

Finanziato dall'Unione europea NextGenerationEU







21cmFASTv4: Multi-Tracer Inference at High Redshift James Davies

Spoke 3 III Technical Workshop, Perugia 28 Maggio, 2025

ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Scientific Rationale: Multi-Tracer Inference

Credit: Aman Chokshi



ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Credit: Aman Chokshi

Scientific Rationale: Multi-Tracer Inference

We have many different methods of observing this epoch:

- Galaxy Surveys via JWST, VLT, Roman
- Line intensity mapping via AtLAST, CCAT-Prime, SPHEREx
- Radio Interferometry via SKA, HERA, LOFAR, MWA
- CMB (optical depth, kSZ...)

Each of these observations probes different scales, physics etc.. So each provides a highly complementary piece of the puzzle, and we must use as many as possible simultaneously to achieve the best possible constraints on the early universe.









Technical Objectives, Methodologies and Solutions



ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Technical Objectives, Methodologies and Solutions



ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Technical Objectives, Methodologies and Solutions

Making progress in this framework means one of two things:

- Accelerate the model: Allowing us to test more hypotheses and relax priors
- Include more physics: Allowing us to test against a more diverse range of observations, adding terms to the likelihood and narrowing the posterior











Methodology: 21cmFASTv4

- Open source code for quickly producing large-scale simulations of the first billion years
- Python-wrapped C Library with openMP threading
- -We have upgraded the code to provide a discrete source population, greatly increasing the range of observables we can use for inference.



Davies, Mesinger, Murray (2025). Submitted to A&A

ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Methodology: HMF Sampling



Davies, Mesinger, Murray (2025). Submitted to A&A

ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Methodology: HMF Sampling



Davies, Mesinger, Murray (2025). Submitted to A&A

ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Methodology: HMF Sampling



ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing







 $\delta = 0.00$

 $\delta = 1.00$



 $\delta = 1.67$

Methodology: HMF Sampling

- Halo populations are sampled from a conditional mass function given as input
- Bulk galaxy properties (Stellar-to-halo mass, specific SFR etc...) are sampled from lognormal distributions which are correlated between descendants and progenitors

 $\delta = -0.90$

• The result is stochastic, self-consistent galaxy populations in a cosmological simulation using only a few core hours. *Meaning we can infer galaxy property distributions using data!*

Davies, Mesinger, Murray (2025). Submitted to A&A

position









Main Results: Galaxy Populations

- We are now able to use galaxy based observables to our inference pipeline. This is improves our analyses in several aspects
 - Adding terms to your likelihood produces tighter constraints
 - Correlations between galaxy and IGM observables are taken into account correctly.
 - Cross-correlation functions can be utilised in inference, which contain information not present in either observable alone











Main Results: Galaxy Cross-Correlation

- We have made forecasts for galaxy x 21cm cross correlation, testing narrow-band, grism, spectroscopic surveys with HERA and SKA
- We showed that Roman HLS grism can detect the cross correlation to >3 sigma with HERA inteferometry. SKA interferometry may additionally detect the cross using deep spectroscopy with VLT



Gagnon-Hartman, Davies, Mesinger (2025). Accepted by A&A









Final Steps: Postprocessing Multi-Tracer

- A discrete galaxy population can be post-processed to produce a wide range of observables
- CII, CO line intensities, Lyman alpha, and many more!
- -Each of these can be produced, analysed and cross-correlated with other results from the simulation with very little additional computation time



Zhang et al. in prep.









Final Steps: Database & SBI

- A database and inference pipeline for Simulation Based Inference (SBI) is being set up
- The database is currently running on Leonardo, which will total ~10^5 simulations with different parameter combinations/random seeds
- -Once finished, an NRE network will be trained to perform inference on high-redshift data



Triantafyllou et al in prep.









Final Steps: Database & SBI

- A database and inference pipeline for Simulation Based Inference (SBI) is being set up
- The database is currently running on Leonardo, which will total ~10^5 simulations with different parameter combinations/random seeds
- -Once finished, an NRE network will be trained to perform inference on high-redshift data











Conclusions

- Interpreting the high-redshift universe will require a multi-tracer approach
- We need flexible, fast simulations to probe the high-dimensional parameter space associated with early galaxies
- -21cmFASTv4 provides such a tool. The code is open souce and available at
- github.com/21cmfast/21cmFAST
- -We are currently developing analysis and inference pipelines to make the most of these new capabilities











Thank you!

ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing









Bonus: Galaxy Property Distributions











Bonus: Stochastic Cell Histories











Bonus: Mass Functions



ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing