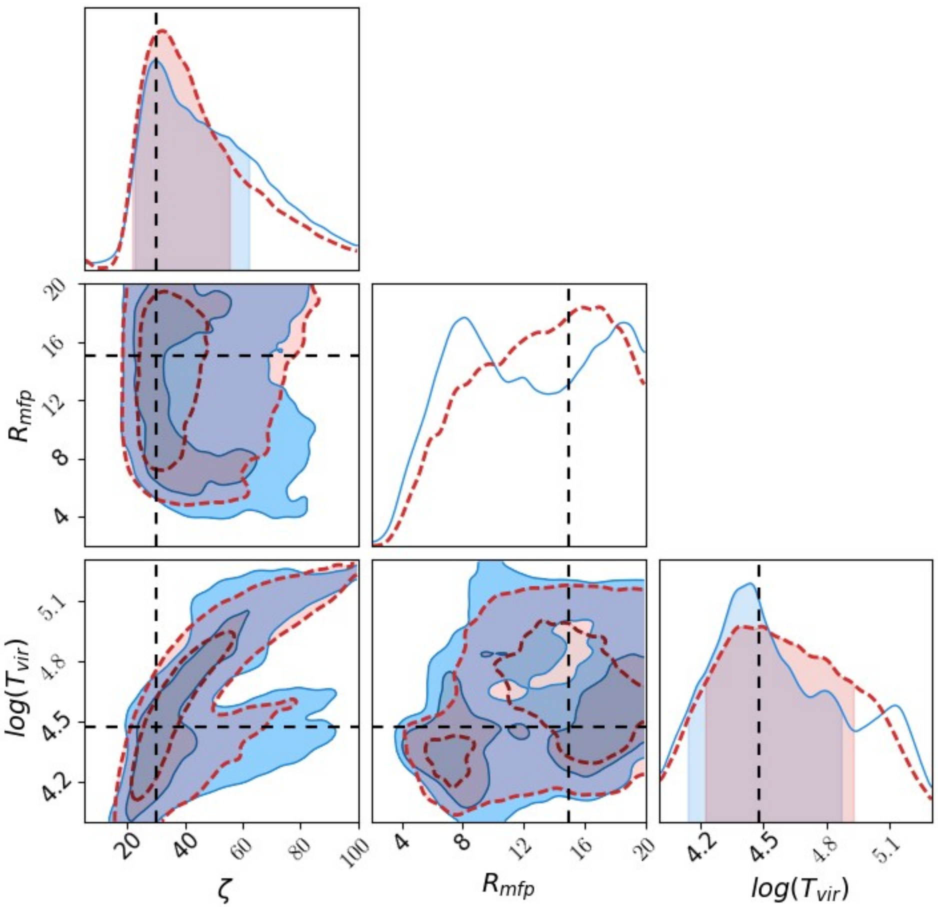
- Sample the likelihood:

