Field level model for HI: mocks and applications

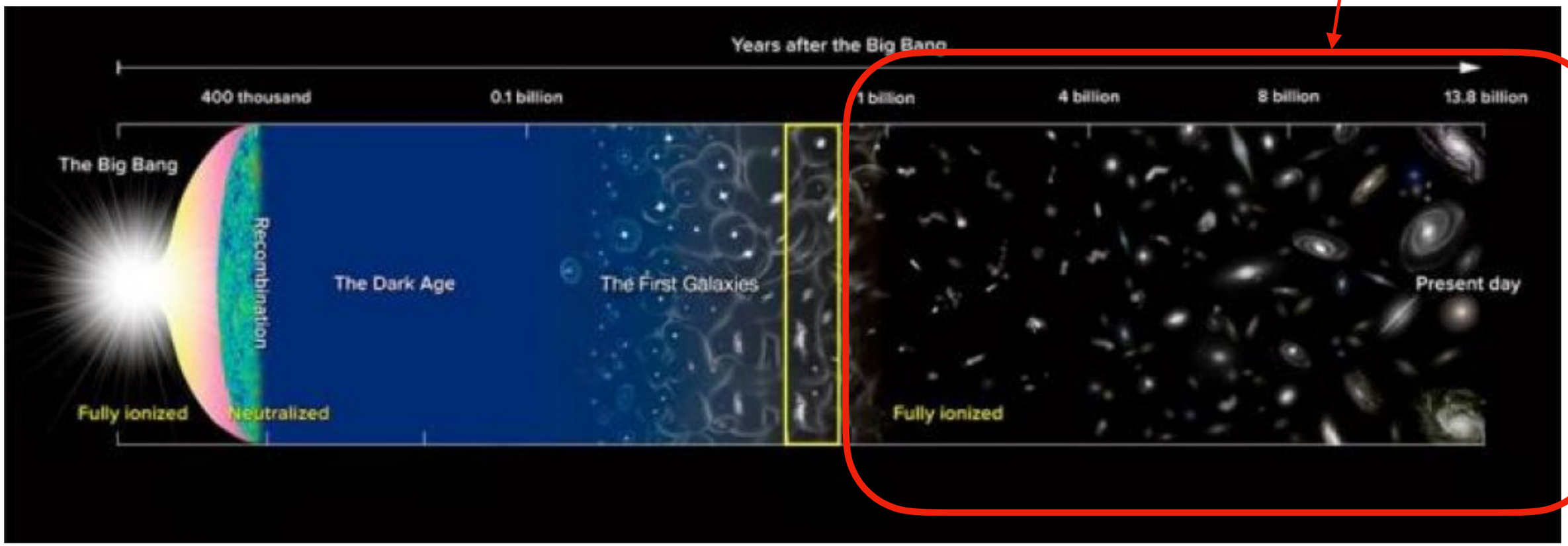
Andrej Obuljen (University of Zurich)

arXiv: 2207.12398

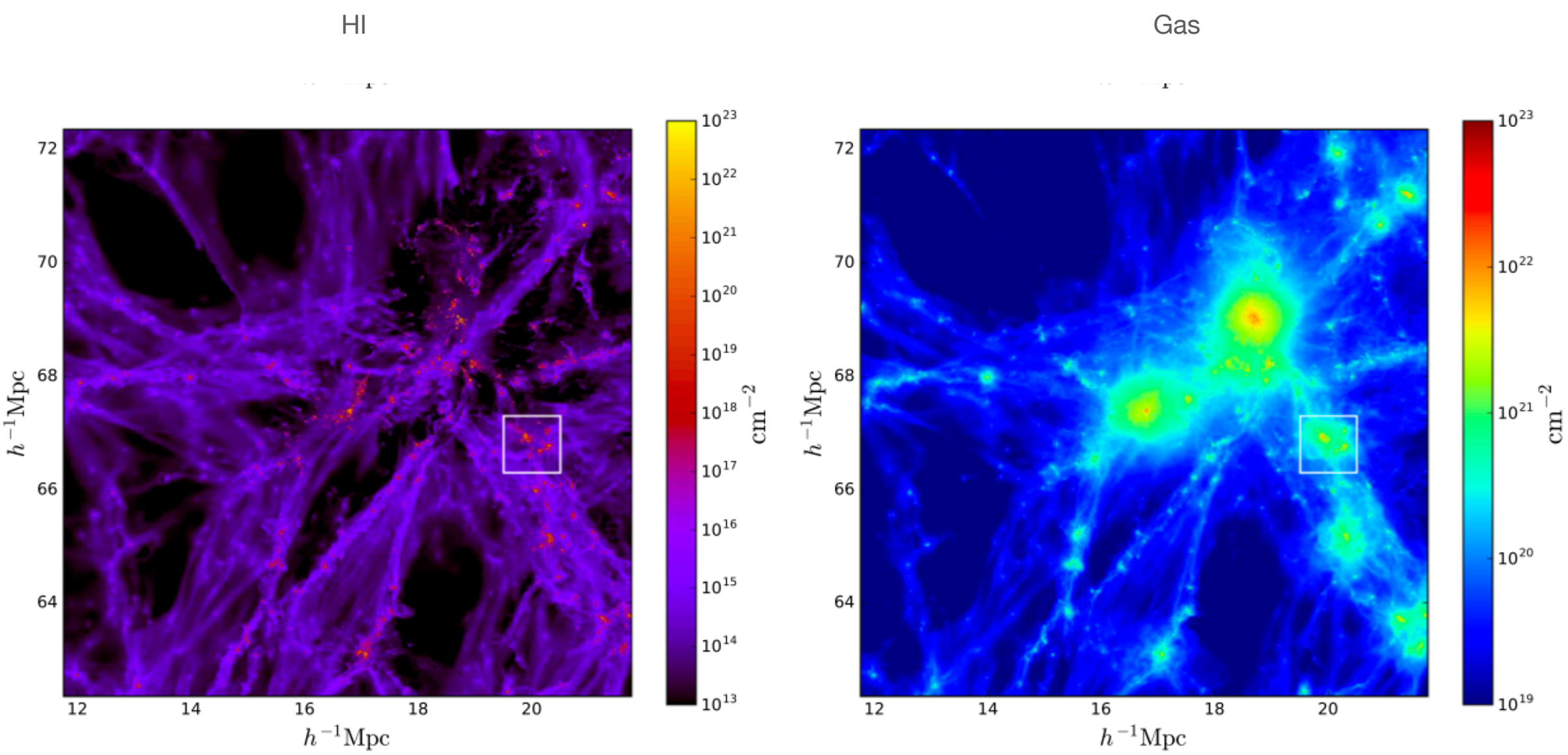
SKA days Zurich
2023

Where is neutral hydrogen (HI)?

We focus on post-reionization

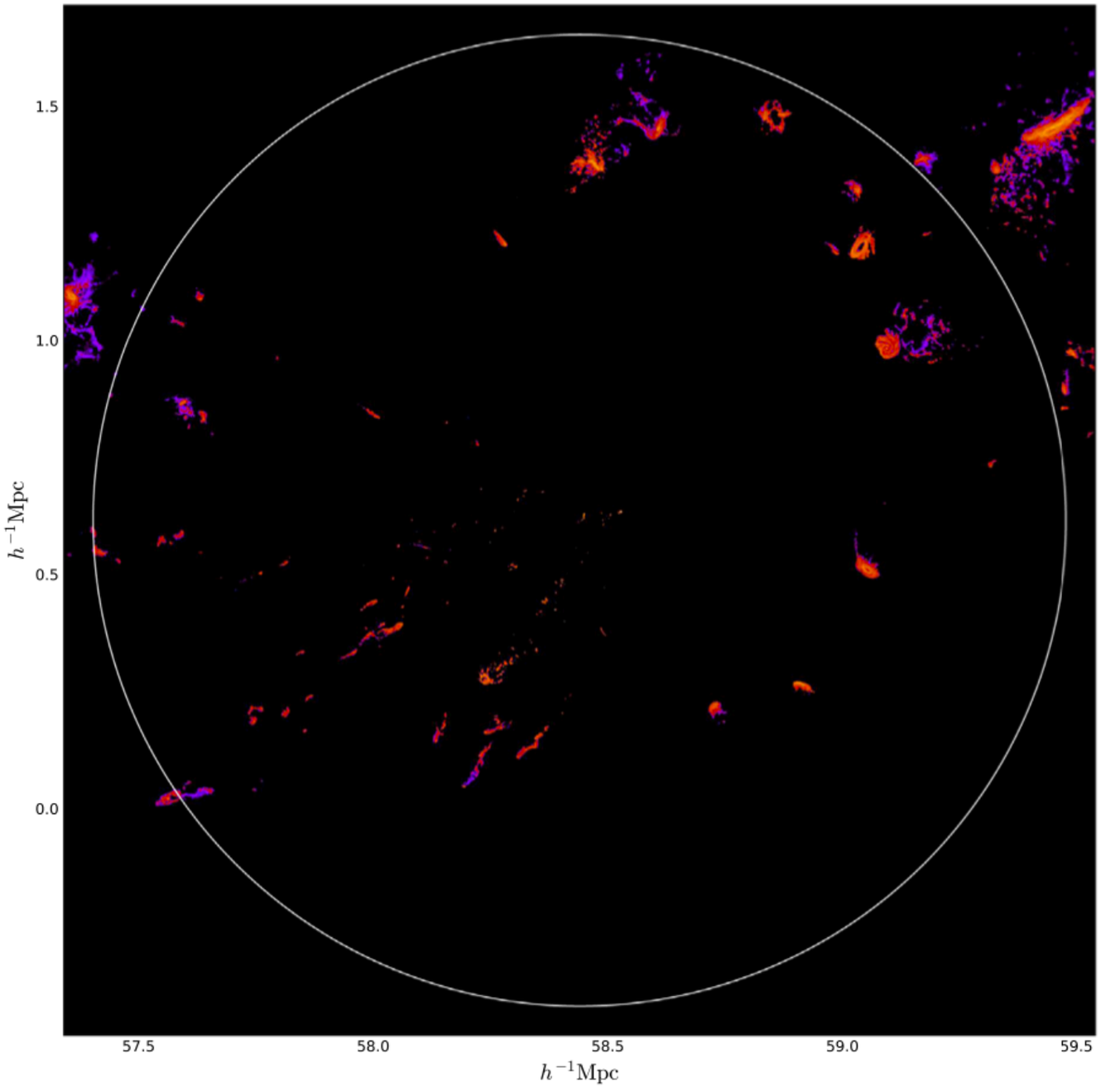


Where is neutral hydrogen (HI)?

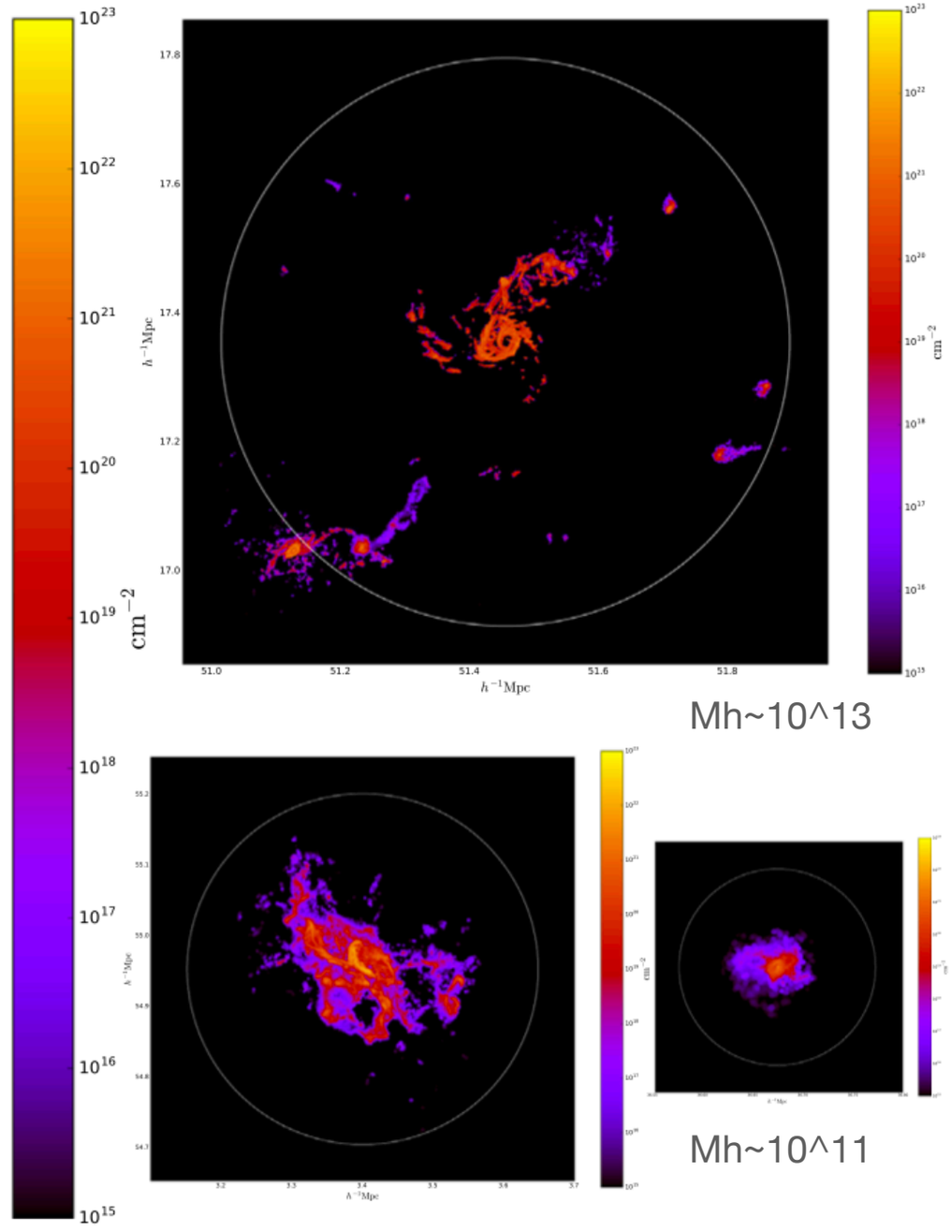


IllustrisTNG (Villaescusa-Navarro+18)

Where is neutral hydrogen (HI)?



