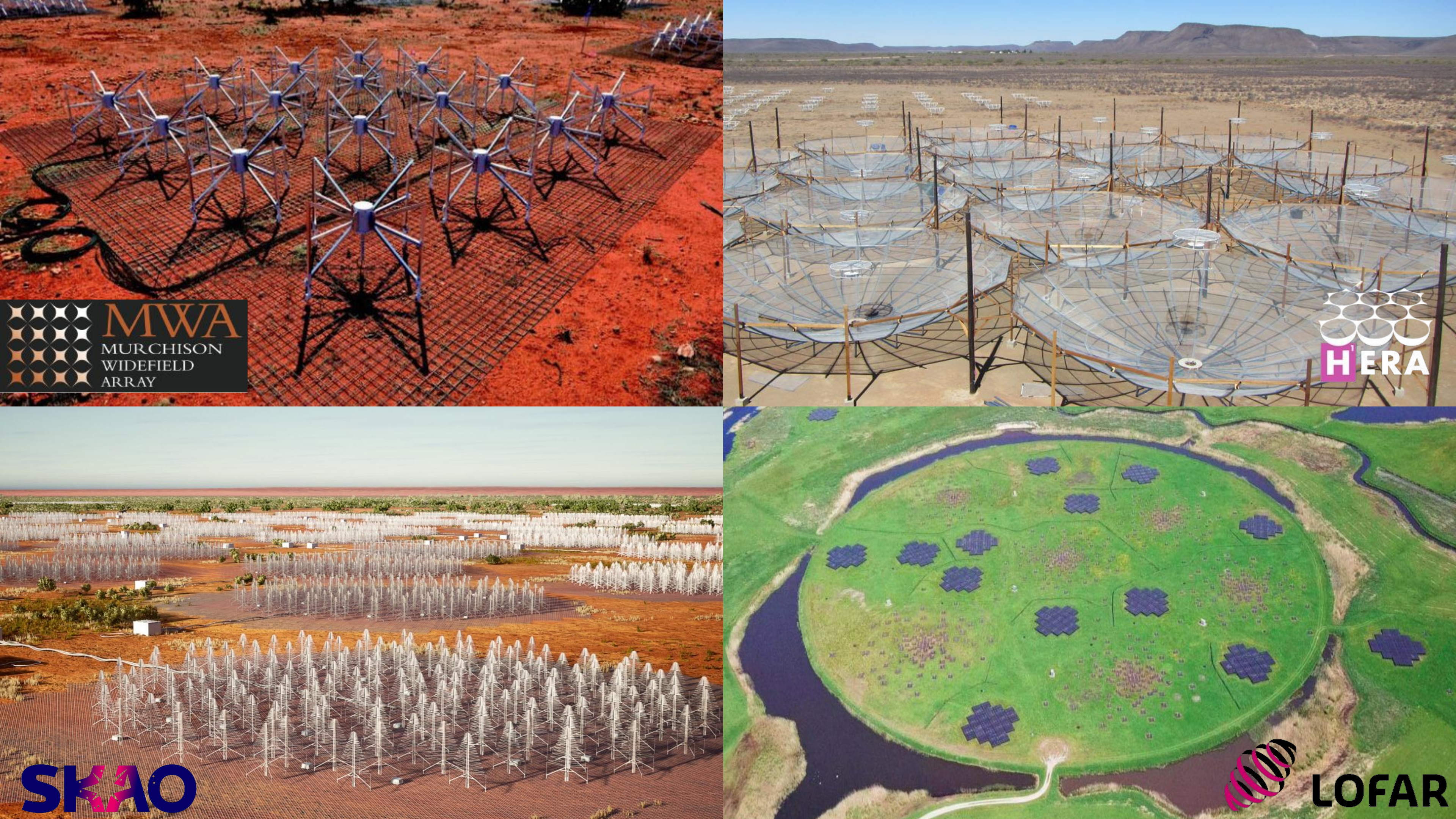
Toward a complete machine learning-based forward modelling pipeline

Daniela Breitman
Scuola Normale Superiore
Supervised by Andrei Mesinger
SKA CD/EoR ST Meeting, Tsinghua
July 18, 2024

SCUOLA NORMALE SUPERIORE

daniela.breitman@sns.it


MWA
MURCHISON
WIDEFIELD
ARRAY
HERA
SKAO
LOFAR

How do we learn from this data?


BAYES' THEOREM
P(θ|d) = P(d|θ)P(θ) / P(d)

Bayes' Theorem

- **What we want** P(θ|d) = updated knowledge on θ provided an observation d
- **Forward model** P(d|θ) = probability of reproducing observation d with model parameters θ
- P(θ) = prior on the model parameters θ

$$P(\theta|d) = \frac{P(d|\theta)P(\theta)}{P(d)}$$

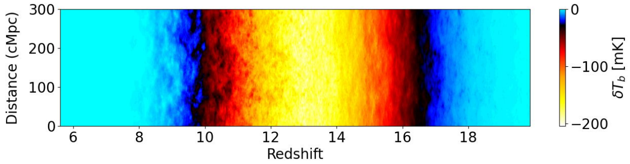
Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

Traditional Pipeline

Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

Traditional Pipeline

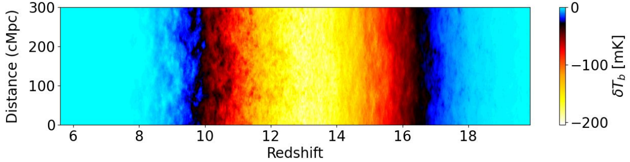
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$



Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

Traditional Pipeline

Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_\parallel, z)$

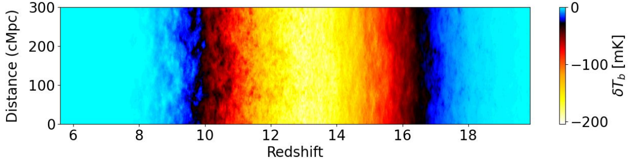


Sample params: $\theta_{astro} \sim P(\theta_{astro}), \theta_{cosmo}$, ICs

Traditional Pipeline

Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

≳ 300k CPU hrs

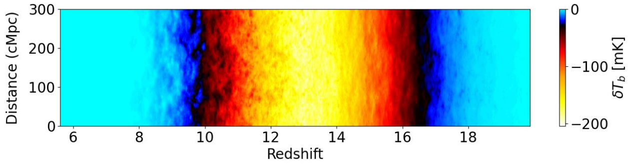


Sample params: $\theta_{astro} \sim P(\theta_{astro}), \theta_{cosmo}$, ICs

Traditional Pipeline

Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

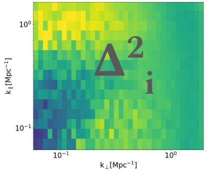
≳ 300k CPU hrs



This work

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

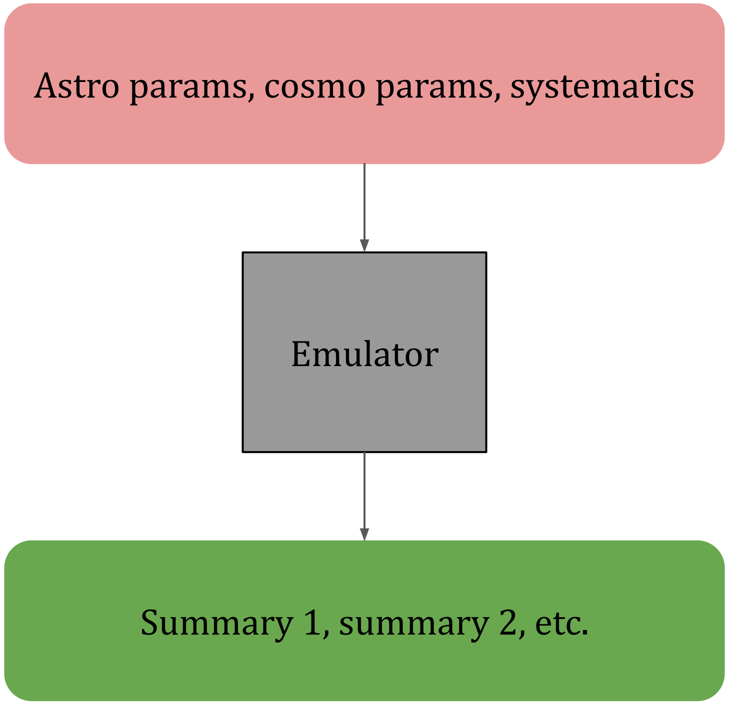
≲ 1 GPU day



, etc.

Breitman+in prep.

Ideal Emulator



Summary Statistics

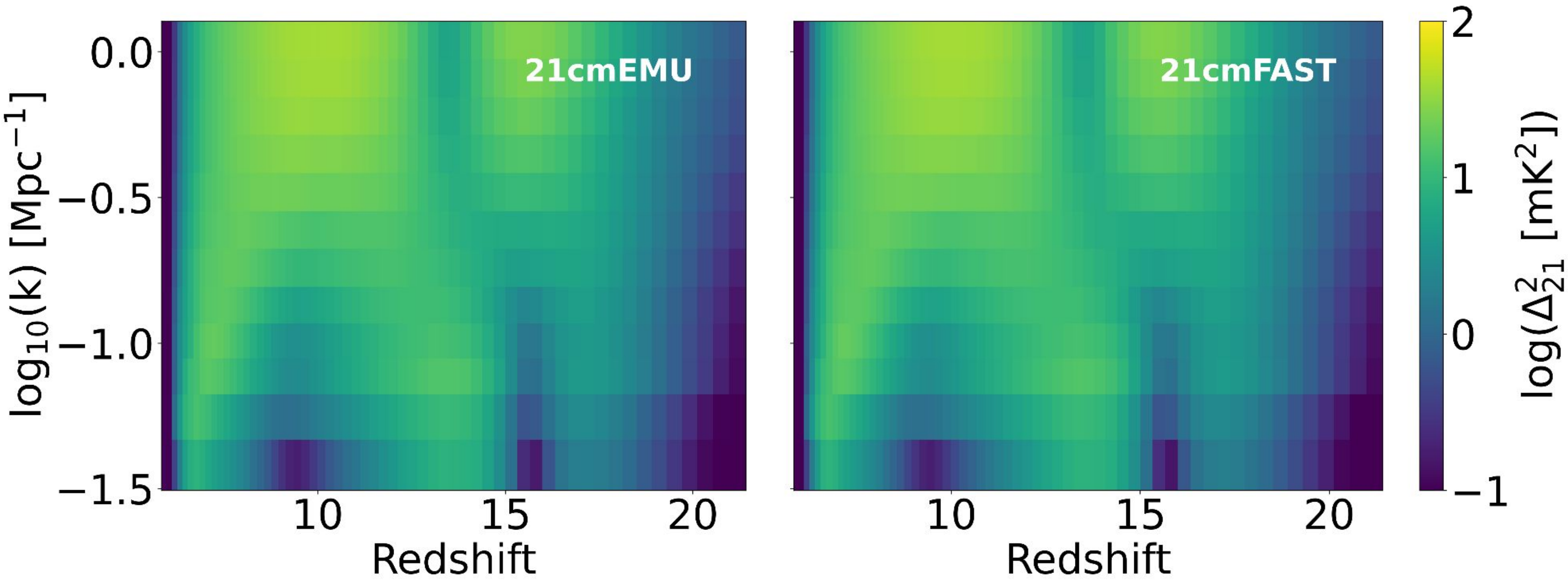
- **xHI (z)**: Global fraction of neutral hydrogen
- **T_S(z)**: Global neutral IGM spin temperature $$T_S^{-1} = \frac{T_\gamma^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}$$
- **T_b(z)**: Global 21-cm brightness temperature [e.g. Bye+22, Bevins+21, Cohen+19]

$$\delta T_b \approx 27 x_{HI} (1+\delta_b) \left(\frac{\Omega_b h^2}{0.023}\right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10}\right)^{1/2} \times \left(\frac{T_S - T_R}{T_S}\right) \left[\frac{\partial_r v_r}{(1+z) H(z)}\right] \mathrm{mK}$$

- **21-cm power spectrum (k, z)**: e.g. Kern+17, Schmit & Pritchard 18, Ghara+20, Mondal+21
- **UV luminosity function:** Source number density per magnitude vs z [Bouwens+15, 16]
- **τ_e**: Thomson scattering optical depth of CMB photons

Breitman+23

Power Spectrum



Breitman+23

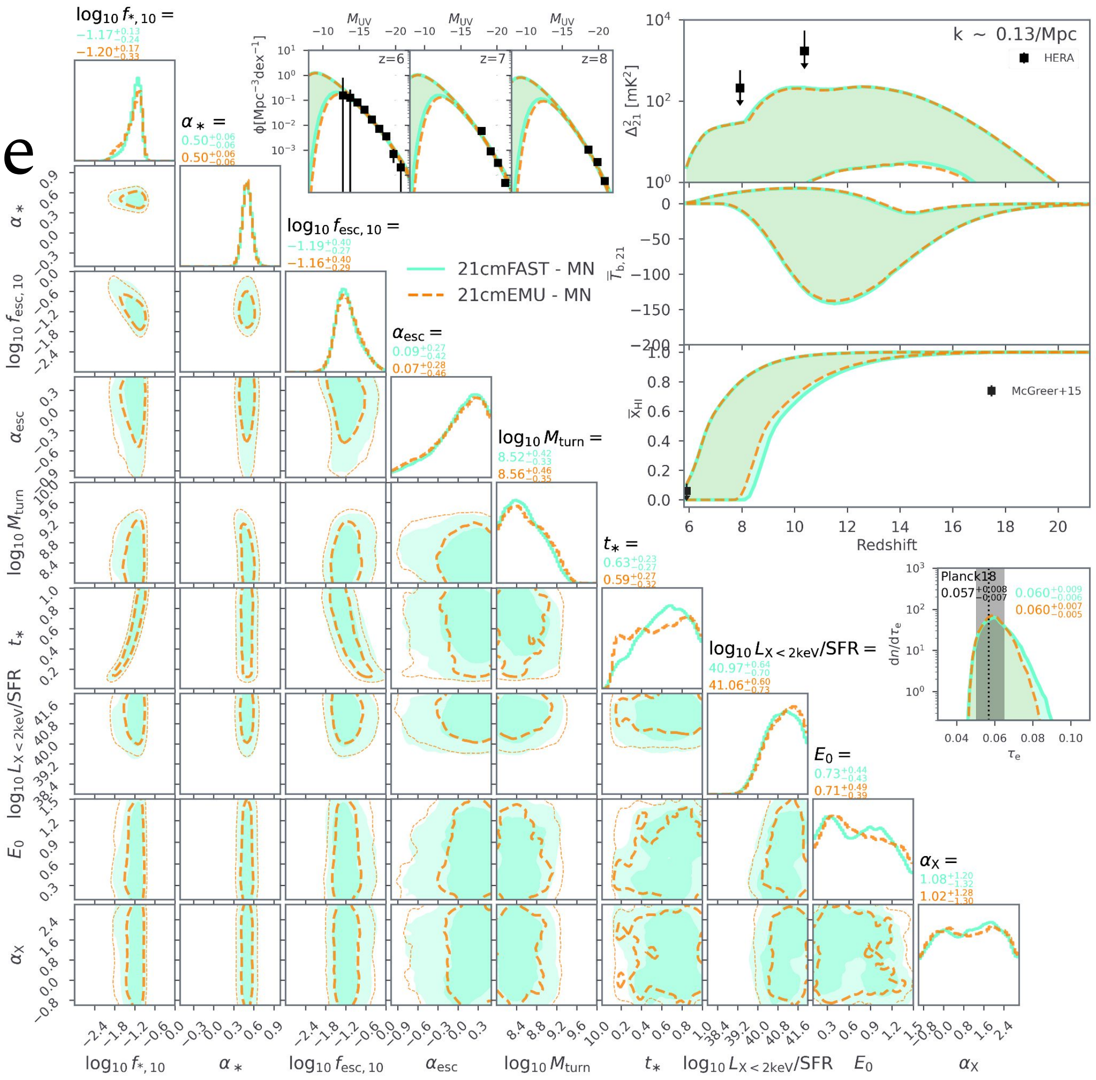
21cmEMU Performance

Summary	Median FE (%)	68% CL (%)
$\log \Delta_{21}^2$	0.55	2.4
$\overline{T}_\mathrm{b}$	0.34	1.2
$\log \overline{T}_\mathrm{S}$	0.032	0.13
$\overline{x}_\mathrm{HI}$	0.0073	0.10
τ_e	0.11	0.26
$\log \phi$	0.50	2.1

Breitman+23

Small errors, but do they impact the inference?