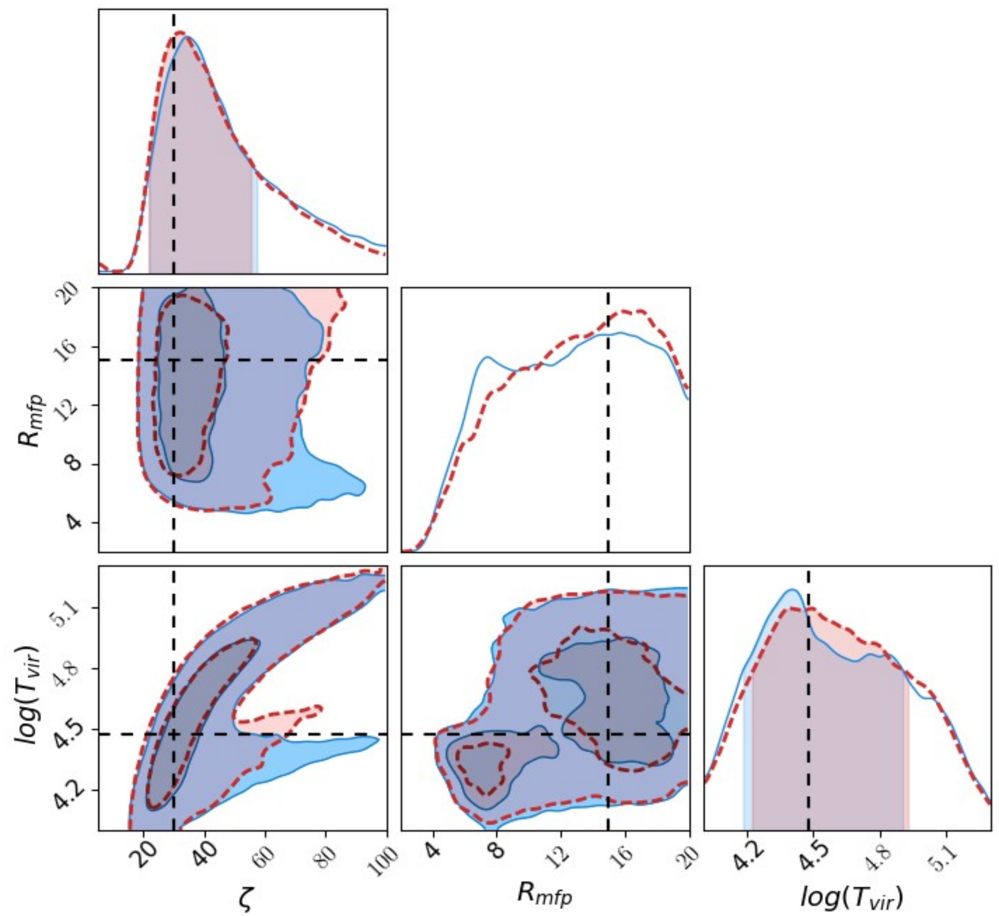
$$\ln \mathcal{L} = \frac{[P_{\rm obs}(k) - P(k|\boldsymbol{\theta})]^2}{2P_N(k)}$$

- Determine best fit parameters:



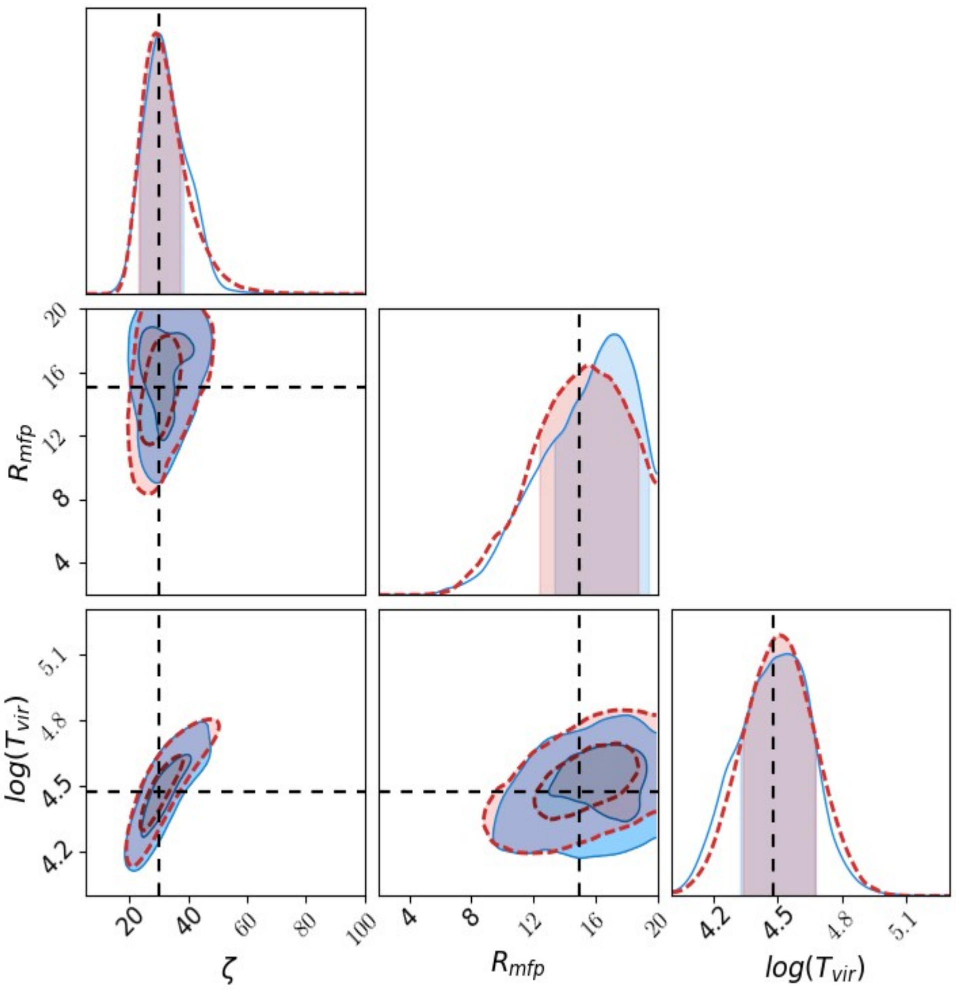
Fitting *single* redshift observations at $z = 9$

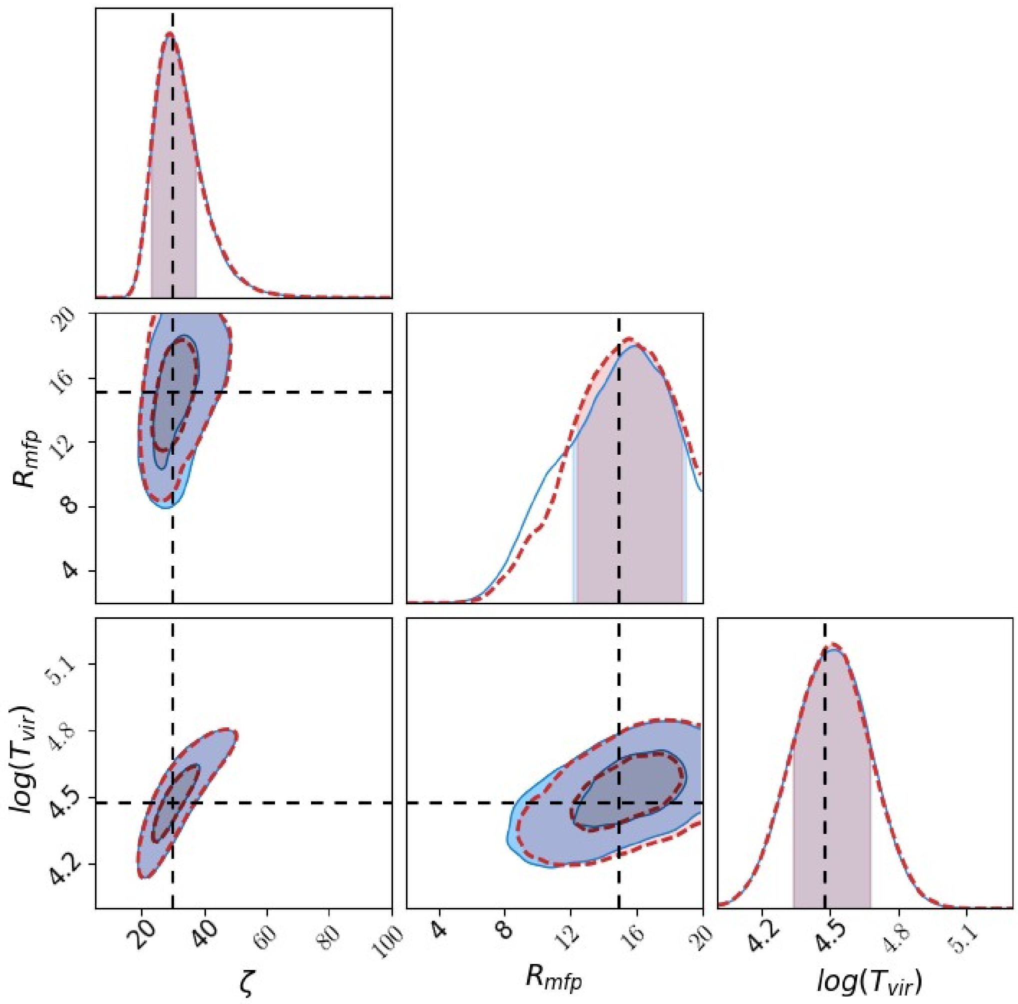




- Using 100 training samples.
- Using 1000 training samples.

Combining z = *8*, *9* and *10*





- Using 100 training samples at 3 redshifts.
- Using 1000 training samples at 3 redshifts.

Results

Table 1. Median values and 68% confidence interval found in the parameter search via the brute-force method (21CMMC) and our ANN emulation at $z = 9$ and $z = 8$. The fiducial parameter values for both redshifts are given by $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

Code - Training Set	z	ζ	$R_{\rm mfp}$	$\log T_{\rm vir}$
21CMMC	9	$41.28^{+24.85}_{-13.43}$	$13.38^{+4.28}_{-5.15}$	$4.59^{+0.37}_{-0.32}$
ANN - 100LHS	9	$45.47^{+25.19}_{-17.18}$	$12.13^{+5.71}_{-5.05}$	$4.54^{+0.47}_{-0.28}$
ANN - 1000LHS	9	$42.52^{+26.18}_{-13.74}$	$12.89^{+4.63}_{-5.29}$	$4.57^{+0.40}_{-0.31}$
ANN - 10000LHS	9	$42.21^{+25.42}_{-14.12}$	$13.18^{+4.46}_{-5.14}$	$4.58^{+0.39}_{-0.31}$
21CMMC	8	$39.64^{+31.90}_{-16.11}$	$14.99^{+2.98}_{-3.64}$	$4.61^{+0.21}_{-0.23}$
ANN - 100LHS	8	$43.06^{+26.16}_{-17.38}$	$14.58^{+3.47}_{-3.90}$	$4.64^{+0.19}_{-0.25}$
ANN - 1000LHS	8	$42.71^{+31.30}_{-18.67}$	$14.67^{+3.19}_{-4.26}$	$4.62^{+0.21}_{-0.23}$