Mh~10^14



Mh~10^13

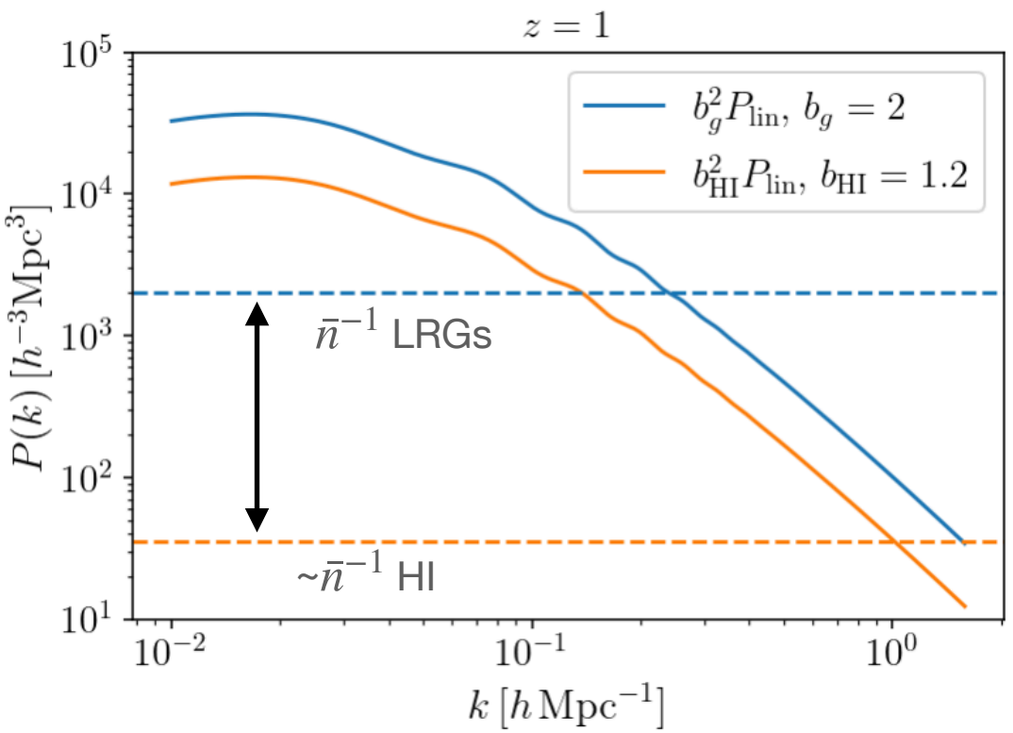
Mh~10^12

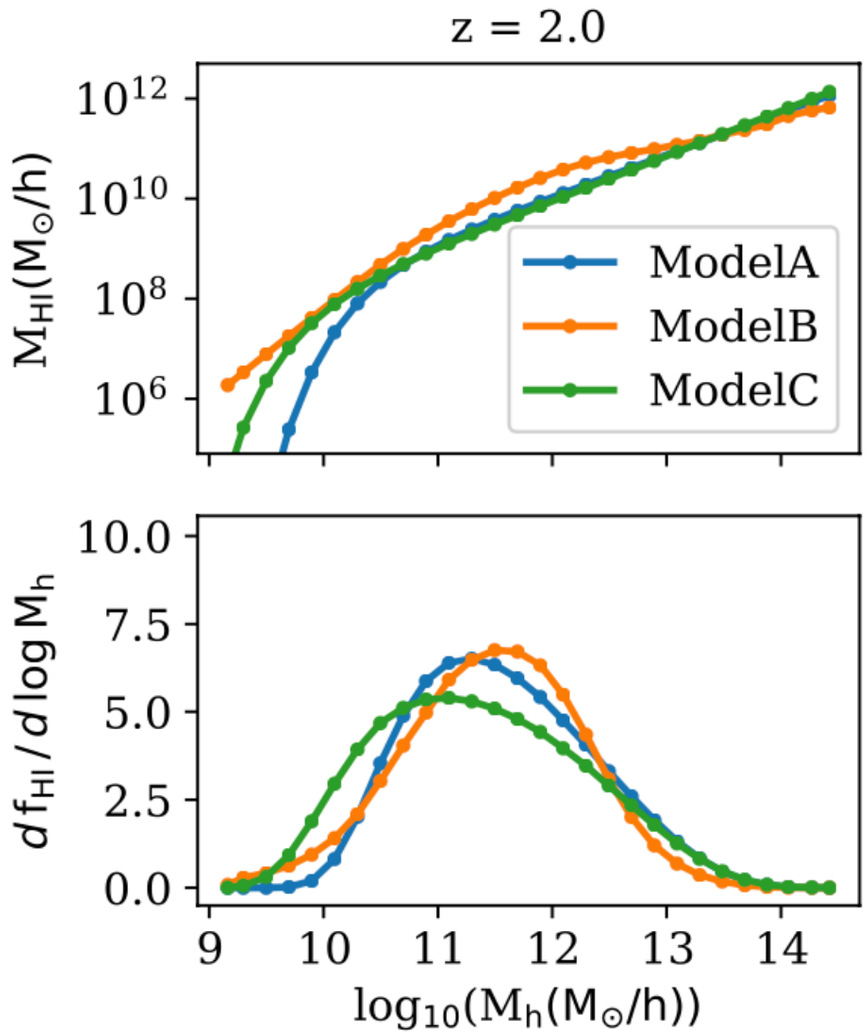
Mh~10^11

IllustrisTNG (Villaescusa-Navarro+18)

Why is HI different/interesting?

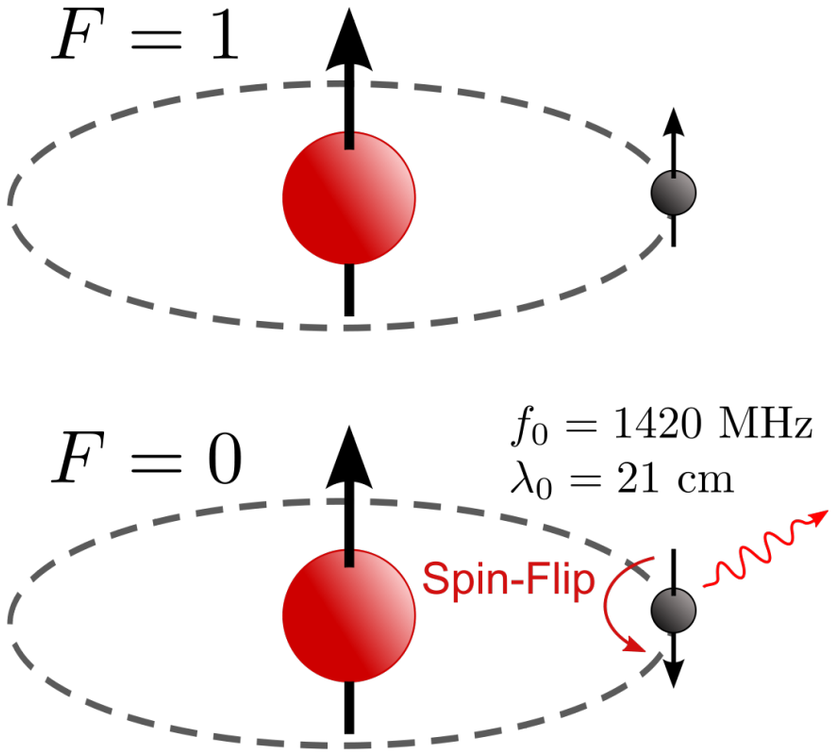
- HI signal from moderate halo masses
- Low shot-noise!
- Moderate redshift space effects
- Shot-noise limiting factor for galaxy clustering (Kobayashi+22)

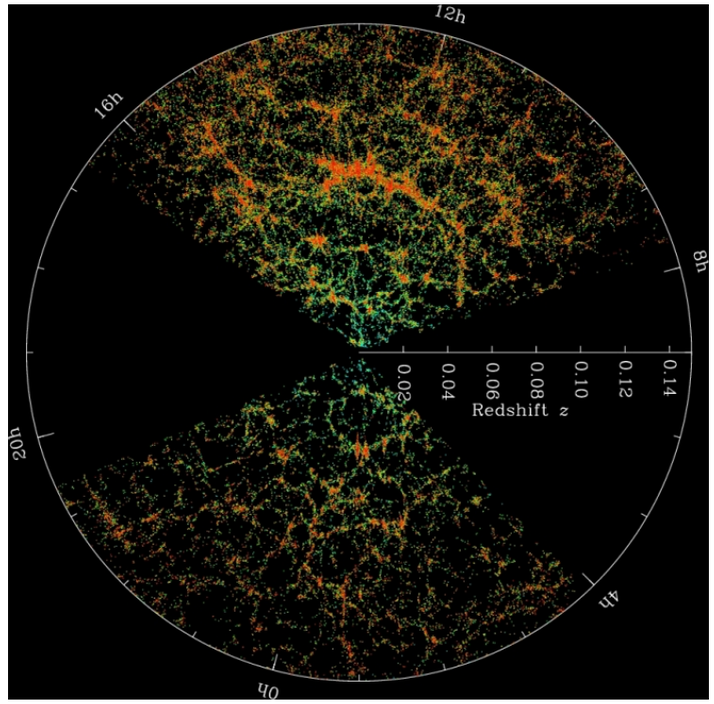




Modi+19

How to observe HI?



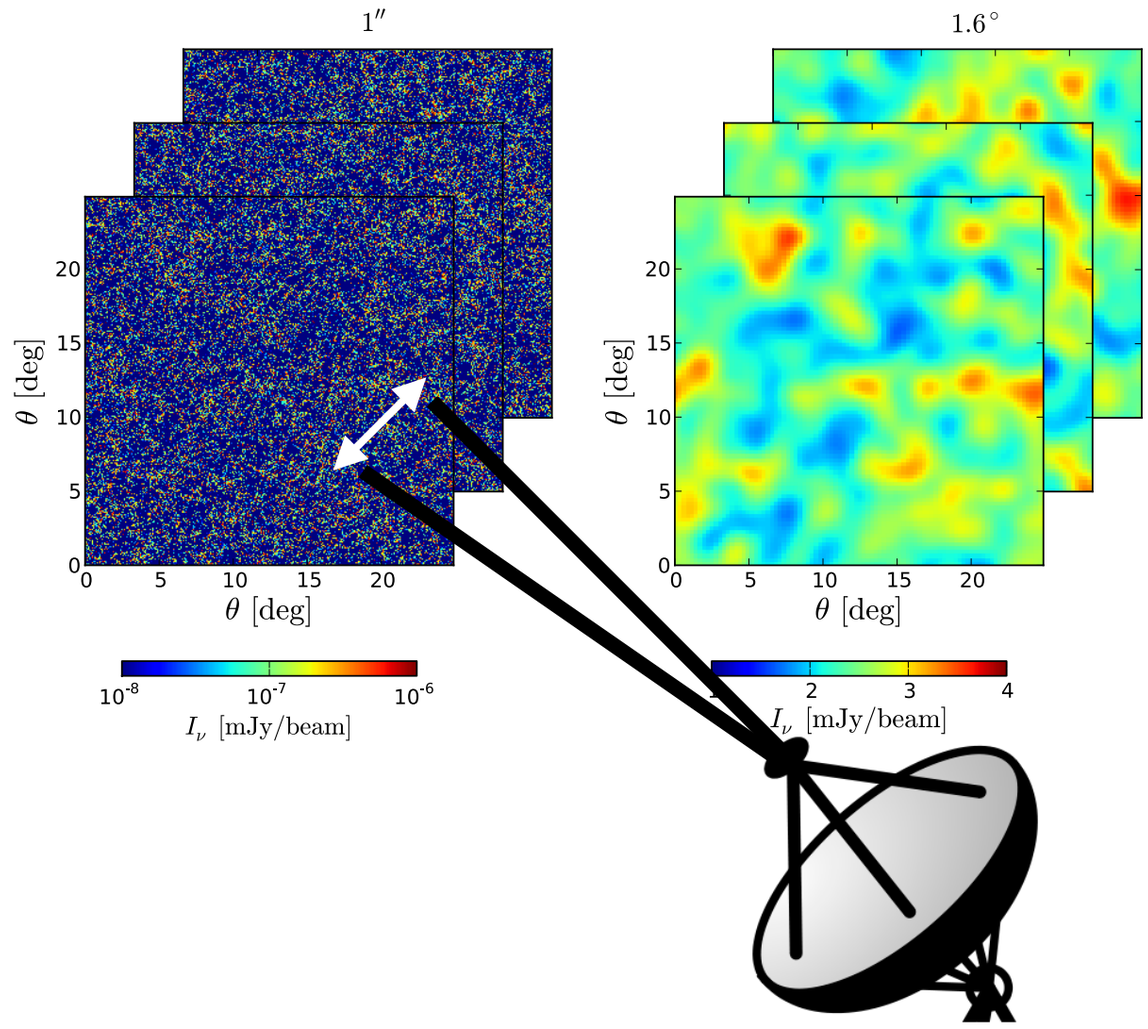


Galaxy redshift survey

Useful to map HI distribution across redshifts

$$\lambda_{\rm obs}(z) = \lambda_0(1+z)$$

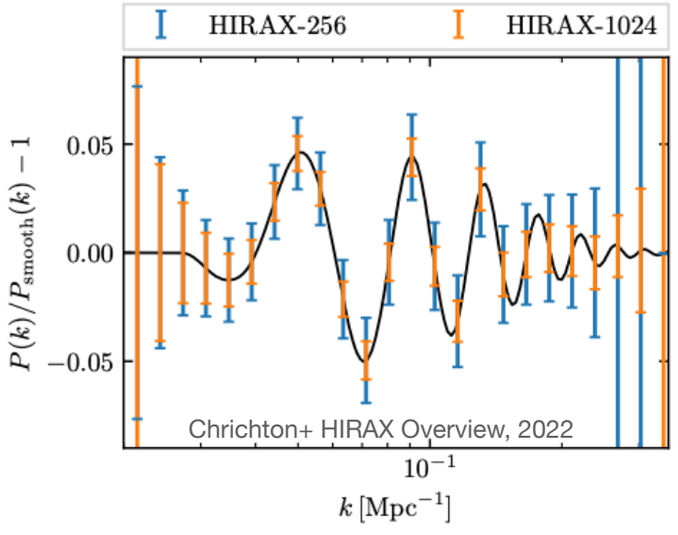
Weak transition, give up resolving galaxies



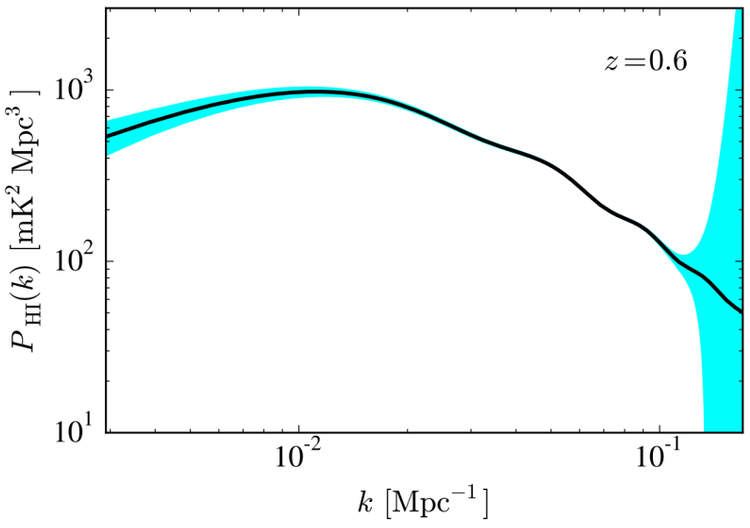
21cm Intensity Mapping (IM) surveys

Current $z \lesssim 2$

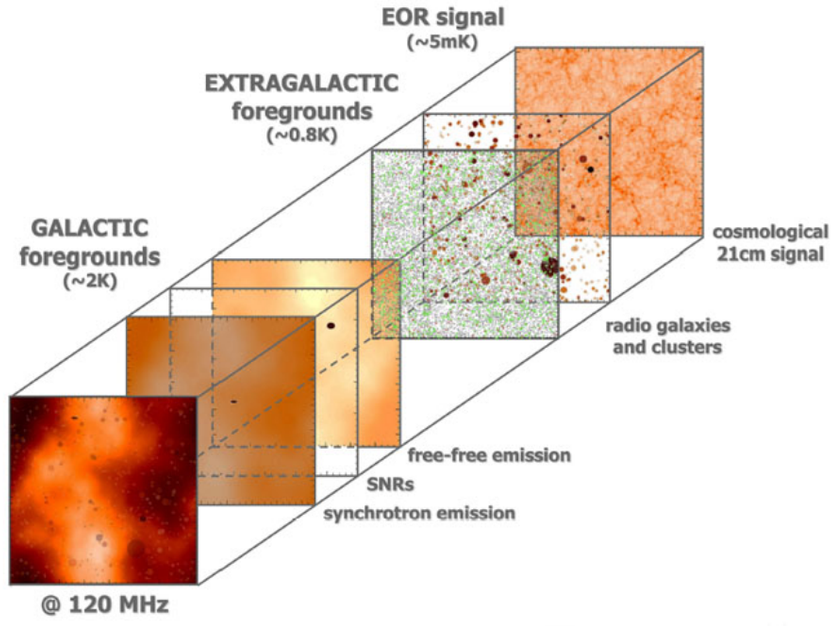
CHIME



HIRAX



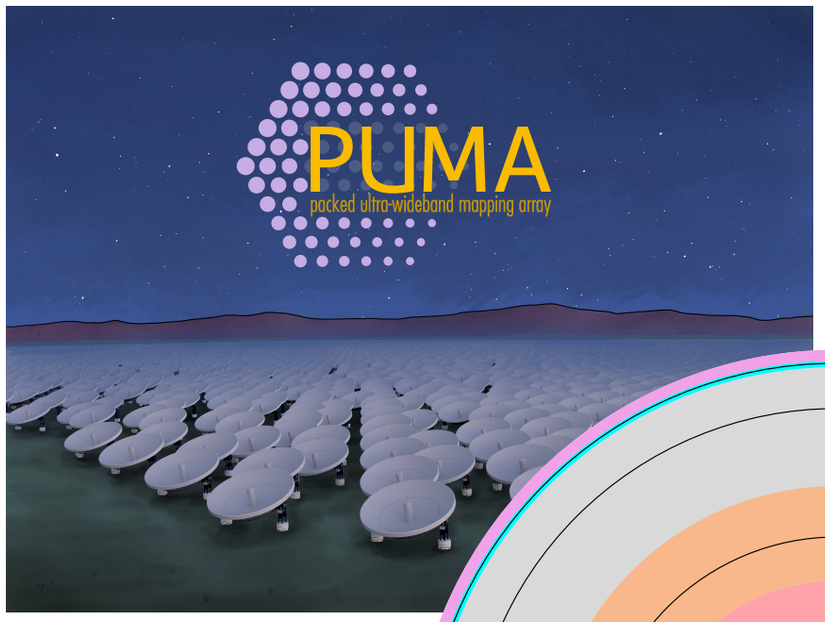
SKAO

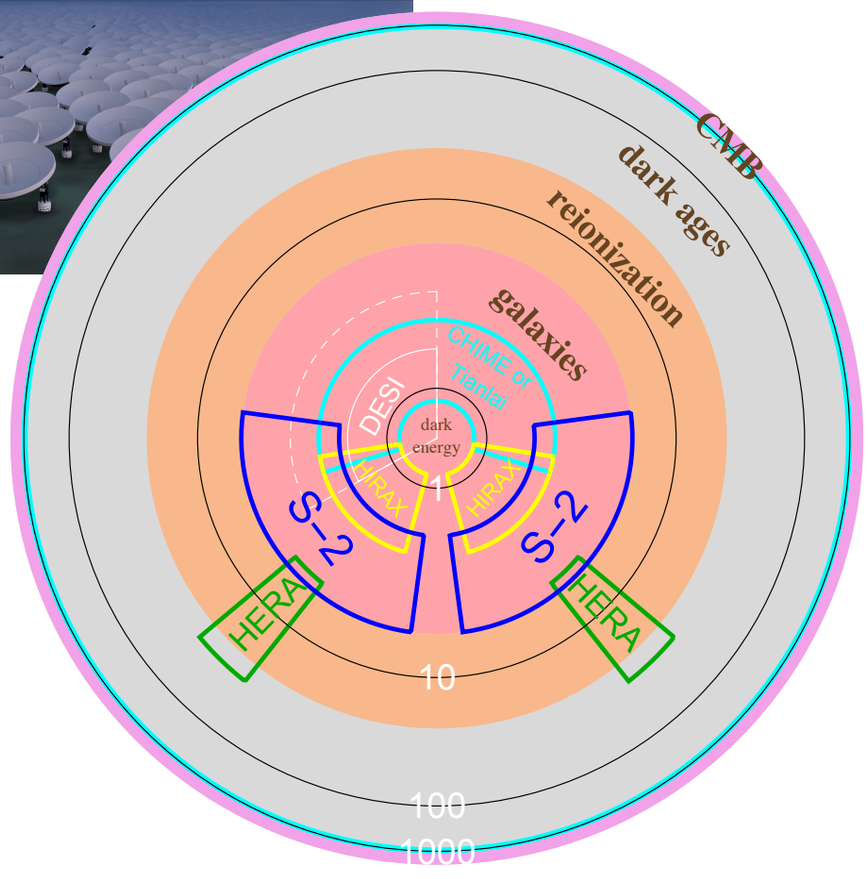


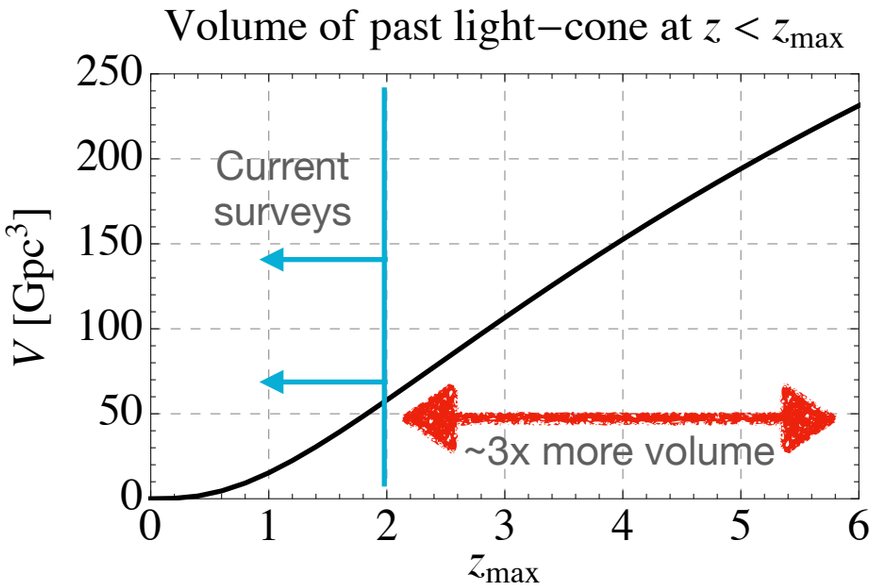
Foregrounds

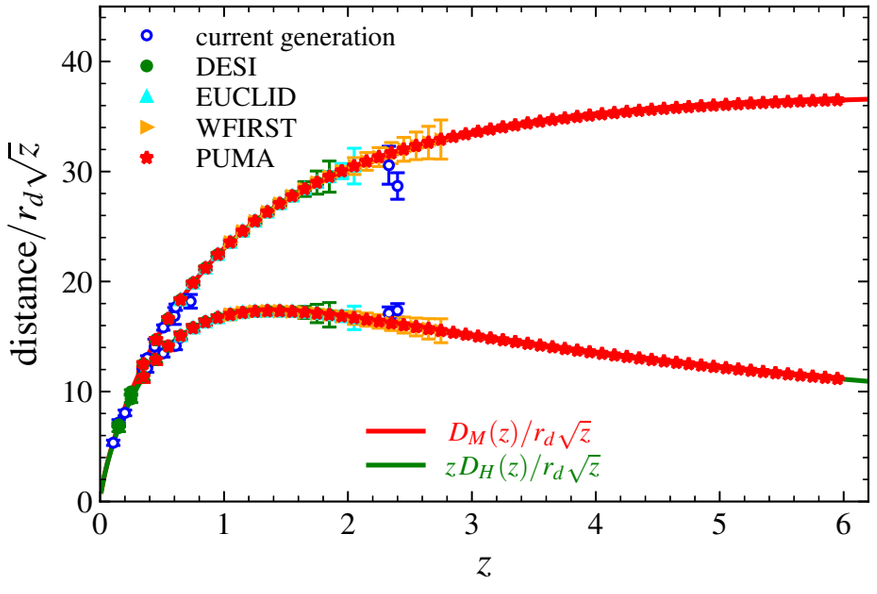
Future 21cm IM surveys

PUMA $0.3 < z < 6$









Figures from Ansari+2018

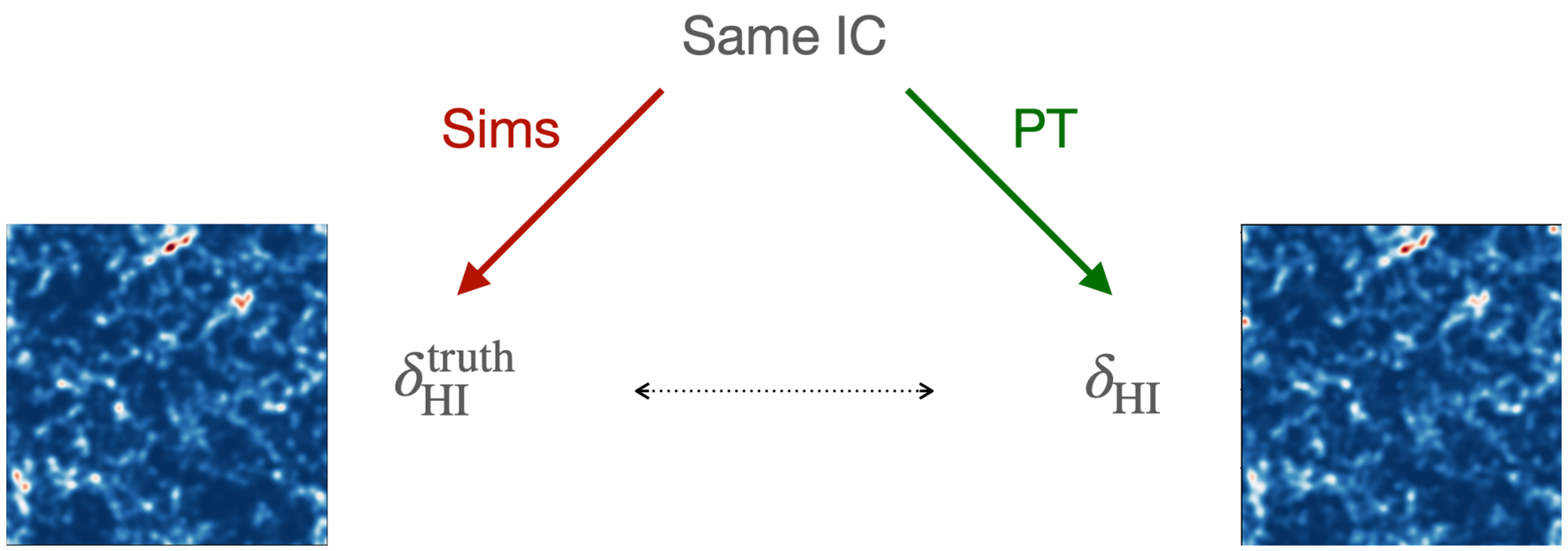
Motivation

- 21cm IM surveys mainly sensitive to perturbative scales
- We can model HI at the field level using PT+bias models (Schmittfull+18)
- HI noise properties –> reconsider cosmological analysis
- Generate fast and accurate 3D mocks: **Hi-Fi mocks**
- Application of Hi-Fi mocks

Field level approach

Advantages

- Pixel-by-pixel agreement → agreement of all summary statistics
- No overfitting
- Easy to isolate and study noise
- No cosmic variance for same IC, no need for large hydro sims.



What is the PT model?

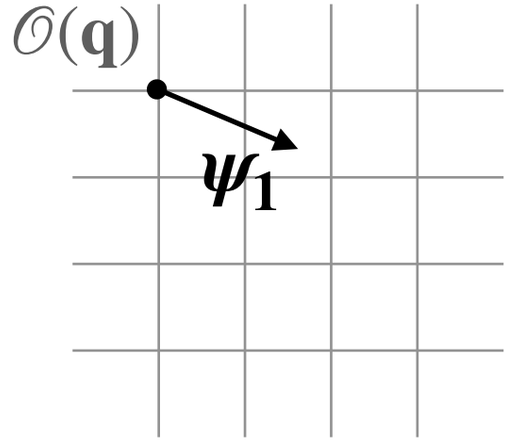
Hybrid Lagrangian and Eulerian scheme, bulk flows included, only linear fields

$$\delta_{\rm HI}(\mathbf{k}) = \int d^3\mathbf{q}(1 + b_1^L\delta_1 + b_2^L(\delta_1^2 - \sigma_1^2) + b_{\mathscr{G}_2}^L\mathscr{G}_2 + \ldots - i\mathbf{k}\cdot\boldsymbol{\psi_2} + \ldots)e^{-i\mathbf{k}(\mathbf{q}+\boldsymbol{\psi}_1)}$$

Zel'dovich displacement

Define *shifted bias operators* in Eulerian space:

$$\tilde{\mathscr{O}}(\mathbf{k}) = \int d^3\mathbf{q}\mathscr{O}(\mathbf{q})e^{-i\mathbf{k}(\mathbf{q}+\boldsymbol{\psi}_1(\mathbf{q}))},$$



where $\mathscr{O} \in \{1, \delta_1, \delta_2, \mathscr{G}_2, \delta_3, \ldots\}$

$$\delta_{\rm HI}(\mathbf{k}) = \beta_1(k)\tilde{\delta}_1(\mathbf{k}) + \beta_2(k)\tilde{\delta}_2^{\perp}(\mathbf{k}) + \beta_{\mathscr{G}_2}(k)\tilde{\mathscr{G}}_2^{\perp}(\mathbf{k}) + \ldots + \text{noise}$$

Transfer functions

Matches 1-loop EFT & CLPT power spectrum

Measure of success

How to compare model & truth at the field level?

- Minimise mean-squared difference/residuals:

$$P_{\rm err}(k) \equiv \left\langle \, | \, \delta_{\rm HI}^{\rm truth}(\mathbf{k}) - \delta_{\rm HI}^{\rm model}(\mathbf{k}) \, |^2 \right\rangle$$

by doing least squares in each k-bin to obtain best-fit transfer functions:

$$\beta_i = \langle \mathcal{O}_i^{\perp} \delta_{\rm HI}^{\rm truth*} \rangle / \langle \, | \, \mathcal{O}_i^{\perp} \, |^2 \rangle$$

- Example:

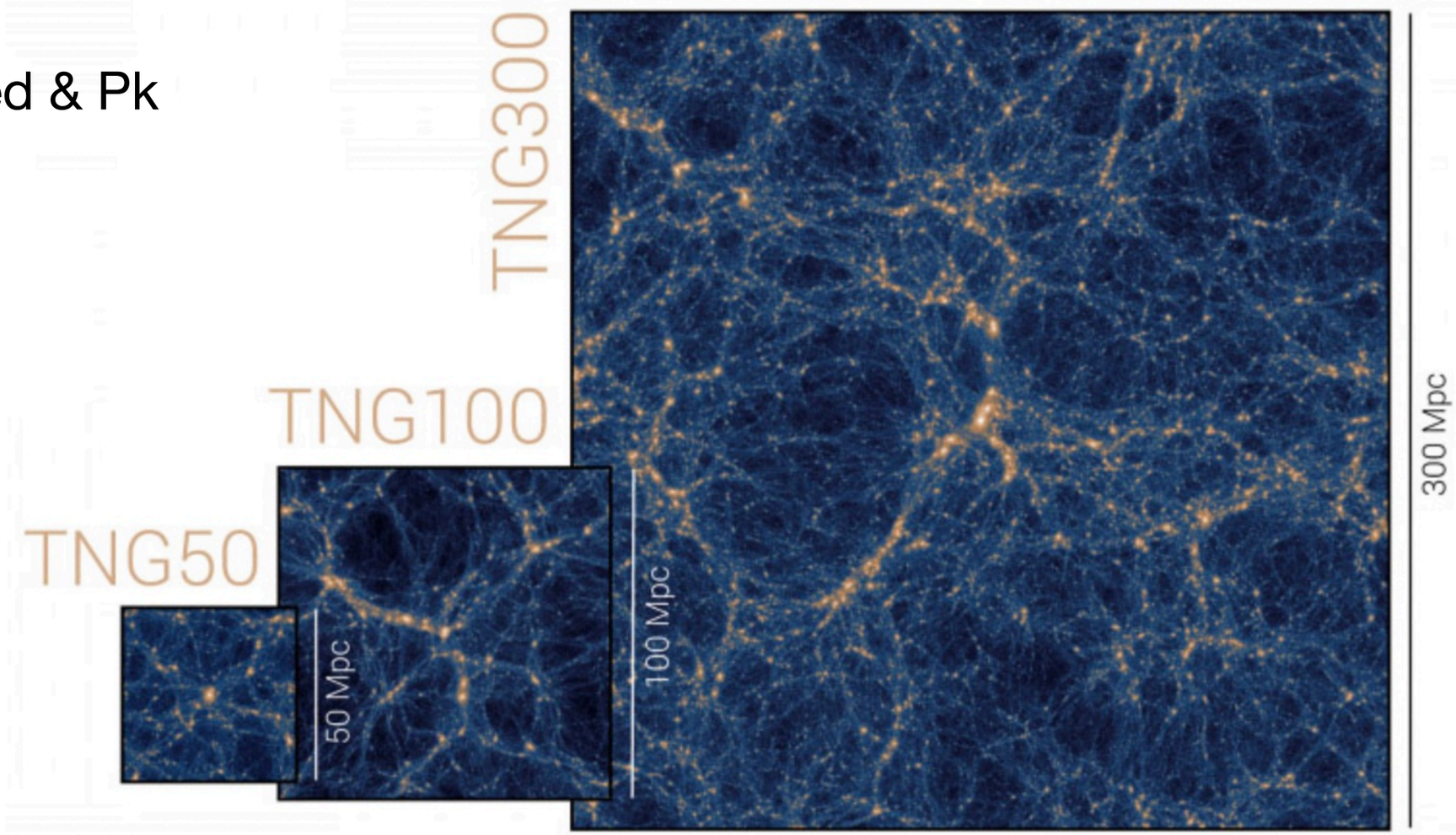
$$\delta_{\rm HI} = b_1 \delta + \epsilon$$

$$b_1(k) = \langle \delta_{\rm HI}^{\rm truth} \delta* \rangle / \langle \, | \, \delta \, |^2 \rangle$$

$$P_{\rm err}(k) = \left\langle \, | \, \epsilon \, |^2 \right\rangle$$

HI from IllustrisTNG

- We apply this approach to a full hydro simulation
- TNG300-1 ($L = 205\,h^{-1}\mathrm{Mpc}$)
- HI in post-processing (Villaescusa+18)
- Same IC: random seed & Pk

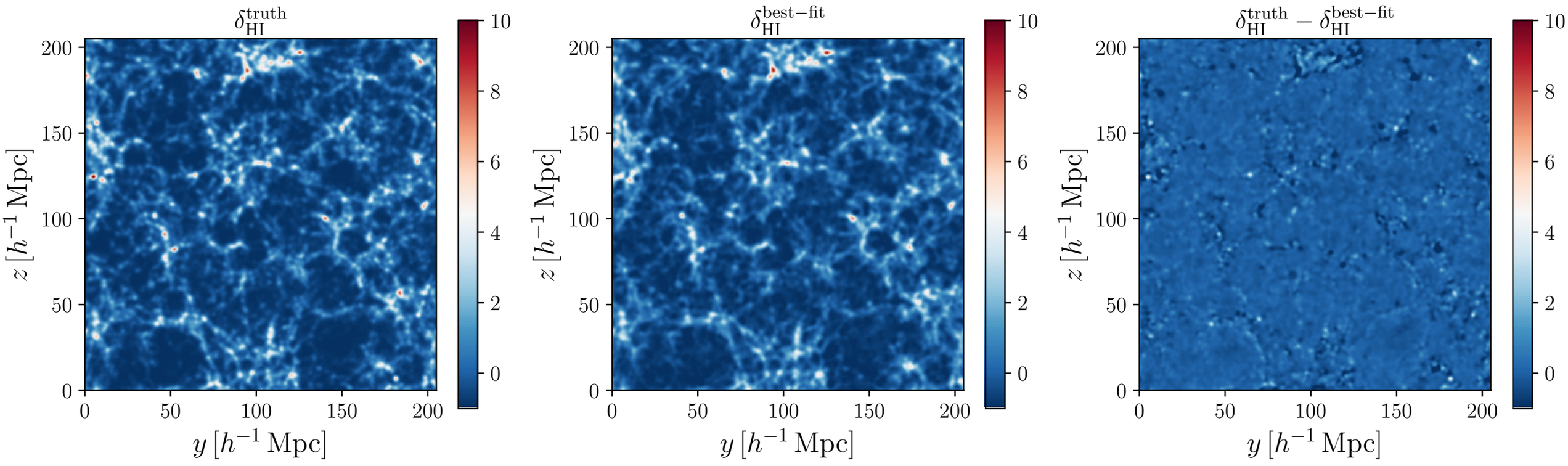


TNG Collaboration

Results – 3D fields

$z = 1$, real space

$$\delta_{\rm HI}(\mathbf{k}) = \beta_1(k)\tilde{\delta}_1(\mathbf{k}) + \beta_2(k)\tilde{\delta}_2^{\perp}(\mathbf{k}) + \beta_{\mathscr{G}_2}(k)\tilde{\mathscr{G}}_2^{\perp}(\mathbf{k}) + \beta_3(k)\tilde{\delta}_3^{\perp}(\mathbf{k}) + \ldots + \text{noise}$$

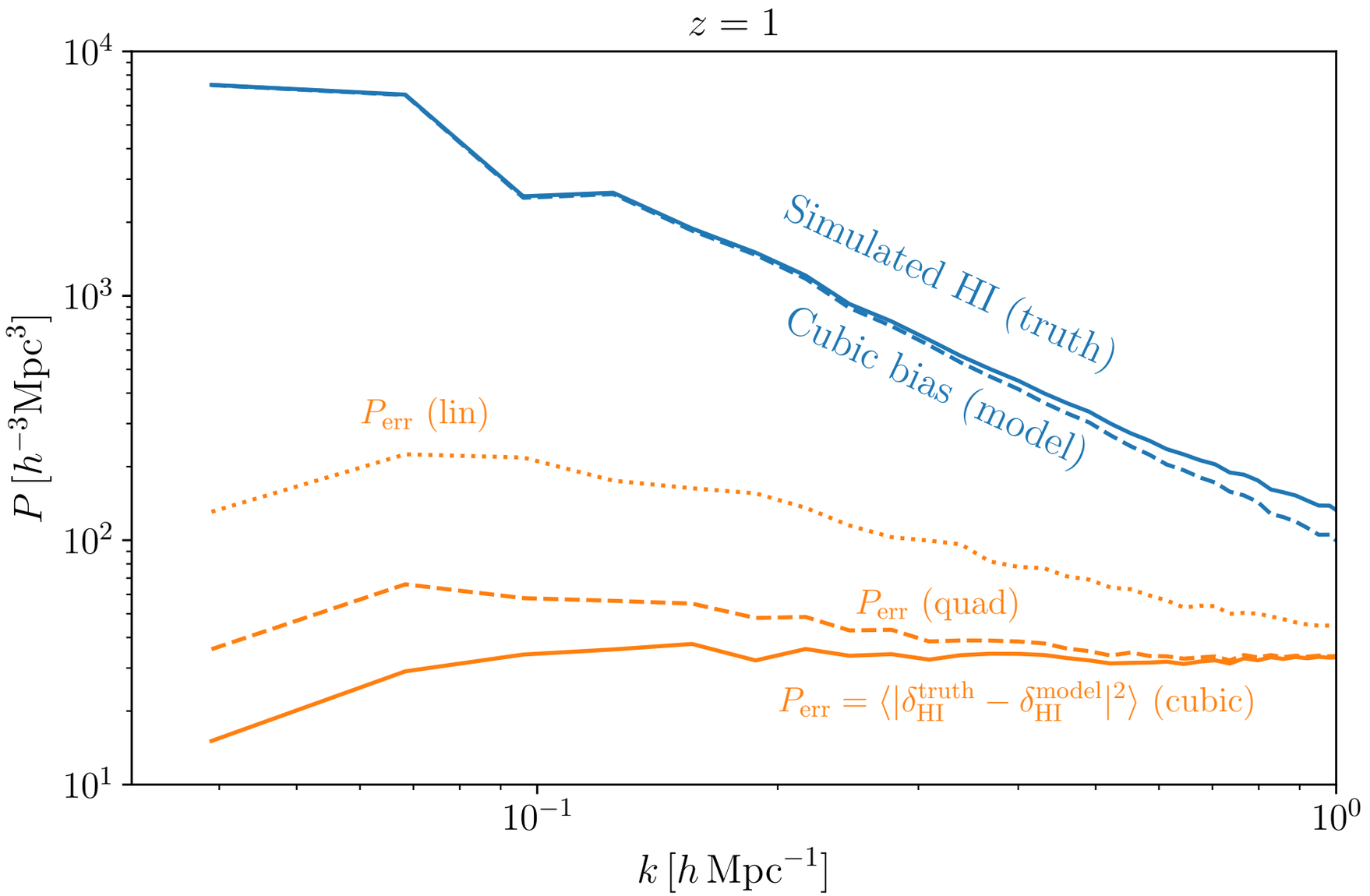


Slice depth 20 Mpc/h, smoothed 1 Mpc/h Gaussian

Results – power spectrum

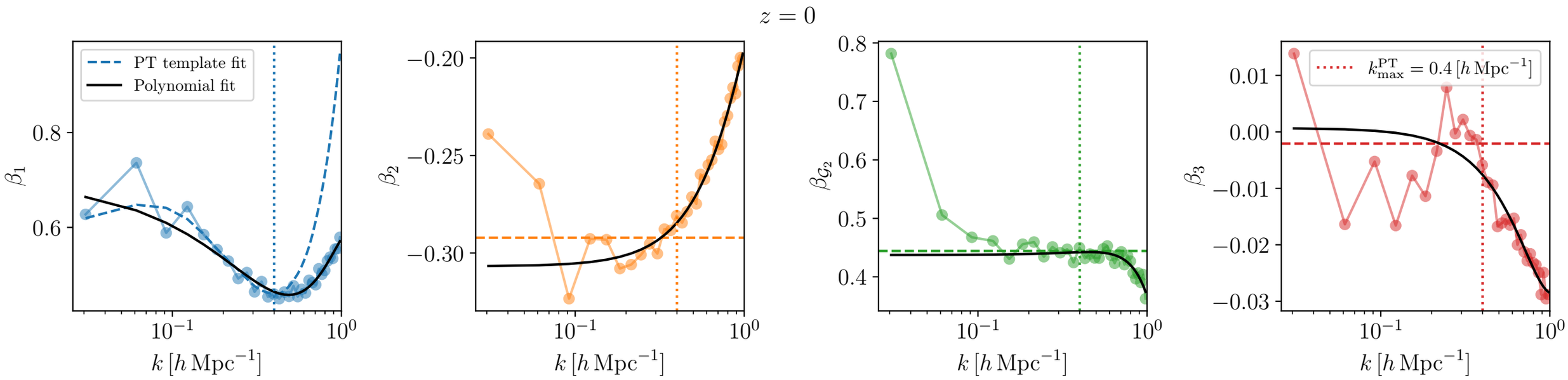
$z = 1$, real space

$$\delta_{\rm HI}(\mathbf{k}) = \beta_1(k)\tilde{\delta}_1(\mathbf{k}) + \beta_2(k)\tilde{\delta}_2^{\perp}(\mathbf{k}) + \beta_{\mathscr{G}_2}(k)\tilde{\mathscr{G}}_2^{\perp}(\mathbf{k}) + \beta_3(k)\tilde{\delta}_3^{\perp}(\mathbf{k}) + \ldots + \text{noise}$$



Transfer functions

$z = 0$, real space



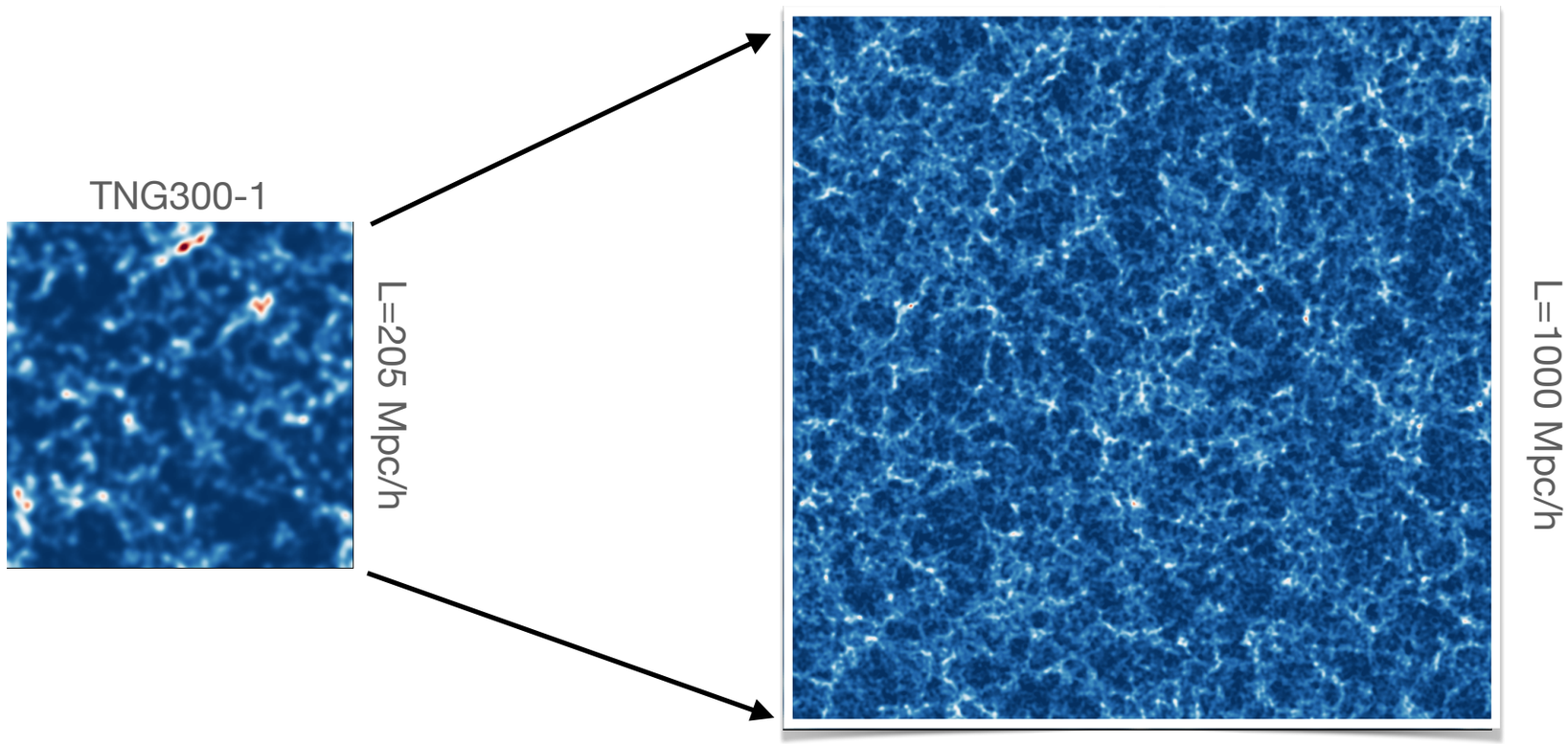
$$\delta_{\rm HI}(\mathbf{k}) = \beta_1(k)\tilde{\delta}_1(\mathbf{k}) + \beta_2(k)\tilde{\delta}_2^{\perp}(\mathbf{k}) + \beta_{\mathcal{G}_2}(k)\tilde{\mathcal{G}}_2^{\perp}(\mathbf{k}) + \beta_3(k)\tilde{\delta}_3^{\perp}(\mathbf{k}) + \ldots + \text{noise}$$

$$\beta_1^{\rm real}(k) = a_0 + a_1 k + a_2 k^2 + a_4 k^4,$$
$$\beta_{i\neq 1}^{\rm real}(k) = a_0 + a_2 k^2 + a_4 k^4$$

Hi-Fi mocks

https://github.com/andrejobuljen/Hi-Fi_mocks

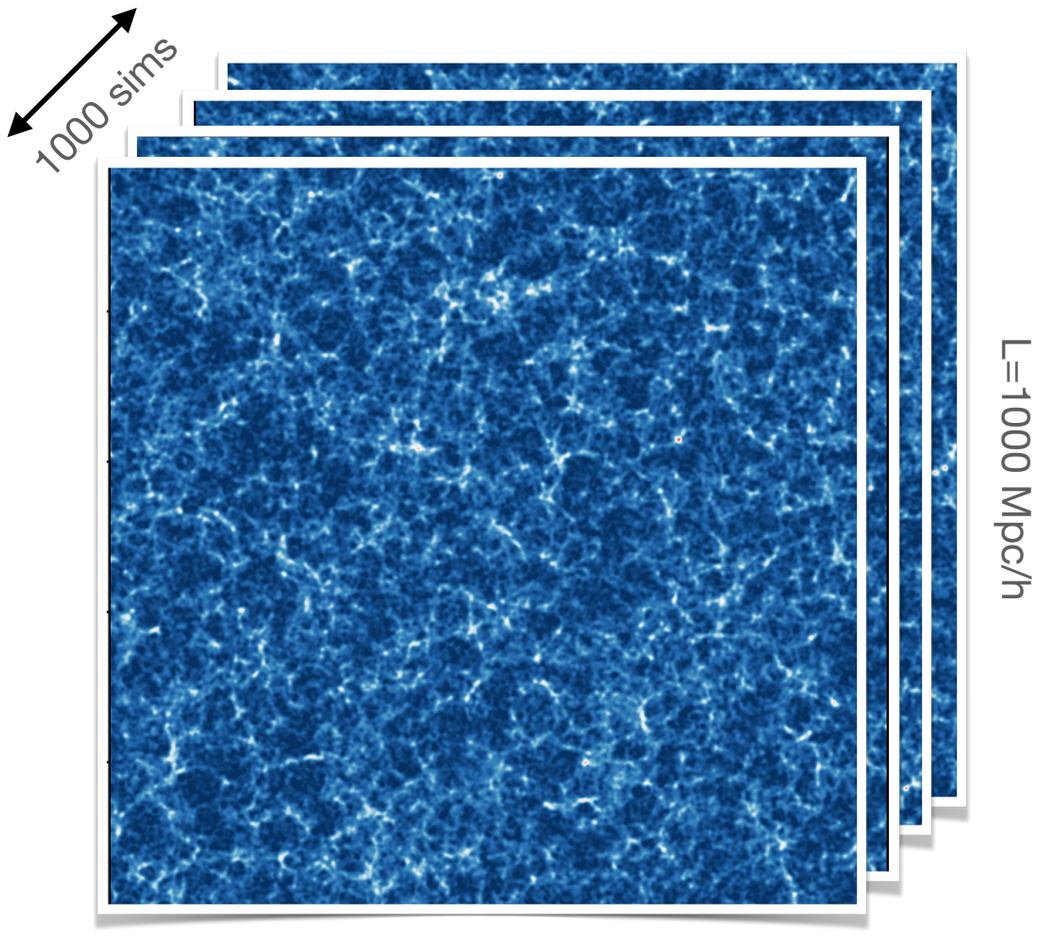
- Generate fast 3D HI field (Hi-Fi) level mocks
- Tuned to TNG HI clustering at z=[0-5]
- Extendable to any volume!
- Real & redshift space
- Publicly available, give it a try!



Hi-Fi mocks

https://github.com/andrejobuljen/Hi-Fi_mocks

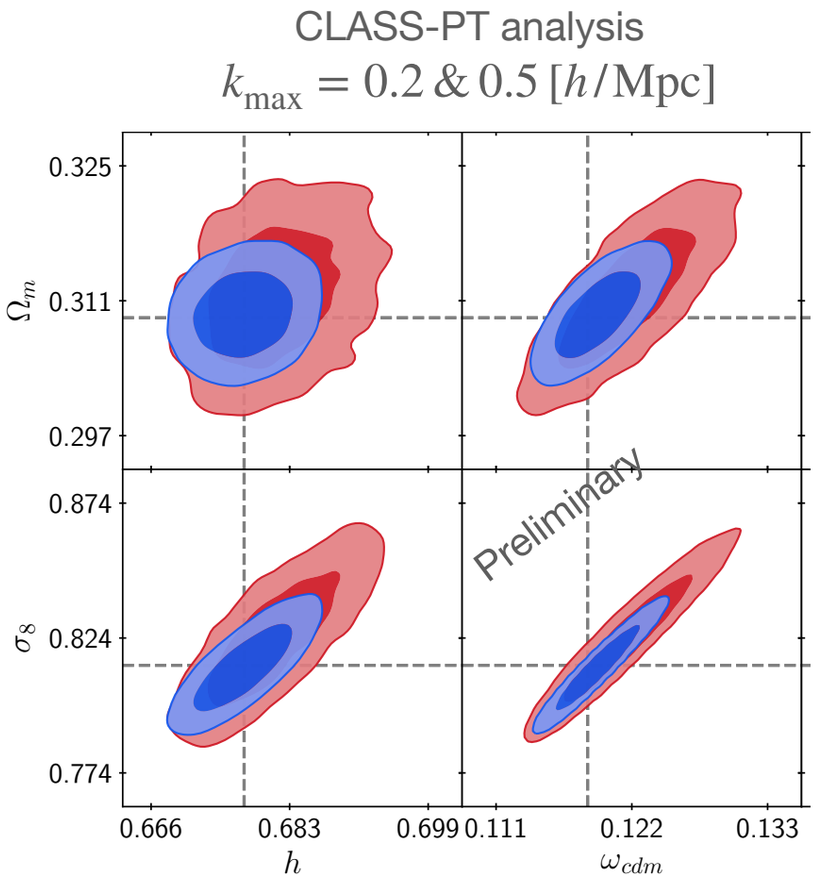
- Generate fast 3D HI field (Hi-Fi) level mocks
- Tuned to TNG HI clustering at z=[0-5]
- Extendable to any volume!
- Real & redshift space
- Publicly available, give it a try!



Hi-Fi mocks

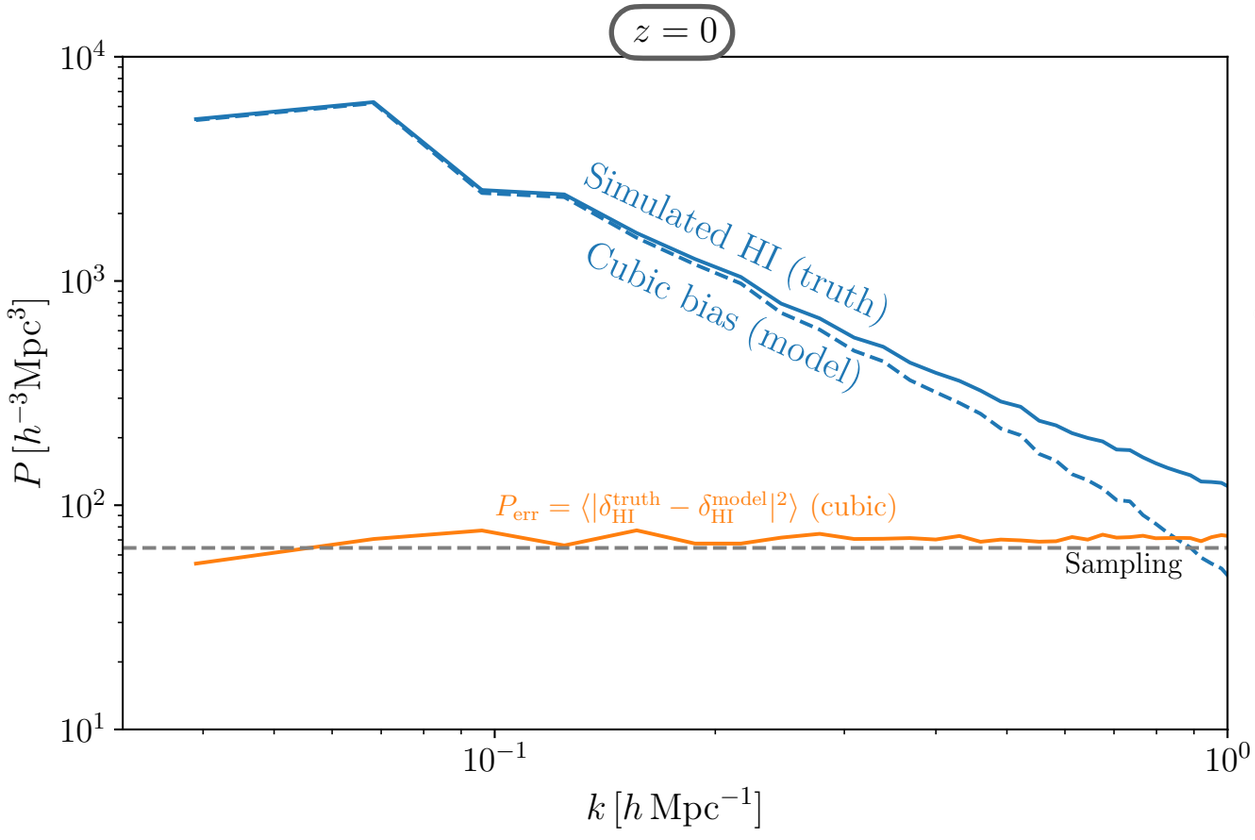
https://github.com/andrejobuljen/Hi-Fi_mocks

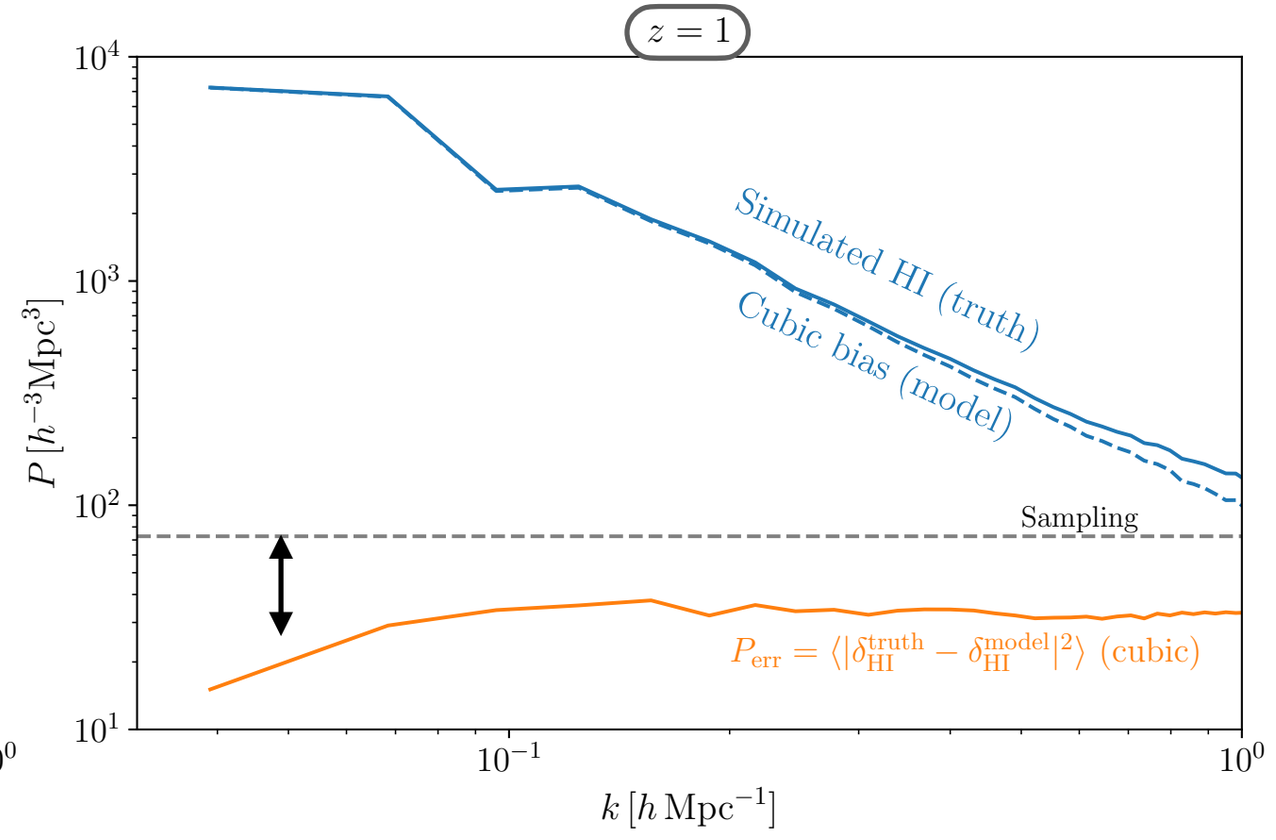
- Preliminary!
- Total volume 10 (Gpc/h)^3
- 1000 HI fields for covariance matrix
- Pk analysis in real space
- Unbiased cosmology $k_{\rm max} = 0.5$
- Soon: Pk in z-space, bispectrum…



HI noise properties

- $P_{\rm err}$ flat after including higher order terms
- $P_{\rm err}$ amplitude comparable to sampling noise $\bar{n}^{-1}$, but not equal