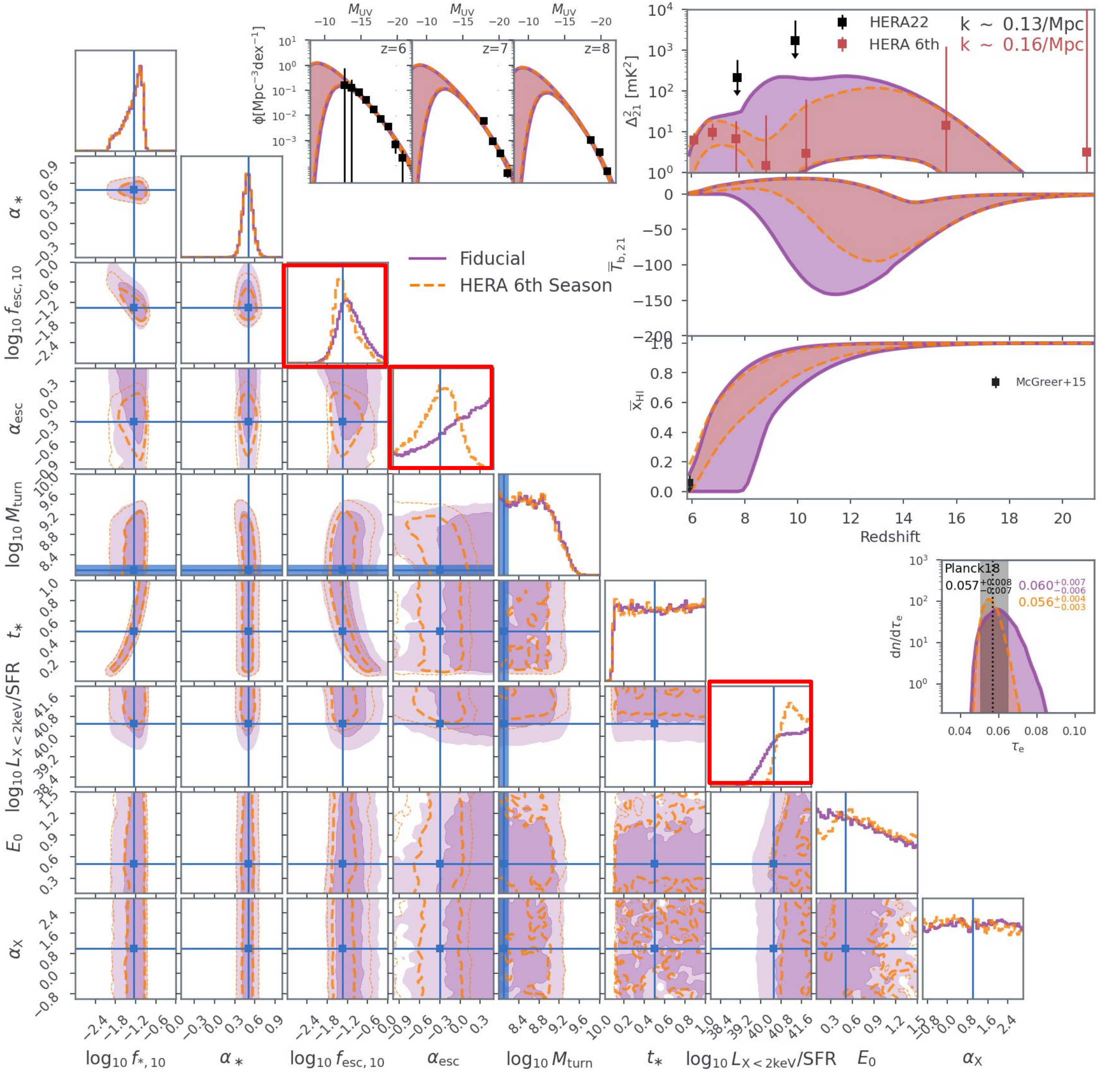
21cmEMU can reproduce 21cmFAST results.



Breitman+23

Forecast for HERA Phase II 6th-Season Observations

New HERA data + tighter prior on systematics improve the constraints for many astrophysical parameters



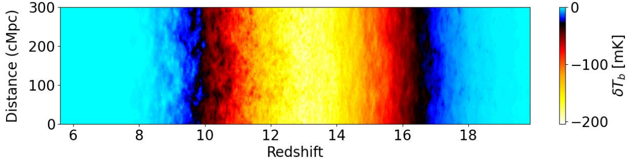
Breitman+23

Sample params: $\theta_{astro} \sim P(\theta_{astro}), \theta_{cosmo}$, ICs

Traditional Pipeline

Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

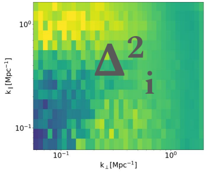
$\gtrsim$ 300k CPU hrs



This work

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

$\lesssim$ 1 GPU day



, etc.

Breitman+in prep.

Sample params: $\theta_{astro} \sim P(\theta_{astro}), \theta_{cosmo}$, ICs

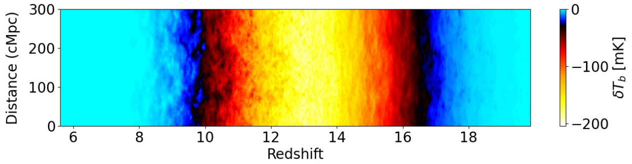
Traditional Pipeline

This work

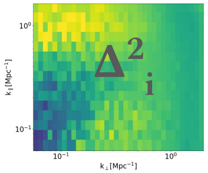
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_\parallel, z)$

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_\parallel, z)$

$\gtrsim$ 300k CPU hrs



$\lesssim$ 1 GPU day

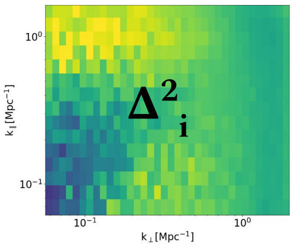


, etc.

Breitman+in prep.

$\mu(\theta) \approx \Delta^2_i$

$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed

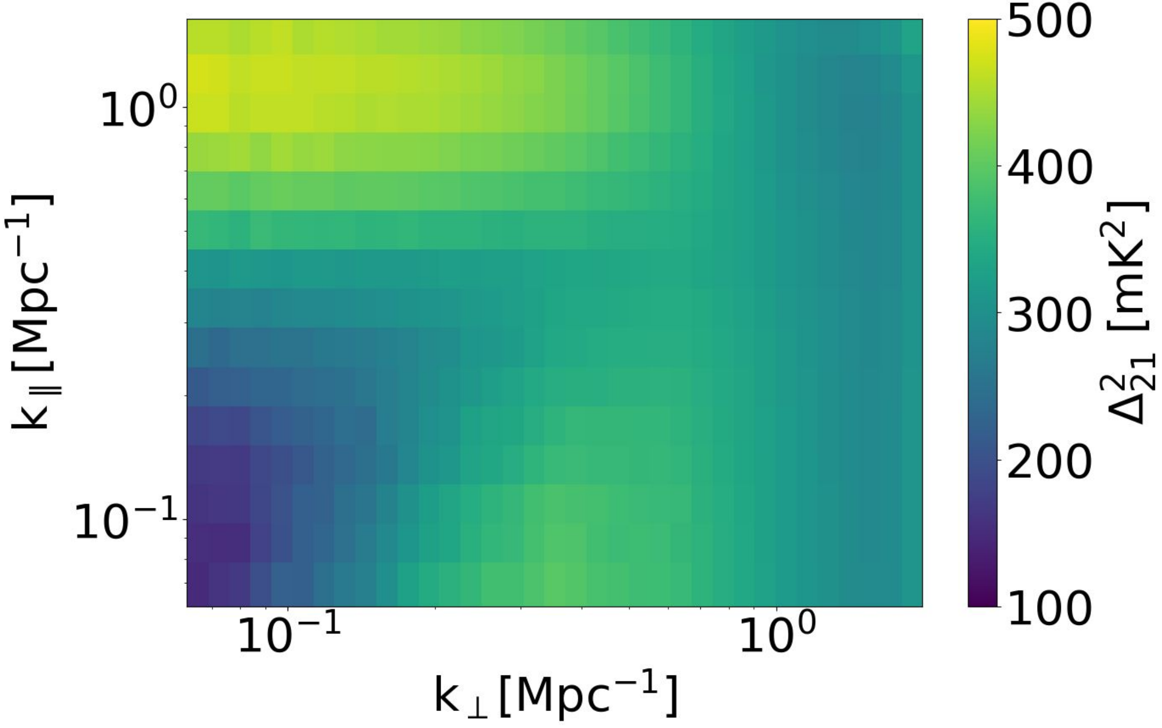


21-cm Power Spectrum

$P(k) \propto \mathcal{F}(T_b)\mathcal{F}^*(T_b)$

$\Delta^2_{21} = P(k)\, k^3/(2\,\pi^2)$ [mK^2]

This is a 3D quantity
1D PS = spherical average
2D PS = cylindrical average

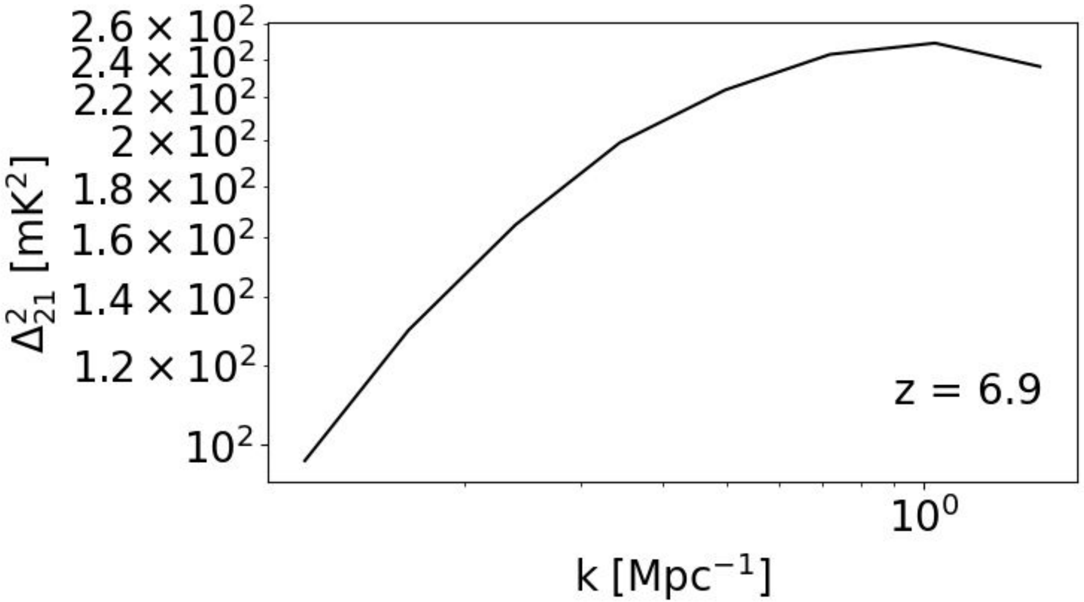


21-cm Power Spectrum

$P(k) \propto \mathcal{F}(T_b)\mathcal{F}^*(T_b)$

$\Delta^2_{21} = P(k)\, k^3/(2\,\pi^2)$ [mK^2]

