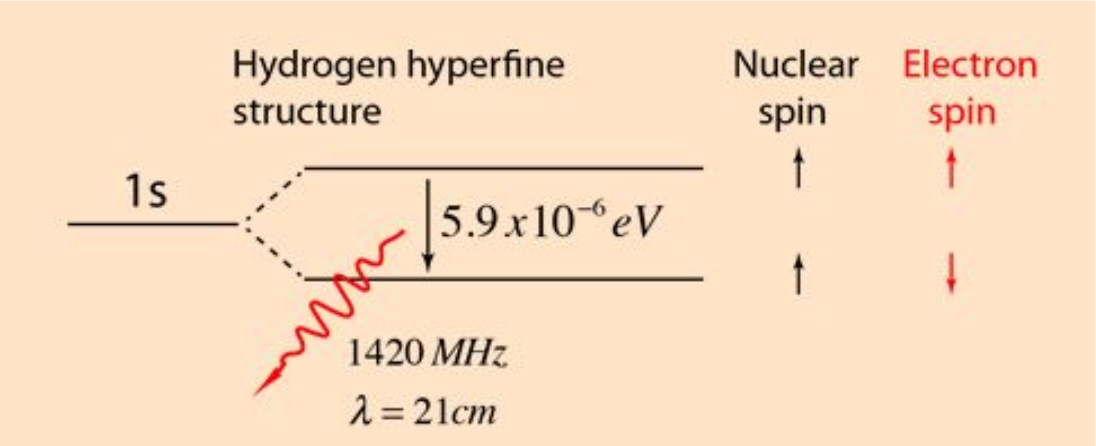
Boosting HI-galaxy cross-clustering signal through higher-order cross-correlations

Eishica Chand[1], Arka Banerjee[1], Simon Foreman[2] and Francisco Villaescusa-Navarro[3,4]

1 IISER PUNE

2 Arizona State University

3,4 Simons Foundation Flatiron Institute; Princeton University



http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/h21.html#c1

➔ This sharp line feature, and its redshifting ($\lambda_o = \lambda_e(1+z)$) , provide a window to **probe HI in both space and time**.