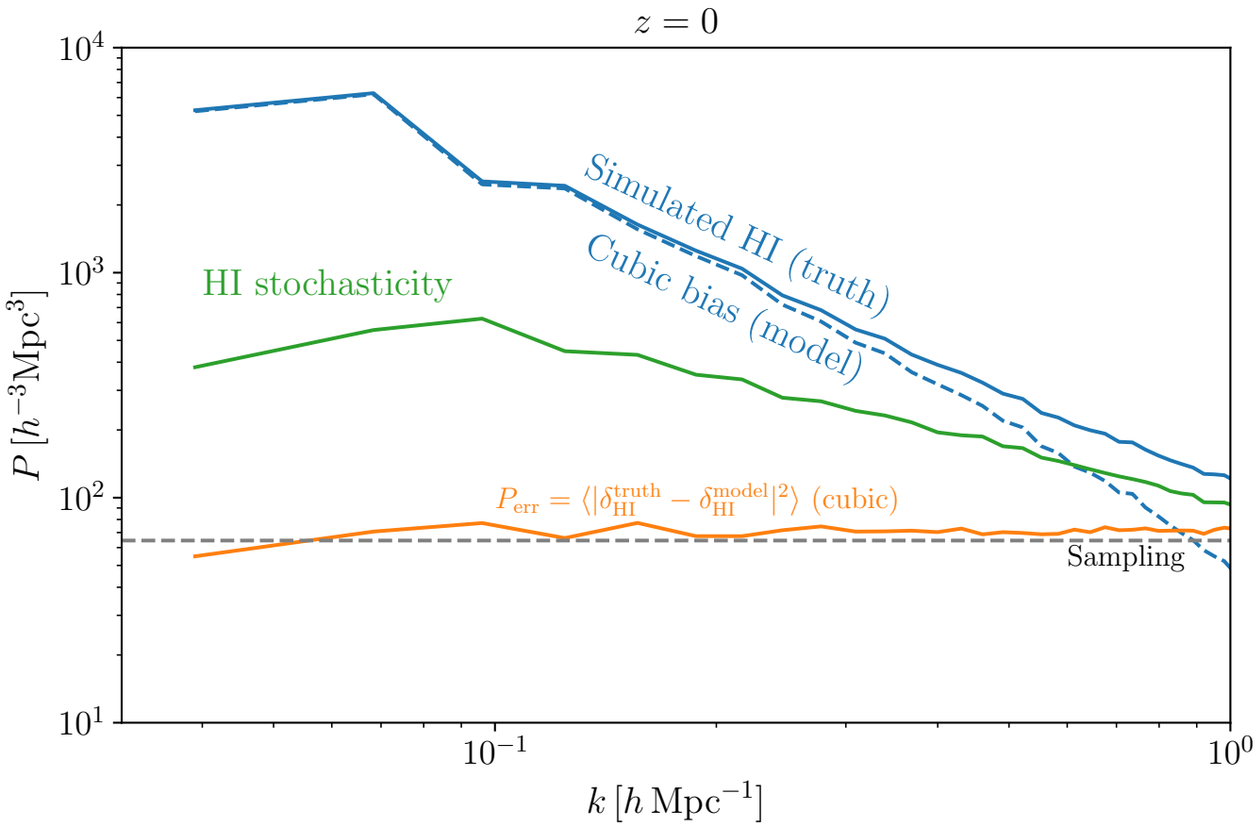


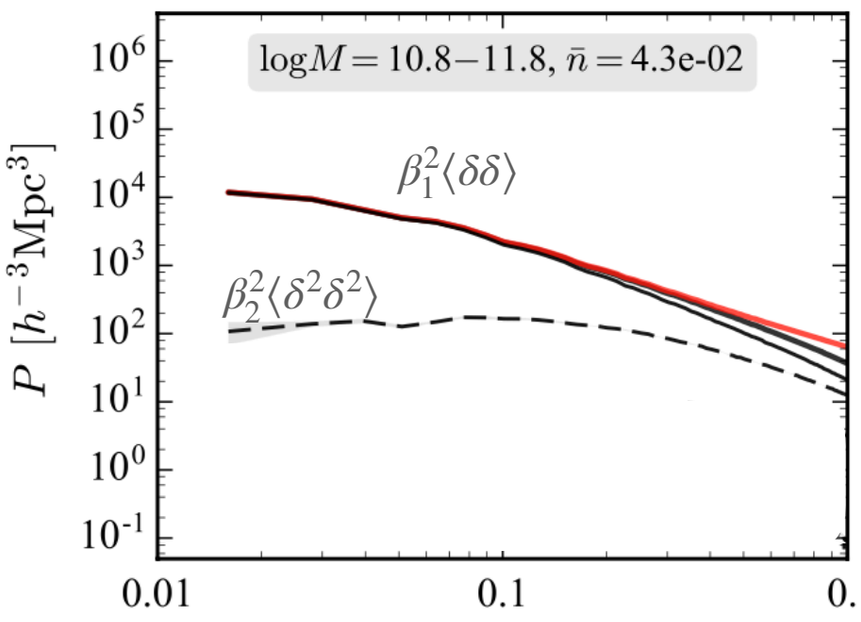
HI noise properties

- HI stochasticity: $\langle\, |\, \delta_{\rm HI}^{\rm truth} - b_1 \delta_{\rm m}\, |^2 \rangle$
- $P_{\rm err}$ lower than stochasticity

$$\delta_{\rm HI}(\mathbf{k}) - \beta_1 \tilde{\delta}_1(\mathbf{k}) = \beta_2 \tilde{\delta}_2^{\perp}(\mathbf{k}) + \beta_{\mathcal{G}_2} \tilde{\mathcal{G}}_2^{\perp}(\mathbf{k}) + \ldots + \epsilon$$

- In contrast to galaxy surveys, higher order terms dominate $P_{\rm err}$





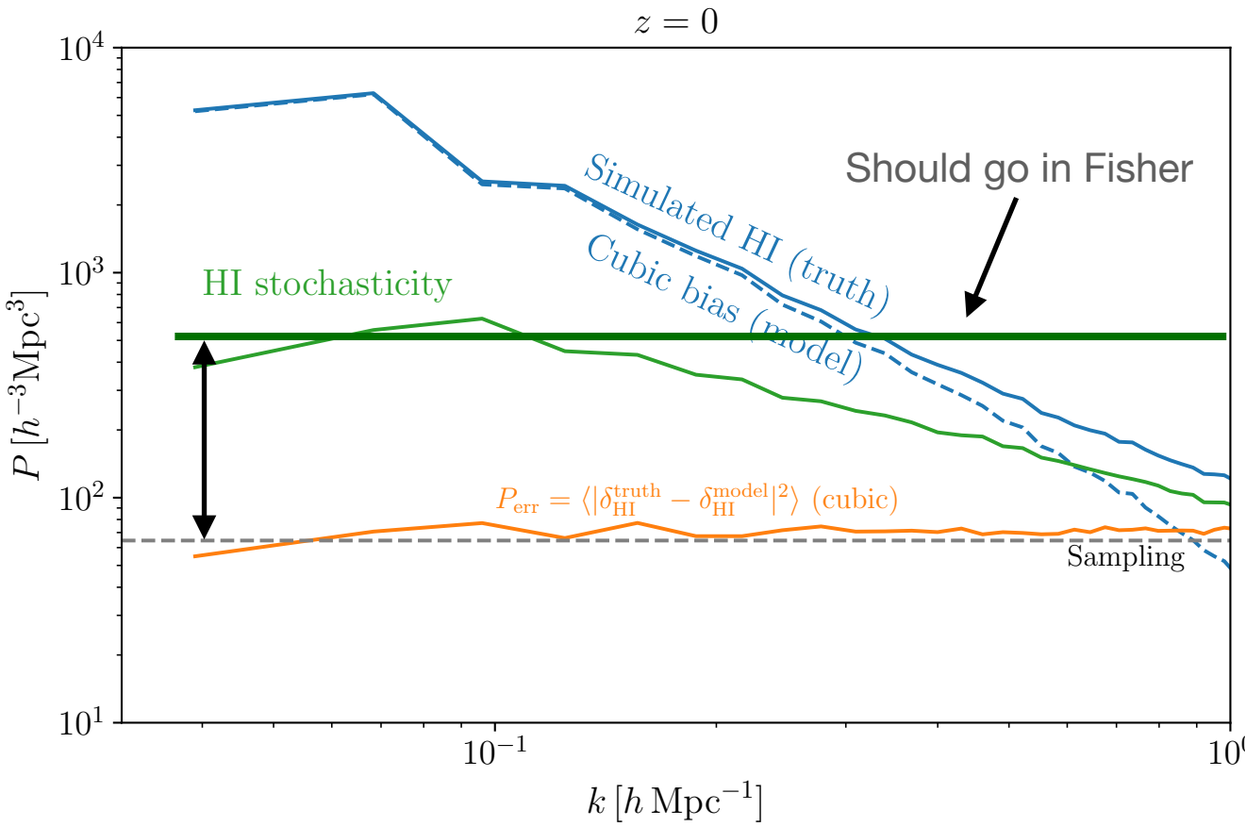
Schmittfull+18

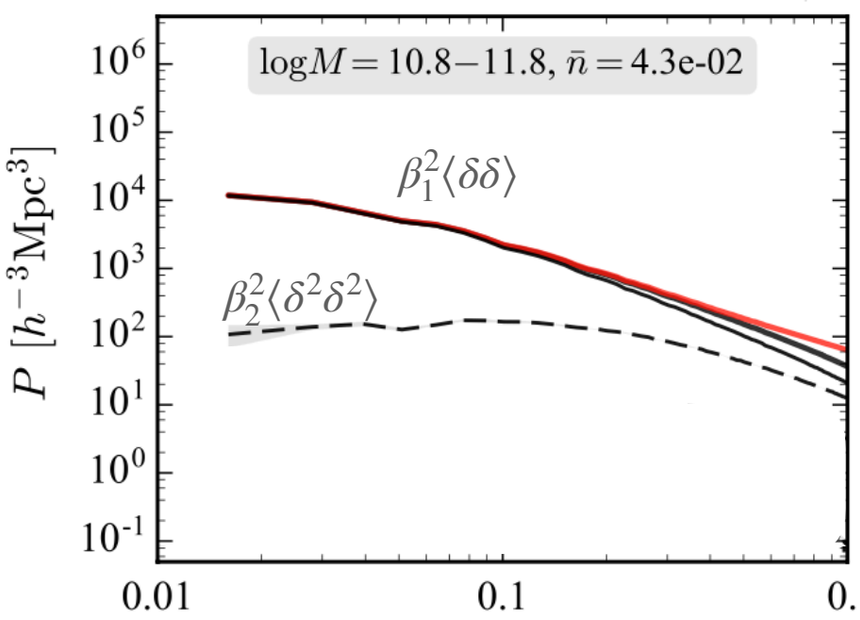
HI noise properties

- For Pk, higher order terms flat & degenerate with noise

- Most Fisher forecasts for Pk assume sampling noise (optimistic)

- Field level may do better!



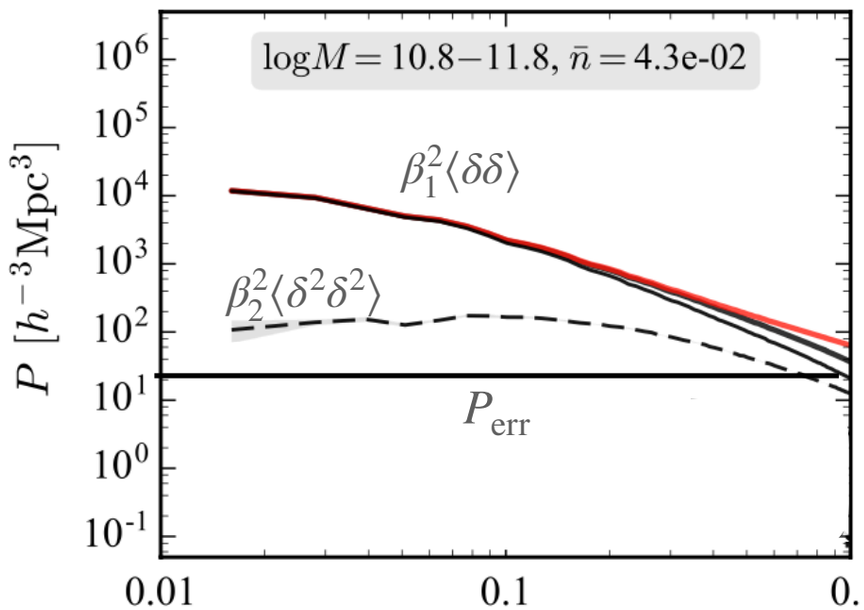


Schmittfull+18

Field-level vs Pk

on large scales

- Typical Pk analysis
- $P_{gg} = b_1^2 P_{\rm lin} + \overbrace{b_2^2 P_{\delta^2\delta^2} + P_{\rm err}}^{\text{Flat on large-scales}} + \dots .$
- $P_{gg} = b_1^2 P_{\rm lin} + \alpha/\bar{n}$
- $\alpha \equiv \bar{n}(P_{\rm err} + b_2^2 P_{\delta^2\delta^2})$
- b_2^2 term degenerate with noise and treated as noise
- Field level analysis doesn't pay this price, as b_2^2 term is calculable