This is a 3D quantity
1D PS = spherical average
2D PS = cylindrical average

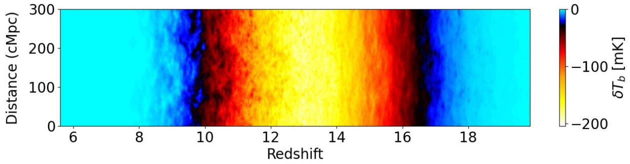


Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

Traditional Pipeline

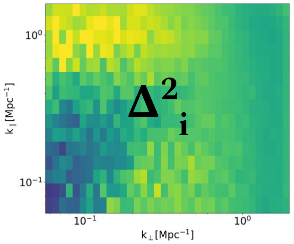
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

≳ 300k CPU hrs



$\mu(\theta) \approx \Delta^2_i$

$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed

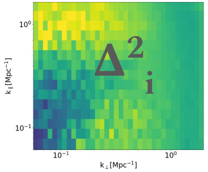


Spherical averaging

This work

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

≲ 1 GPU day



, etc.

Breitman+in prep.

Sample params: $\theta_{astro} \sim P(\theta_{astro}), \theta_{cosmo}$, ICs

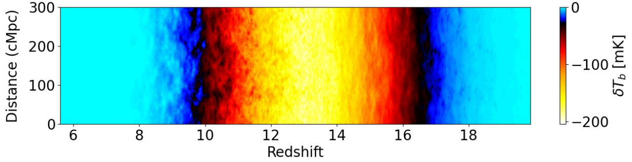
Traditional Pipeline

This work

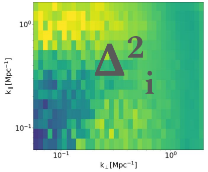
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

≳ 300k CPU hrs



≲ 1 GPU day

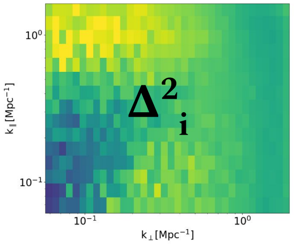


, etc.

Breitman+in prep.

$\mu(\theta) \approx \Delta^2_i$

$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed



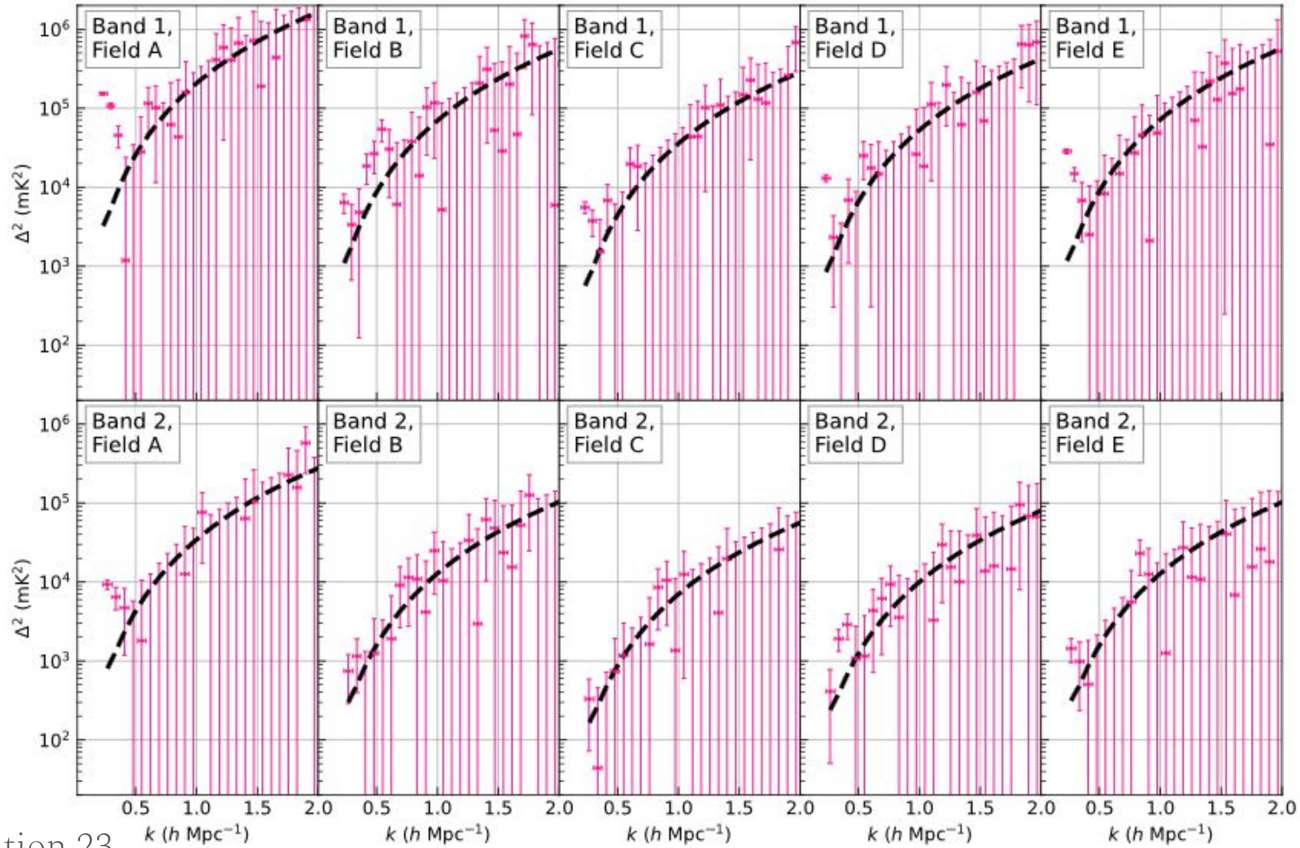
Spherical averaging

Gaussian likelihood: $P(d|\theta)$

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

21-cm Power Spectrum

Δ^2_{21}
obs



HERA Collaboration 23

But are we actually comparing like to like in the forward modelling process?

A few caveats...

- Forward model volume << survey volume

Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

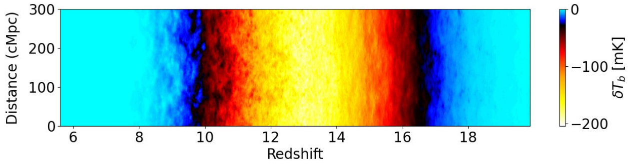
Traditional Pipeline

This work

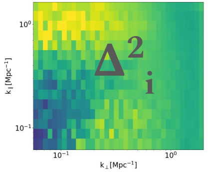
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

≳ 300k CPU hrs



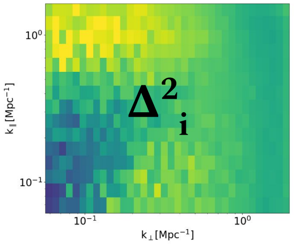
≲ 1 GPU day



, etc.

Breitman+in prep.

$\mu(\theta) \approx \Delta^2_i$
$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed

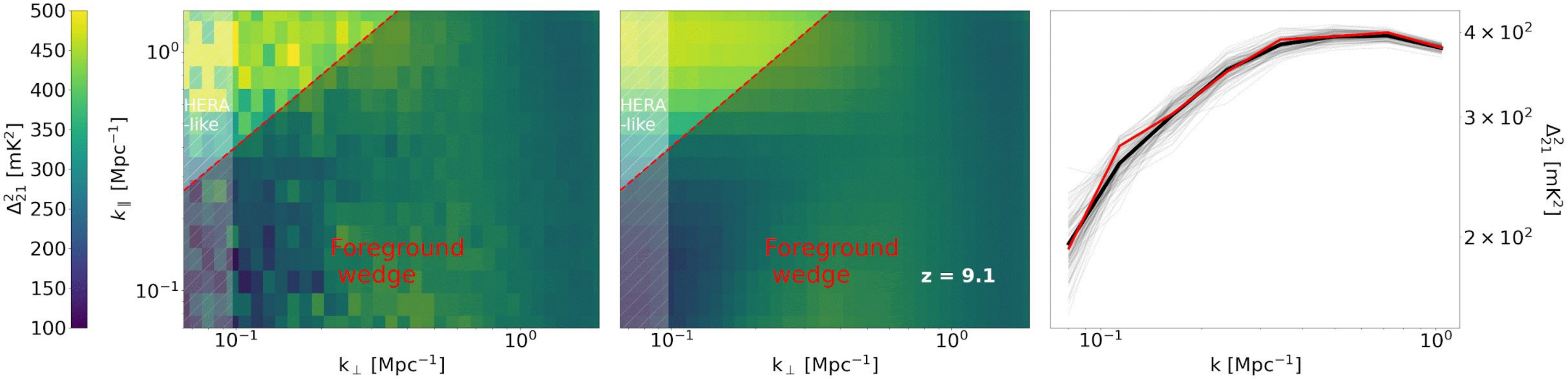


Spherical averaging

Gaussian likelihood: $P(d|\theta)$

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

Sample Variance



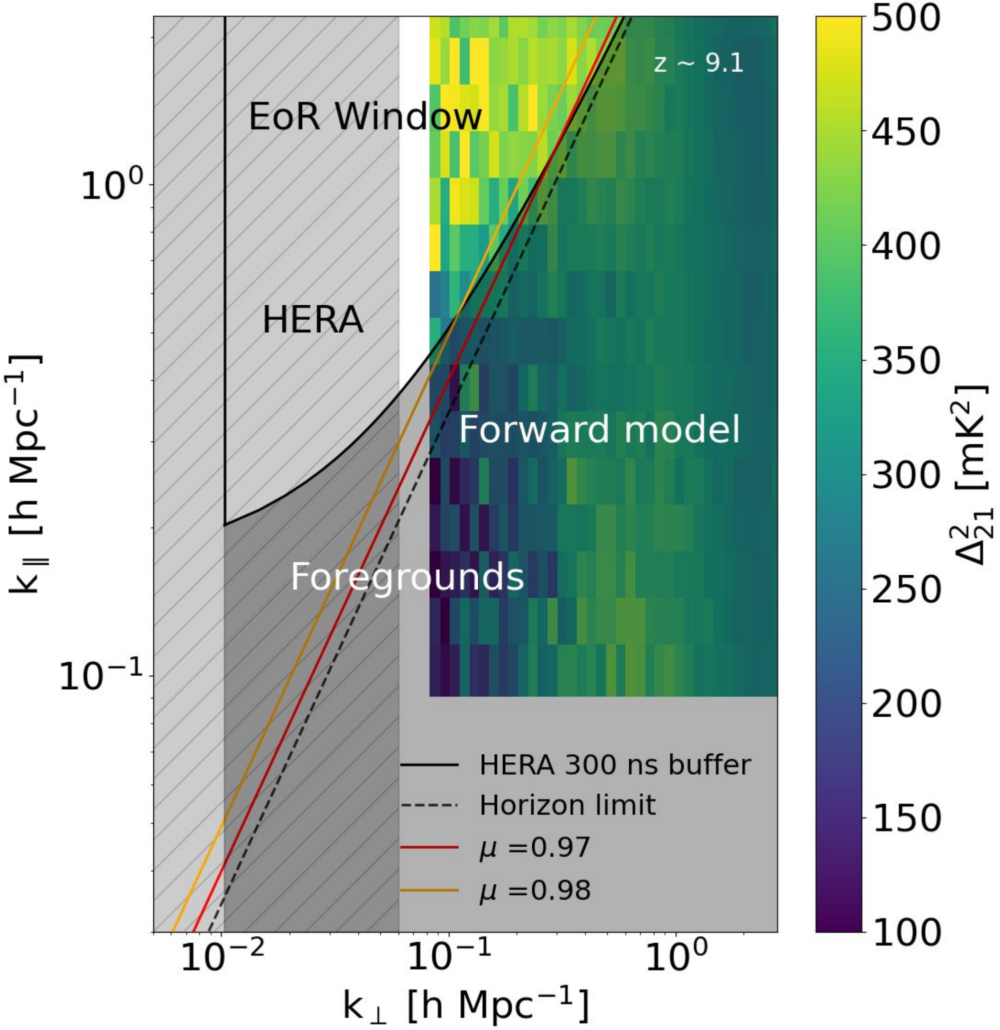
One sample

Mean of ~ 200 samples

Why not increase the model volume?

A typical inference requires >> 100k forward models.

Even with state-of-the-art simulators, it is not practical to forward-model volumes as large as a typical survey in an inference context.



Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

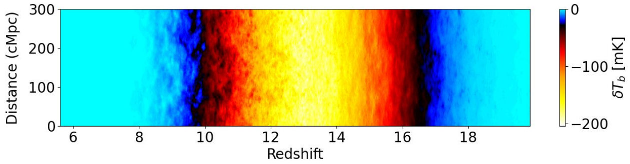
Traditional Pipeline

This work

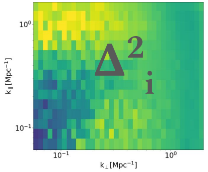
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_\parallel, z)$

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_\parallel, z)$

$\gtrsim$ 300k CPU hrs



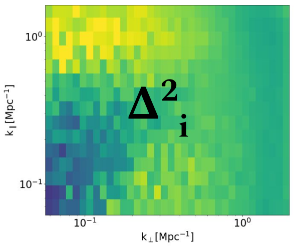
$\lesssim$ 1 GPU day



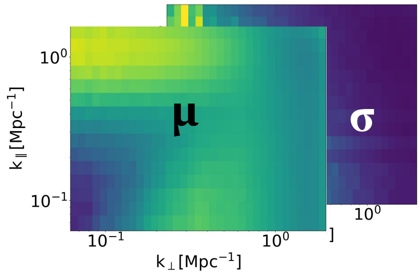
, etc.
Breitman+in prep.

$\mu(\theta) \approx \Delta^2_i$
$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed

$\mu(\theta) \approx NN(\Delta^2_i)$
$\Sigma(\theta) \approx NN(\Delta^2_i)$



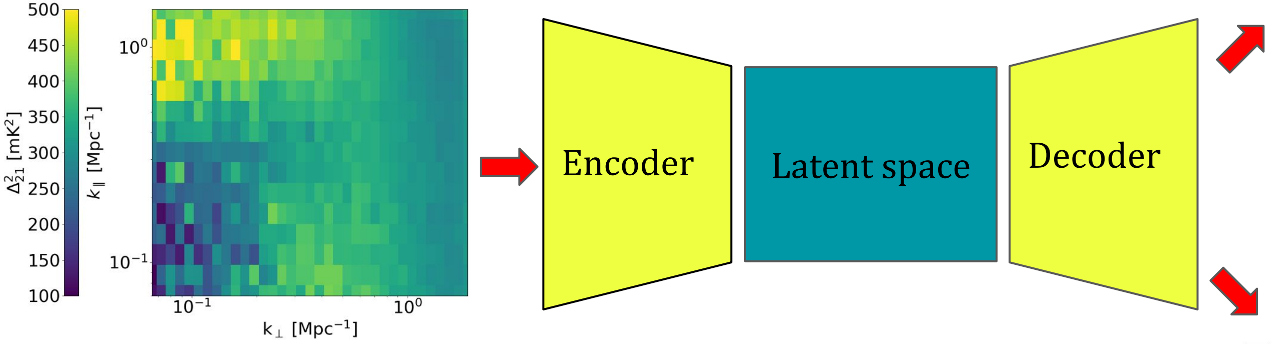
Spherical averaging

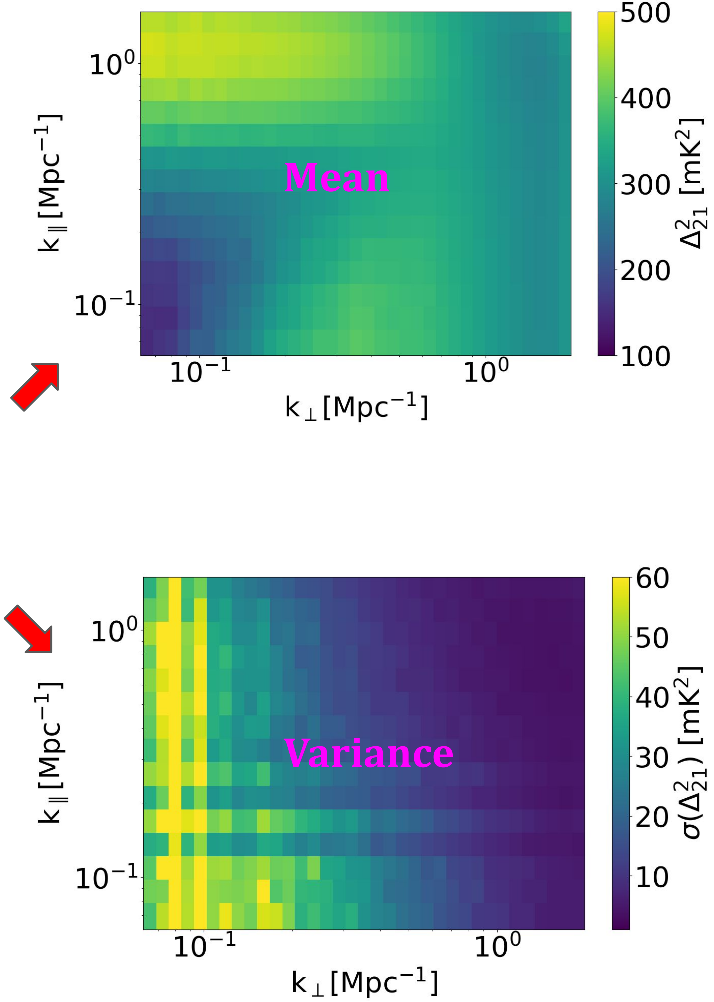


Gaussian likelihood: $P(d|\theta)$

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

Sample Variance Denoiser Network





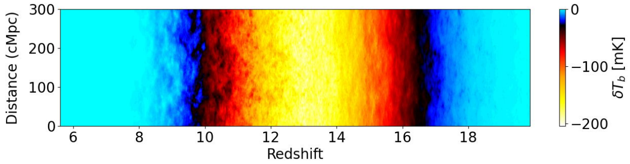
Vary θ_{astro} and ICs

Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

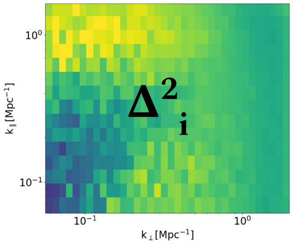
Traditional Pipeline

Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

≳ 300k CPU hrs



$\mu(\theta) \approx \Delta^2_i$
$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed

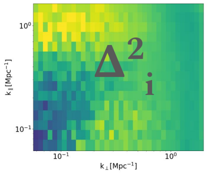


Spherical averaging

This work

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

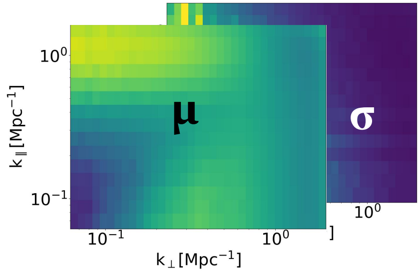
≲ 1 GPU day



, etc.

Breitman+in prep.

$\mu(\theta) \approx NN(\Delta^2_i)$
$\Sigma(\theta) \approx NN(\Delta^2_i)$



Gaussian likelihood: $P(d|\theta)$

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

A few caveats...

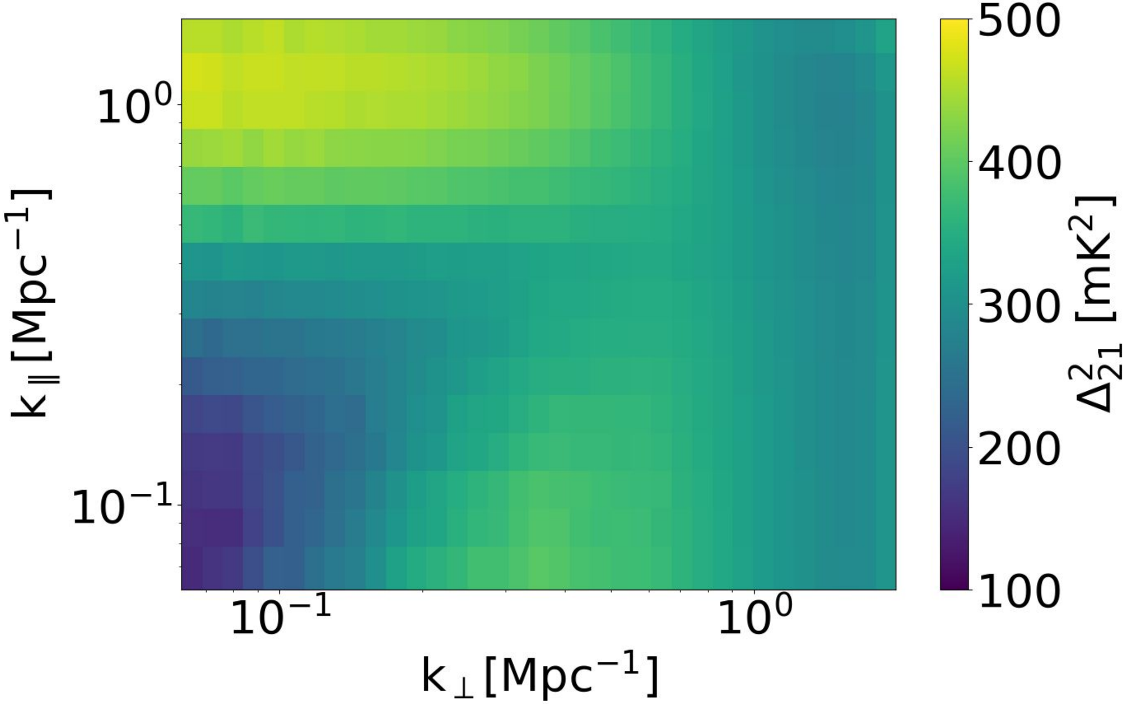
- Forward model volume << survey volume
- Spherically-averaging the (non-isotropic) PS is the same for the forward model and the data

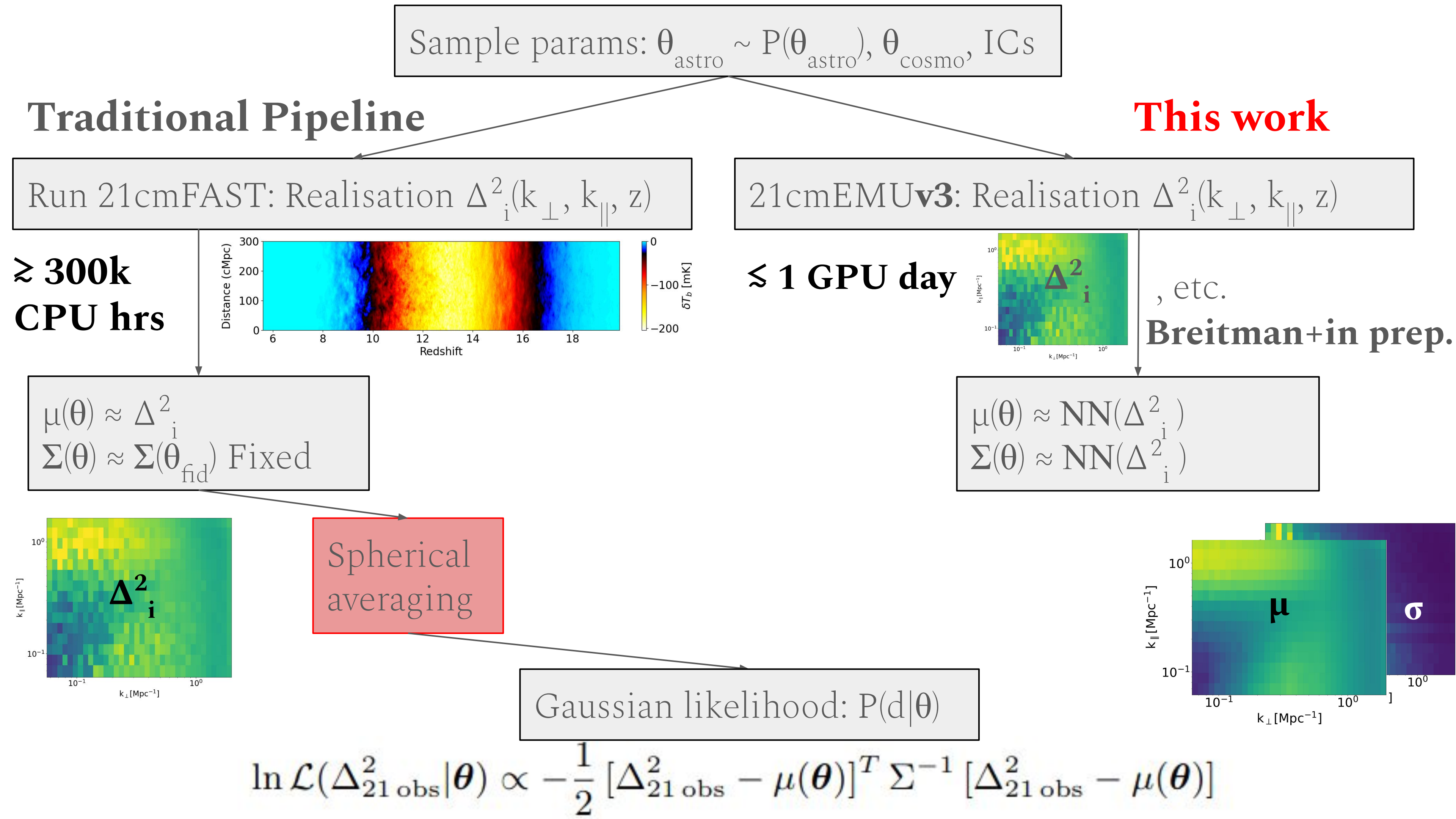
The 21-cm PS is not isotropic

Two main causes:

- Redshift space distortions (RSDs)
- Lightcone evolution along the line of sight

Both contribute to increasing the power along the line of sight.

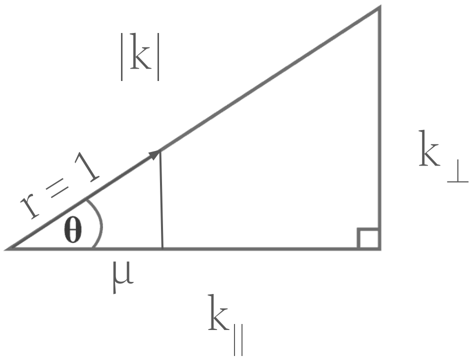


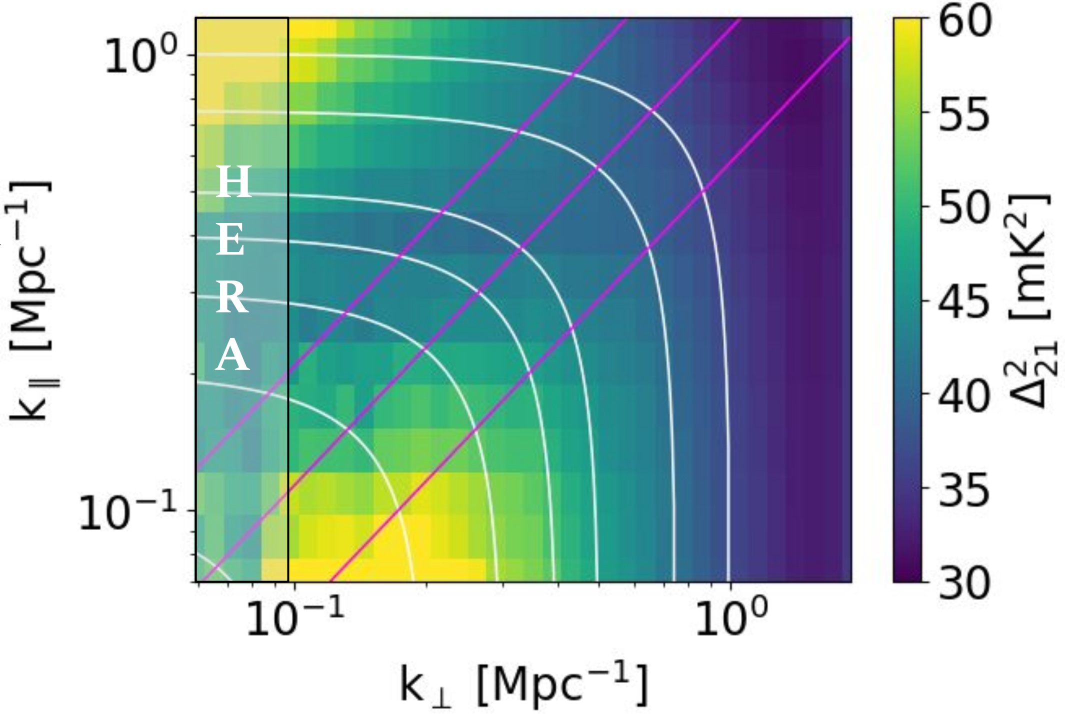


Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs
Traditional Pipeline
This work
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_\parallel, z)$
21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_\parallel, z)$
≳ 300k CPU hrs
Distance (cMpc)
Redshift
δT_b [mK]
≲ 1 GPU day
Δ^2_i
, etc.
Breitman+in prep.
$\mu(\theta) \approx \Delta^2_i$
$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed
$\mu(\theta) \approx NN(\Delta^2_i)$
$\Sigma(\theta) \approx NN(\Delta^2_i)$
Δ^2_i
$k_\parallel$ [Mpc^{-1}]
$k_\perp$ [Mpc^{-1}]
Spherical averaging
μ
σ
Gaussian likelihood: $P(d|\theta)$
$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

Anisotropy of the 21-cm Cylindrical Power Spectrum

- White: Constant |k| contours
- Magenta: Constant $\mu = \cos(\theta)$
  - θ = angle between $k_{\parallel}$ and $k_{\perp}$

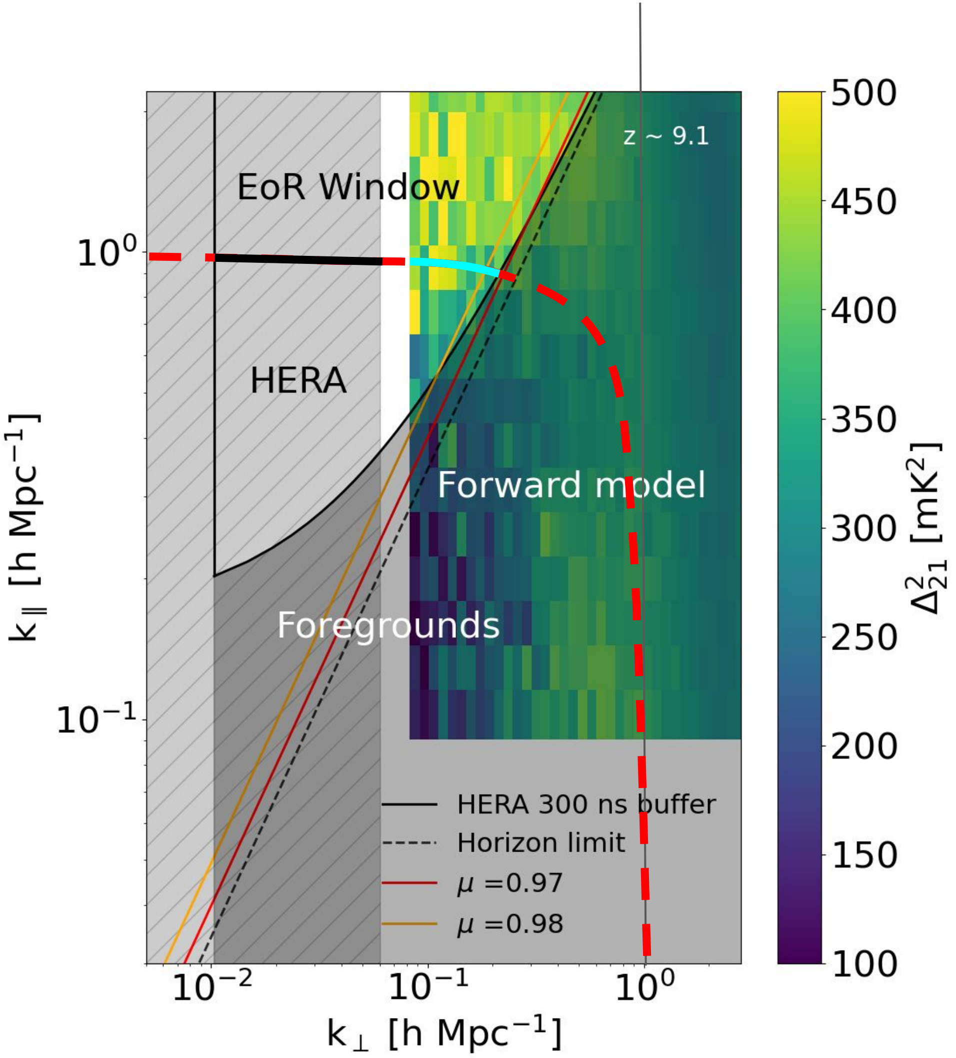




Forward Model

Anisotropy ⇒spherically averaging over the entire power spectrum introduces a bias in the forward model.

Solution: Put a cut μ_{min} = 0.97 on the forward-modelled data to reduce the bias in the spherically-averaged 21-cm PS.



Sample params: $\theta_{astro} \sim P(\theta_{astro})$, θ_{cosmo}, ICs

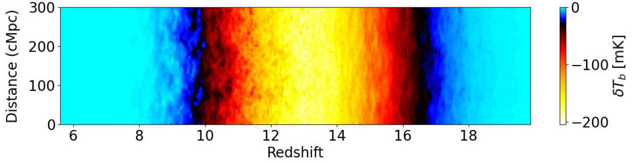
Traditional Pipeline

This work

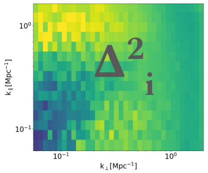
Run 21cmFAST: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

21cmEMU**v3**: Realisation $\Delta^2_i(k_\perp, k_{||}, z)$

≳ 300k CPU hrs



≲ 1 GPU day

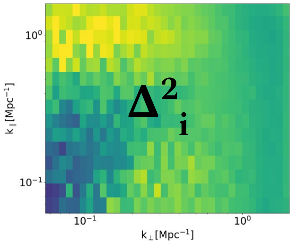


, etc.

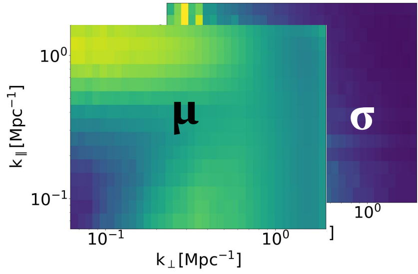
Breitman+in prep.