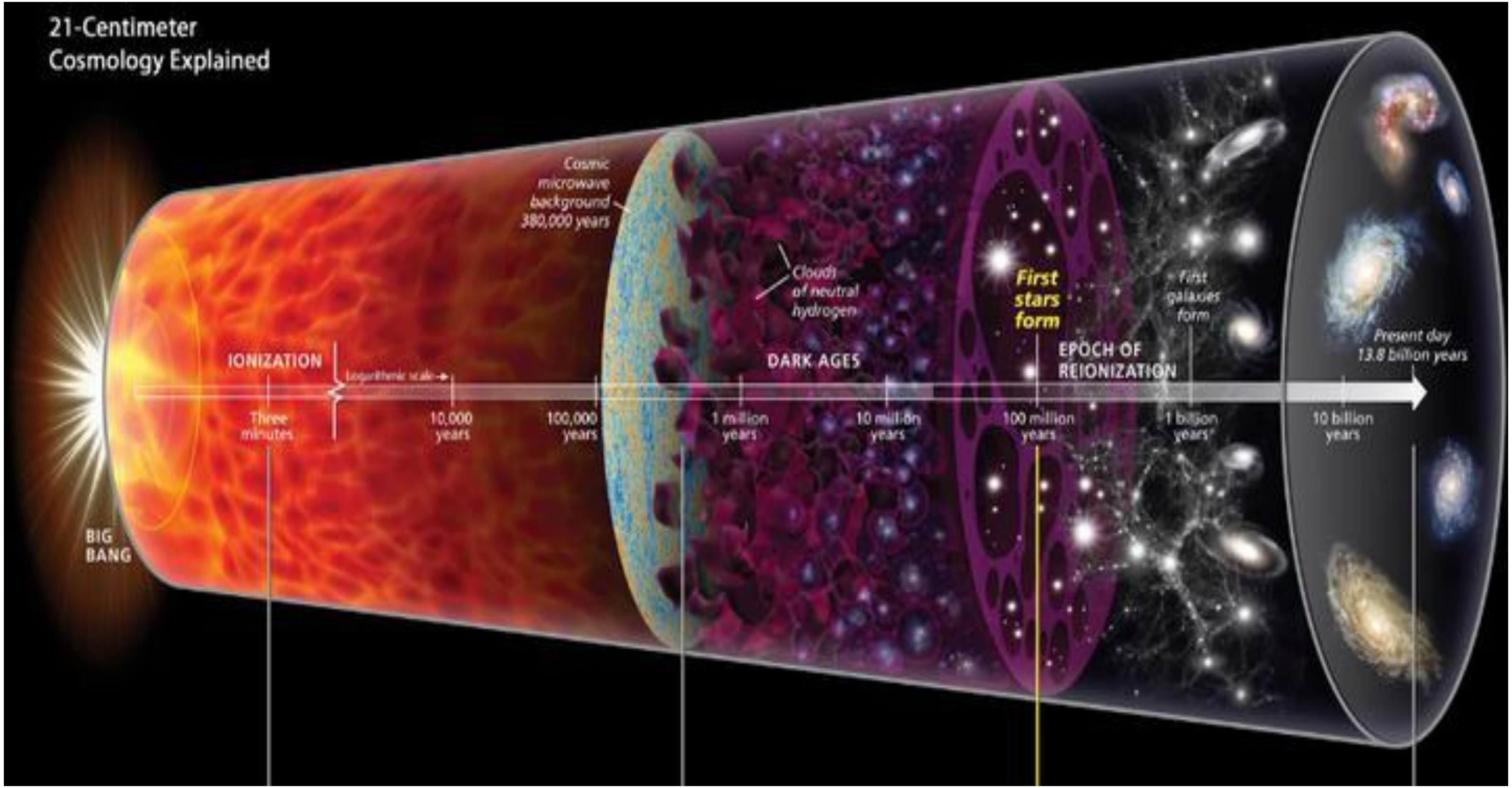
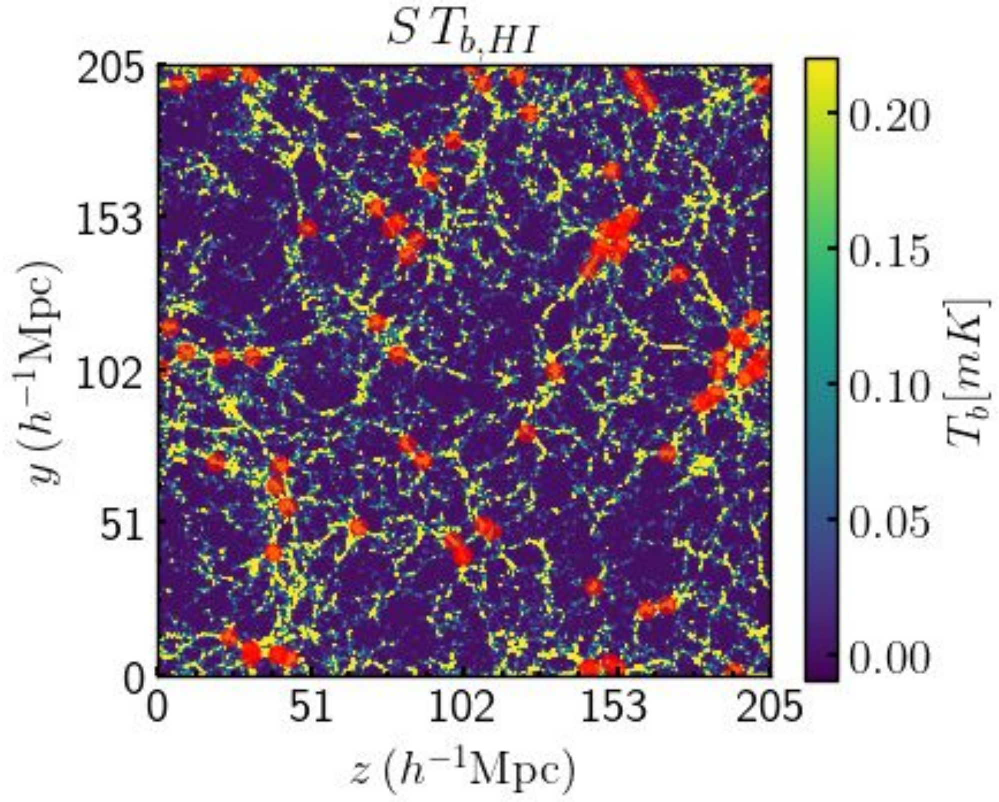