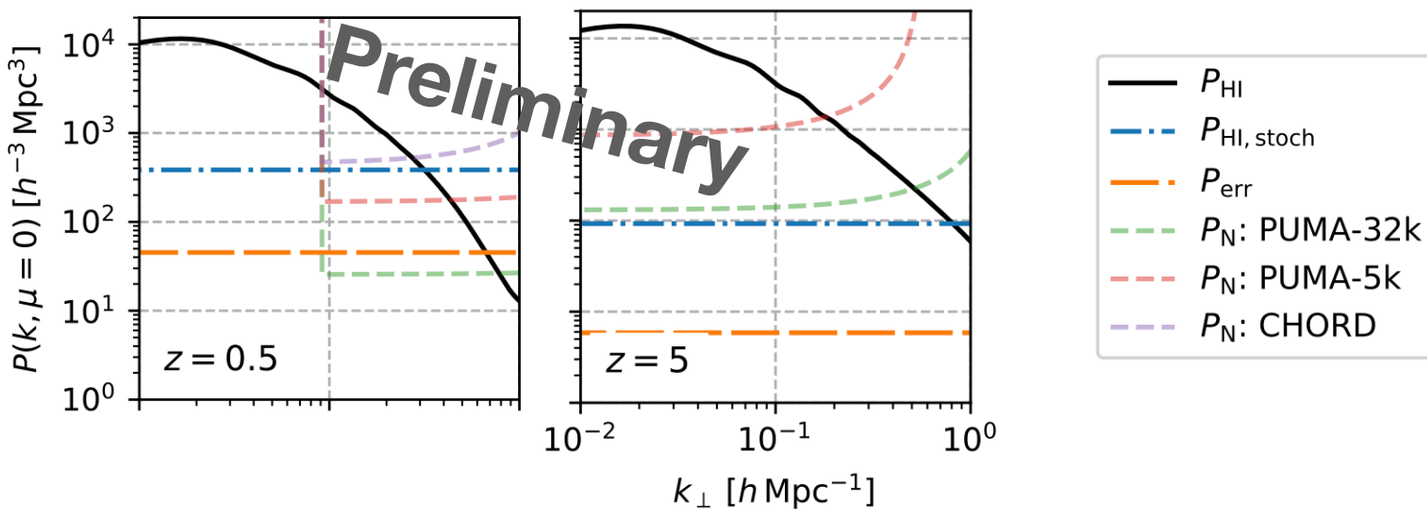
Schmittfull+18

Field-level vs Pk

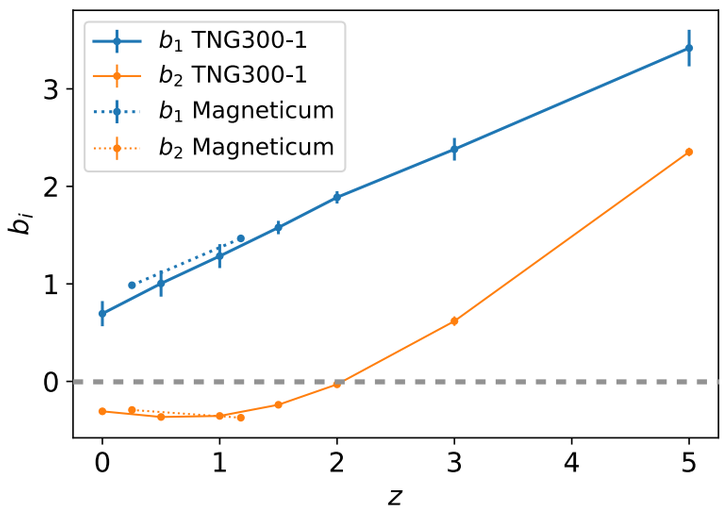
How big is the improvement?

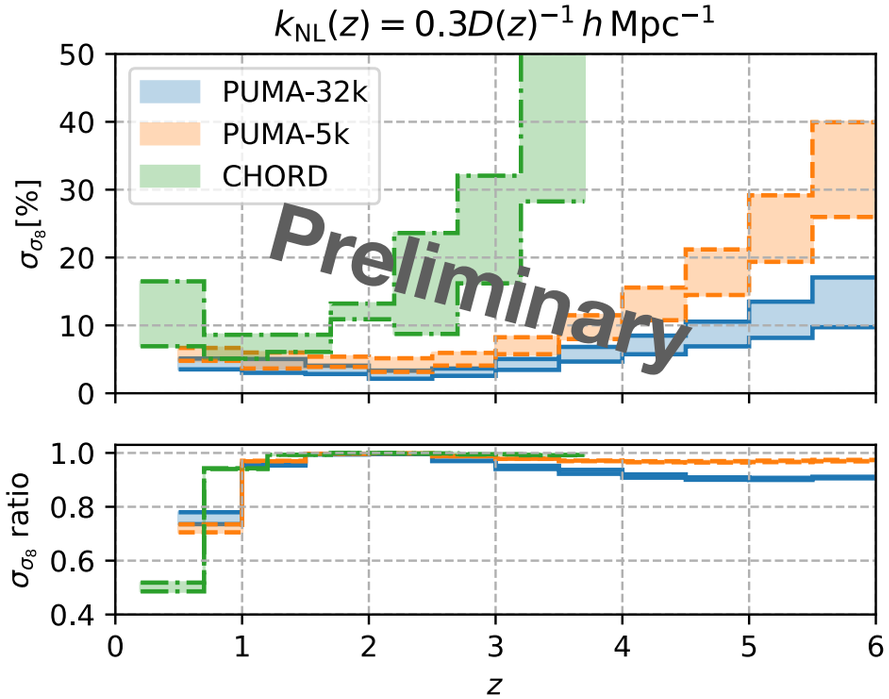
w/ S. Foreman & M. Simonović (in prep)

- Fisher forecasts with realistic noise levels
- Two forecast sets:
  1. remove $< \delta^2 \delta^2 >$ from covariance
  2. keep it in the covariance
- Mimic FL vs Pk analysis on large scales



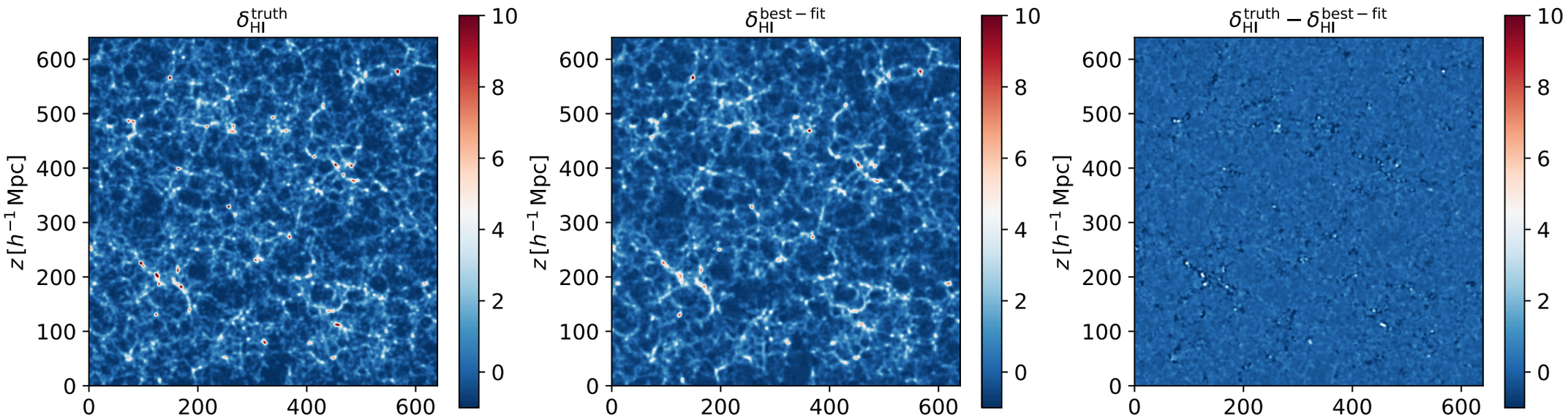
- ~20% improvement on cosmological parameters given **current** noise levels
- Larger improvement for "**optimal**" HI survey



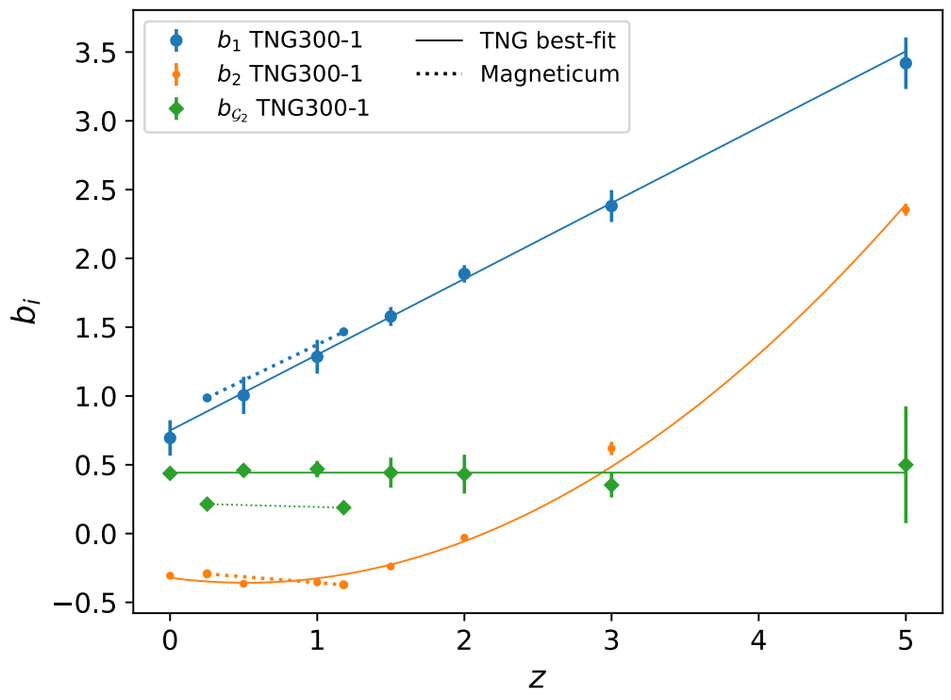


How about other hydro sims?

Magneticum Box2bhr (640 Mpc/h)



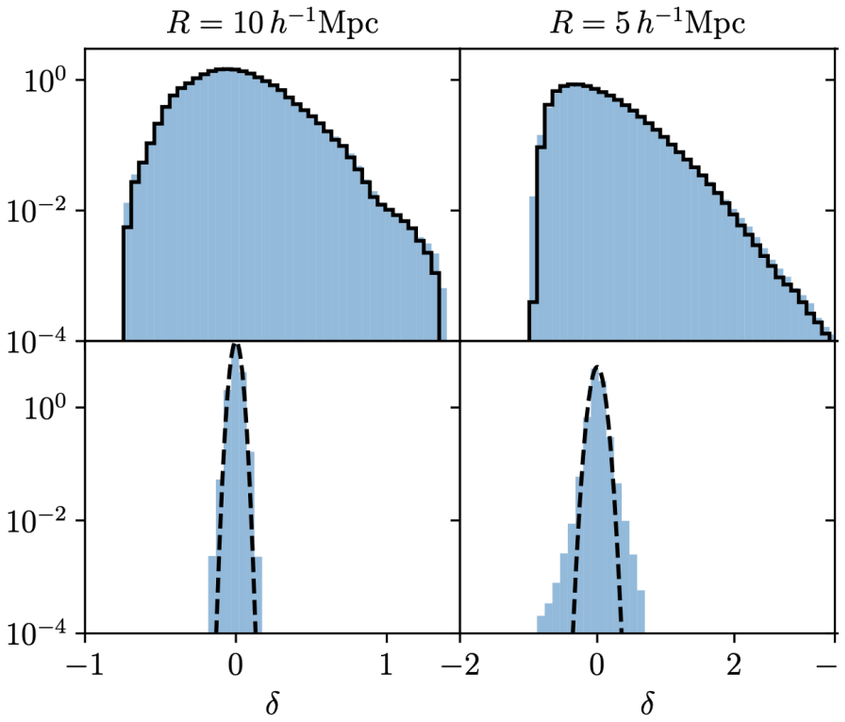
Slice depth 20 Mpc/h,
smoothed 2 Mpc/h Gaussian



How about other statistics?

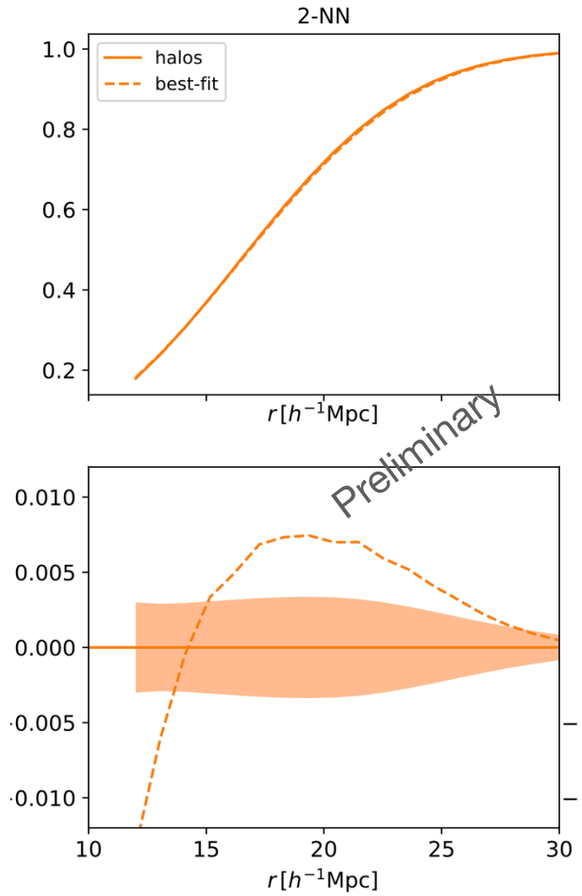
Void size function

1D PDFs



kNNs

with A. Banerjee



Looking ahead

- Easily adaptable to any biased tracer
- Realistic Fisher forecasts for HI: Pk vs ‘field-level’ analysis (in prep)
- Consistently combining constraints from non-Pk observables (1d PDFs, void-size function, kNNs etc) using field-level mocks (ongoing)
- Differentiable forward model in the full field-level inference… (prelim. results)
- Use ML to describe the residuals…

Conclusions

- PT + bias model at the field level works well for HI
- HI noise properties motivate beyond Pk analysis

- We provide code to generate fast and accurate HI field level mocks: **Hi-Fi mocks**
- Directly simulate HI field, no need to find halos –> differentiable
- Useful for forecasts, mocks, covariances, future data analysis

Thank you!