Simulation Based Inference with galaxy clustering

🧪 Background

The non-Gaussian distribution of galaxies carries significant cosmological information, but its extraction relies on having accurate mock catalogs. As the volume and depth of modern galaxy surveys increase, it becomes necessary to use machine learning-based techniques to accelerate the simulations that mimic observations. In this project, we will employ such an accelerated forward model to go beyond simple 2-point statistics of galaxy clustering in both mock and real observations.

◎ Project Goal

We will use transformer-based accelerated forward models (AFMs) to create fast mock simulations of Sloan Digital Sky Survey galaxies. First, we will transform a large-volume, low-resolution particle mesh simulation into a high-resolution halo catalog. We will then perform interpretability exercises to understand how the transformer model learns the conditional probability distributions of halo properties. Subsequently, we will use this AFM to generate mock galaxy catalogs via a halo-galaxy connection model. Finally, we will perform simulation-based inference of cosmological and astrophysical parameters by measuring and analyzing various clustering statistics.

📖 Necessary Concepts

Basic cosmology, halo clustering, Generative machine learning, Simulation based inference (SBI) methods

💻 Computing Need

Initial analysis can be run on a laptop/google colab. Access to GPU would be useful. Might need cluster access towards the end of the module.
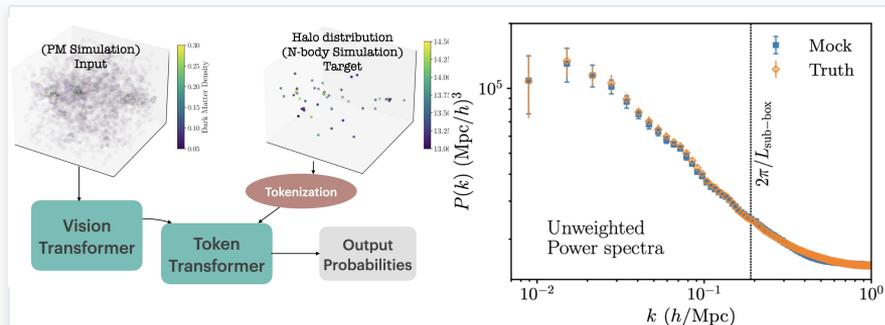
Left: Architecture of the AFM model. Right: Performance of AFM in predicting the power spectra in test simulations.

⭐ Background Resources

- Paper on related AFMs : 2409.11401, 2511.08438, 2409.09124
- AFM implementation: Github repo
- SBI papers: 2402.05137 , 2310.15246
- SBI repo: Github repo

Shivam Pandey

University of Arizona

🌐 Email: shivampcosmo@gmail.com