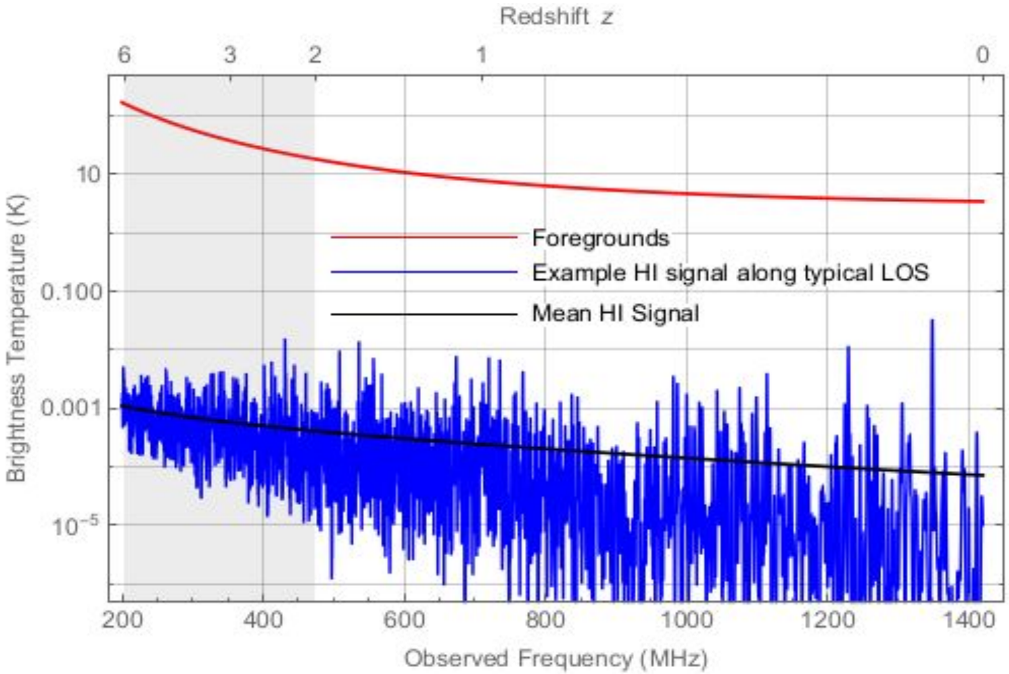
Image credit: [Roen Kelly/Discover magazine]

IllustrisTNG300 Simulation Data



HI observational challenges

1. **Foregrounds (FGs):**



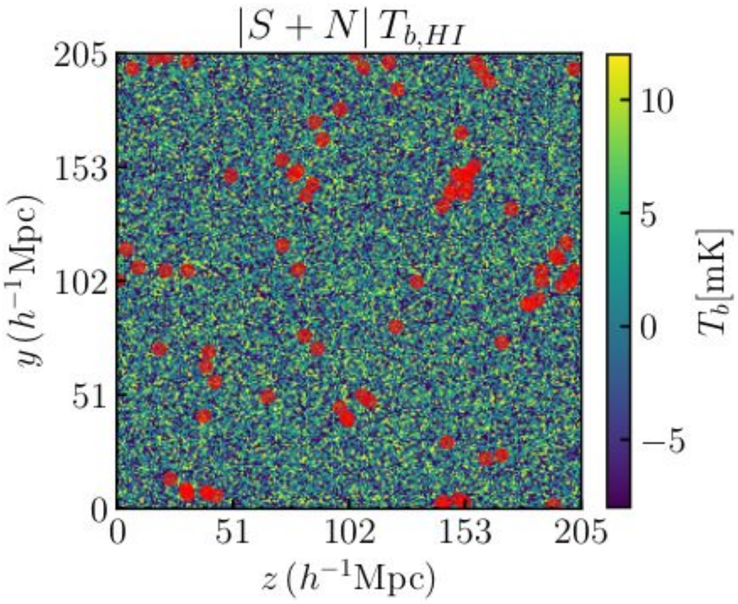
Loss of small-$k_{\parallel}$ modes

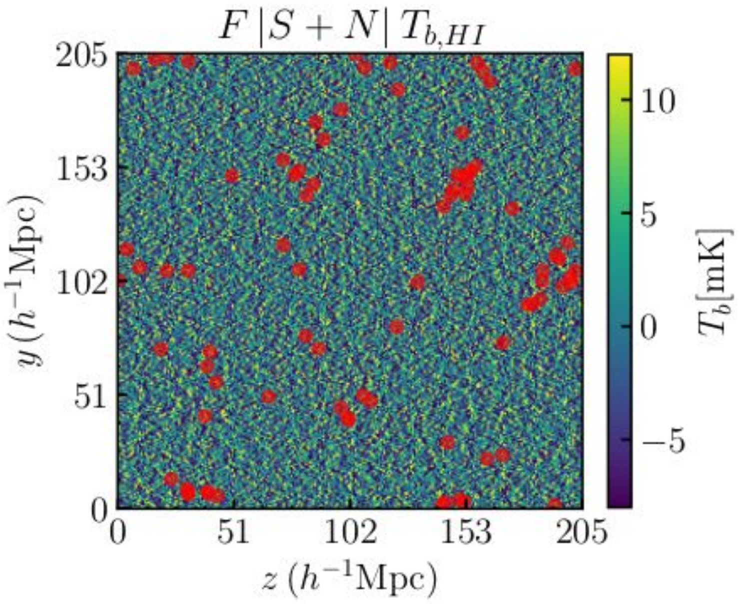
Cosmic Visions 21cm Collaboration, doi:10.48550

2. Thermal Noise Component:

In radio observations the thermal noise component, incorporates the effects of **sky and instrument temperature**, which contaminates the observed 21cm maps.

- ➔ Due to HI survey contaminants, the HI IM signal **hasn't been** yet **detected in auto-correlation** studies.
- ➔ Detection achieved via **cross-correlating** (two-point) HI field with **other LSS tracers**.
- ➔ Two-point statistics limited by non-Gaussian clustering information, thus higher-order cross-correlation techniques are of importance.





$|S+N|\ T_{b,HI}$: $S\ T_{b,HI}$ + Thermal Noise , $F|S+N|T_{b,HI}$: Filtered $|S+N|\ T_{b,HI}$ field

[CHIME Noise Model, t_{obs} = 22 days] **[F : 1 - exp(-$k_z/k_{zmin})^4$, k_{zmin} = 0.3 h/Mpc]**

k Nearest Neighbor(NN)-Field Framework

- Consider **two datasets**: one of discrete points and the other as a continuous field (ρ), both confined within a bounded volume. Finely sample the volume with another set of points (referred to as **'queries'**), **ensuring their number is much larger than that of the data points.**
- **Joint clustering** between the datasets at distance *r*, **P(≥*k*,ρr*):** the joint probability of finding at least ***k*th nearest neighbor** data point to a query point within r, along with the **continuous field smoothed over *r* to exceed a threshold** (**ρr** > **ρr***). This is a measure of Joint-Cumulative Distribution Functions (CDFs).
- **Cross-clustering:** defined as the Excess CDF over the auto-clustering of individual datasets, and given by **P(≥*k*,ρr*)**/ (**P(≥*k*)** × **P(ρr*)**).