$\mu(\theta) \approx \Delta^2_i$
$\Sigma(\theta) \approx \Sigma(\theta_{fid})$ Fixed

$\mu(\theta) \approx NN(\Delta^2_i)$
$\Sigma(\theta) \approx NN(\Delta^2_i)$

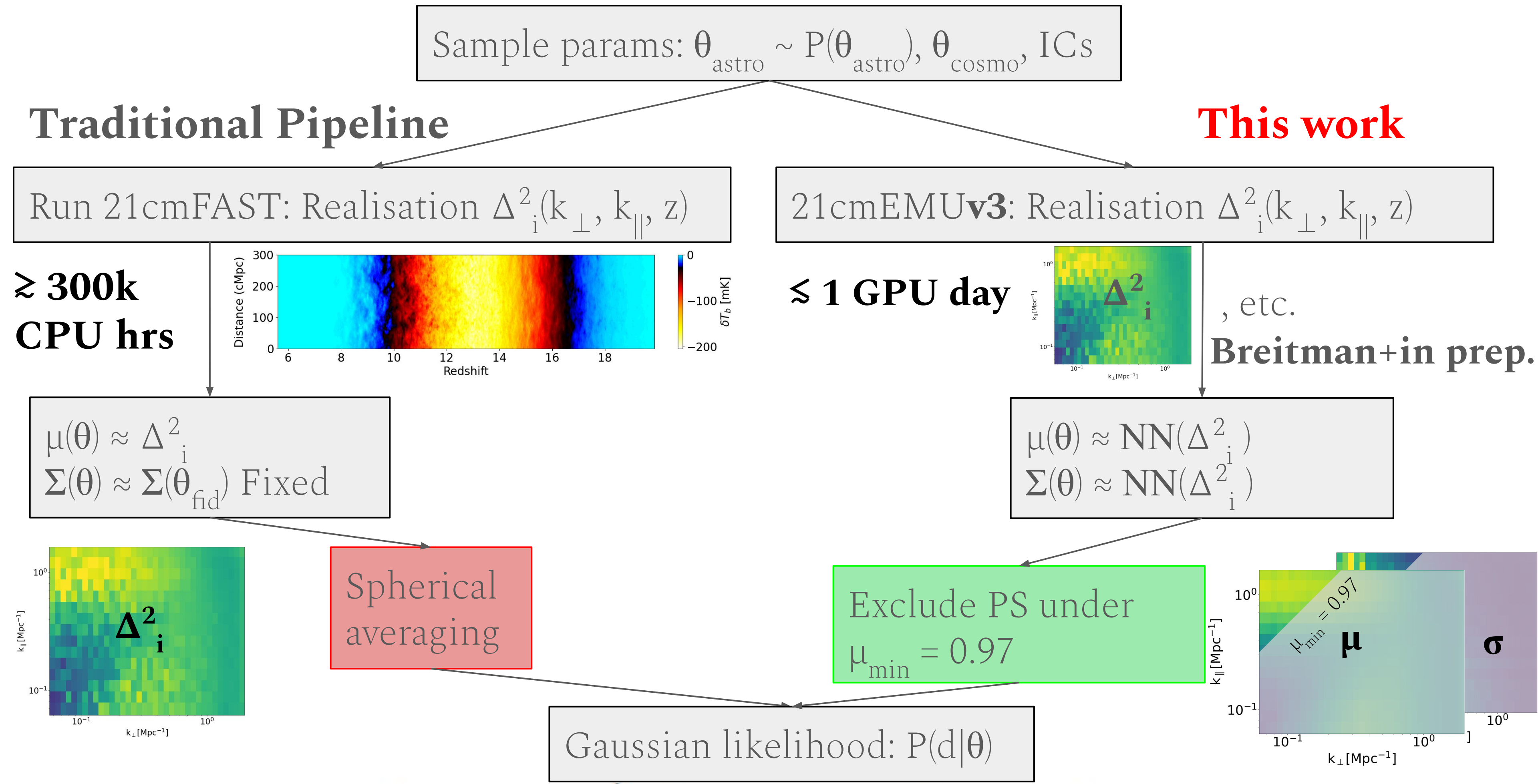


Spherical averaging



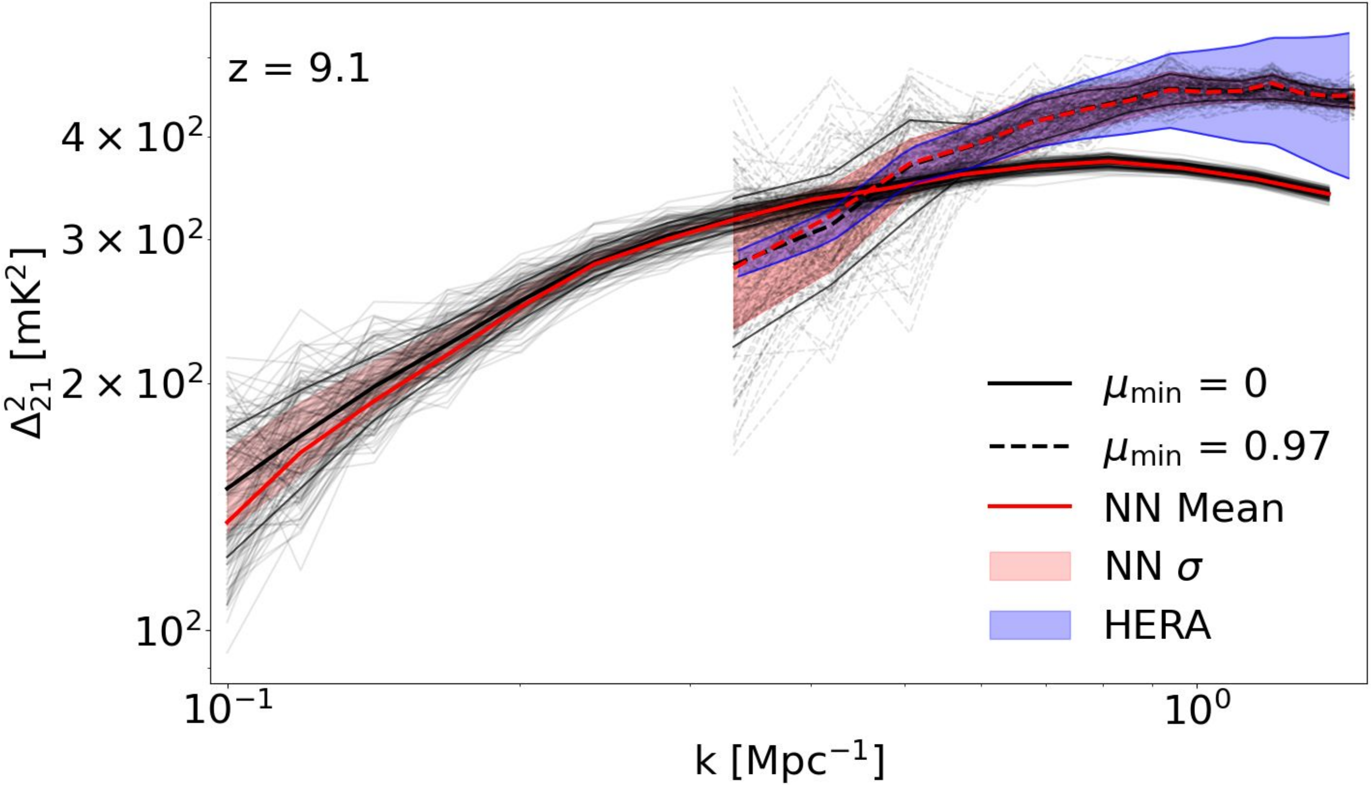
Gaussian likelihood: $P(d|\theta)$

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$



$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

Performance on a single θ_{astro} from the test set



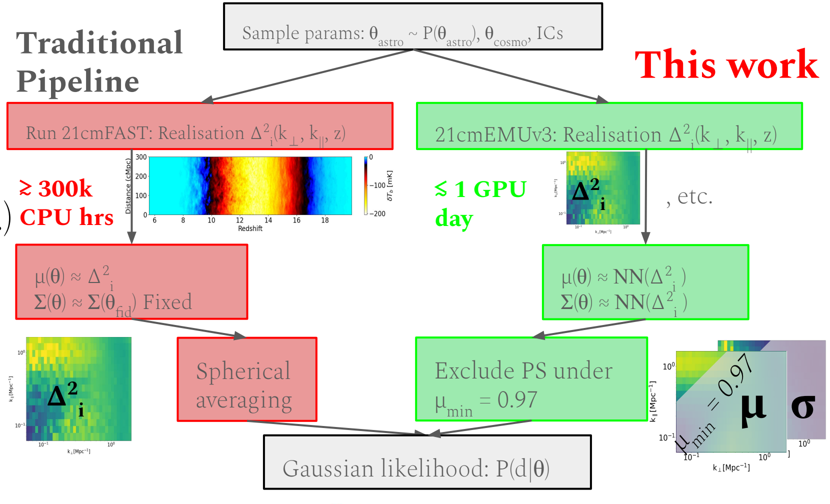
Conclusions

Problems:

- **21cmFAST simulator:** Too slow to perform inferences as new data comes out.
- **Sample variance:** We cannot perform Bayesian inference while forward modelling boxes large enough to cover the survey volume.
- **Anisotropy of the 21-cm PS:** RSDs and lightcone evolution make the 21-cm PS anisotropic. As a consequence, averaging over different regions of PS space than the observation leads to a bias in the model used in the inference ⇒ biased posterior.

Solutions:

- **21cmEMUv3**: Emulate 21-cm summaries (2D PS, etc.)
- Neural network to correct for sample variance:
  - **Provides var(θ) estimate**
  - **> 10^5 times faster than traditional methods**
- Exclude signal above μ_{min} = 0.97, just like the observation, to **reduce the bias due to the anisotropy**.



$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

Now available!!

pip install py21cmemu

https://github.com/21cmfast/21cmemu

Breitman+23