ANN - 10000LHS	8	$39.78^{+31.68}_{-16.22}$	$14.61^{+3.15}_{-4.05}$	$4.60^{+0.22}_{-0.23}$
21CMMC	8,9,10	$31.08^{+8.70}_{-6.04}$	$15.15^{+2.86}_{-3.21}$	$4.51^{+0.17}_{-0.17}$
ANN - 100LHS	8,9,10	$31.51^{+8.57}_{-6.32}$	$15.86^{+2.47}_{-3.62}$	$4.49^{+0.16}_{-0.19}$
ANN - 1000LHS	8,9,10	$31.18^{+8.47}_{-6.08}$	$14.97^{+2.91}_{-3.78}$	$4.51^{+0.16}_{-0.17}$

- We are able to recover the true parameters of the observations within a 68% confidence limit.
- Comparing to 21CMMC which evaluates the model at each point, we get a speed-up of **three** orders of magnitude.
- This technique potentially enables us to use radiative transfer simulations for parameter inference.
- 21SSD: Radiative transfer simulation database (Semelin et al. arXiv: 1707.02073) can be used as training set.

Concluding Remarks

- Modelling the Epoch of Reionization is difficult and evaluating models can be slow.
- Future experiments will collect huge amounts of data which needs to be processed fast and reliably.
- Machine learning is key to process data fast.
- We find speed-ups of 3 orders of magnitude using Neural Networks for bayesian parameter inference.