Two-point Statistics (2PCFs)

- Discrete tracer-field correlations using **2PCFs** involves stacking of field over positions of discrete tracers.

- One has to compute the density of field enclosed within thin shell (radius r , thickness dr) at the positions of discrete tracers, denoted by $\rho_{enc}(\boldsymbol{r})$.

- The average of the enclosed density over all tracer positions, denoted by $\langle \rho_{enc}(\boldsymbol{r}) \rangle$ captures the two-point **cross-correlation** between tracer and field.

Chi-Square Analysis

Quantification of total SNR and Hypothesis Testing:

- ❑ Null hypothesis - No correlations between halos and HI fluctuations.
- ❑ Alternate hypothesis - Correlations present.
- ❑ Covariance Matrix:

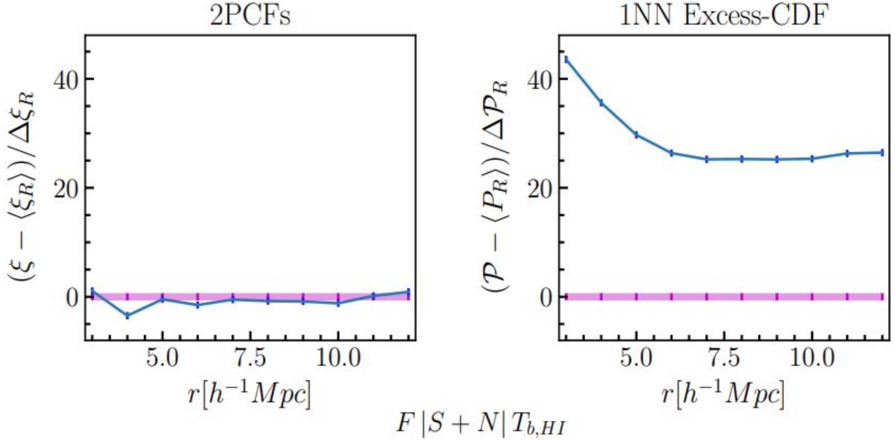
$$C(i,j) = \left\langle \left(D^p_R(i) - \left\langle D^p_R(i) \right\rangle \right) \left(D^p_R(j) - \left\langle D^p_R(j) \right\rangle \right) \right\rangle$$

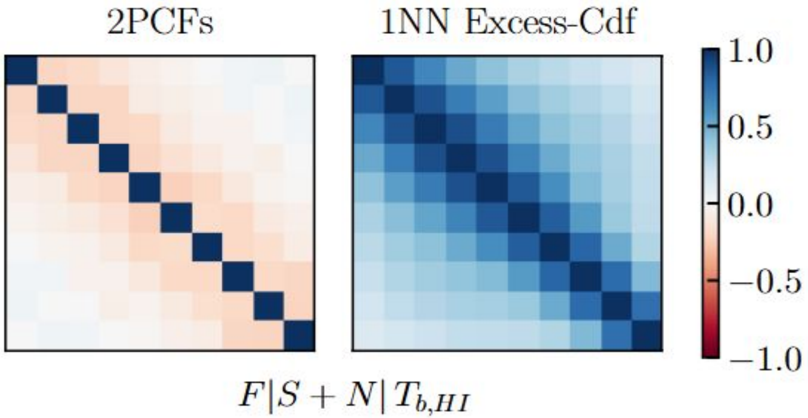
- ❑ Chi-Square value:

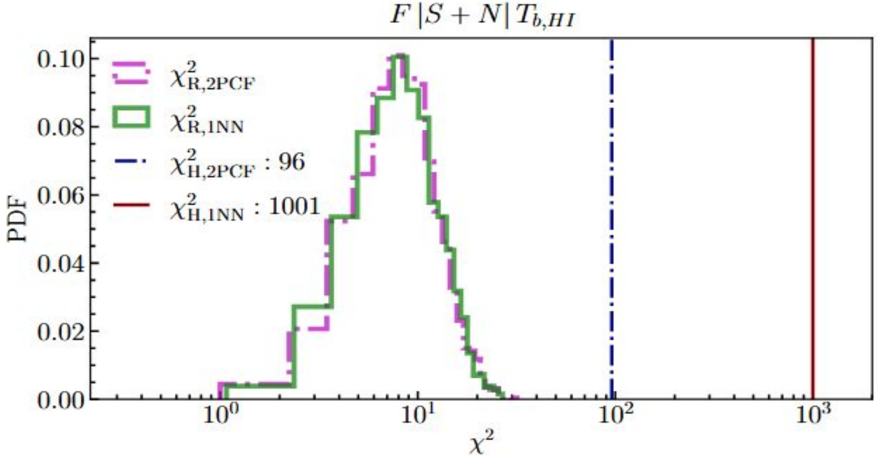
$$\chi^2_y = \sum_{i,j=1}^{n_{\rm bin}} \left(D_y(i) - \left\langle D^p_R(i) \right\rangle \right) C^{-1}(i,j) \left(D_y(j) - \left\langle D^p_R(j) \right\rangle \right)$$

Here D refers to data vector which for the case of kNN-field framework is 1NN Excess-Cdf measure, whereas it's the two-point cross correlation for the 2PCFs.

Results







$\sigma_{2PCF} = 8$

$\sigma_{1NN} = 31$

➔ Gains from Improved Level of FG filtering and Thermal Noise levels

$T_{b,\mathrm{HI}}$ **Field**	F_{P_2} : $1-\exp\left(-k^2/k^2_{\parallel,\mathrm{min}}\right)$		F_{P_4} : $1-\exp\left(-k^2/k^4_{\parallel,\mathrm{min}}\right)$	
	σ_{2PCF}	σ_{1NN}	σ_{2PCF}	σ_{1NN}
$F_{0.05h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{4yrs}}\rvert T_{b,\mathrm{HI}}$	75	173	80	173
$F_{0.1h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{4yrs}}\rvert T_{b,\mathrm{HI}}$	60	176	58	167
$F_{0.15h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{4yrs}}\rvert T_{b,\mathrm{HI}}$	43	181	44	167
$F_{0.2h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{4yrs}}\rvert T_{b,\mathrm{HI}}$	27	177	33	163
$F_{0.25h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{4yrs}}\rvert T_{b,\mathrm{HI}}$	18	181	41	162
$F_{0.3h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{4yrs}}\rvert T_{b,\mathrm{HI}}$	15	180	57	163
$F_{0.3h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{1yr}}\rvert T_{b,\mathrm{HI}}$	10	128	38	114
$F_{0.3h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{6mos}}\rvert T_{b,\mathrm{HI}}$	9	98	29	85
$F_{0.3h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{2mos}}\rvert T_{b,\mathrm{HI}}$	2.9	57	16	49
$F_{0.3h\mathrm{Mpc}^{-1}}\lvert S+N_{\mathrm{22d}}\rvert T_{b,\mathrm{HI}}$	0.8	37	8	31

Conclusions

- The ***k*NN-Field framework** consistently achieves **higher HI-galaxy cross-clustering compared to 2PCFs**, offering greater constraining power and reduced sensitivity to foreground filtering or instrumental noise.

- Applying this framework to 21 cm data from **surveys like CHIME**, alongside galaxy data from optical surveys such as **eBOSS and DESI**, shows significant potential.

- Ongoing work includes **modeling the signal**, which is particularly interesting due to the **highly non-linear scales involved**.

Thank You!

Acknowledgements

- **IISER PUNE**, and it's supercomputing facility (**PARAM Brahma**).
- Prime Minister Research Fellowship (**PMRF**), Ministry of Education, Government of India.