

# First Year Project Proposal

Helen Shao

December 14, 2025

## 1 Project Abstract

I am working with Daniel Eisenstein and the Abacus N-body simulation team to build the next generation of mock catalogs that will complement data from DESI, Euclid, Rubin/LSST, and Simons Observatory. The goals of the AbacusAurora simulations include studying the growth of structure through redshift-space distortions (RSD), weak lensing analysis, galaxy population statistics, and cross-correlations with CMB datasets such as Sunyaev-Zel'dovich maps. To achieve these objectives, the simulation requires (a) large volumes of  $8 h^{-1}$  Gpc boxes that can contain light cones up to  $z = 2$ , and (b) a novel galaxy assignment method via Massless Aggregating Particles (MAPs) that enables on-the-fly construction of merger trees. The former poses computational challenges due to the need for large-scale parallelization and efficient memory usage, which is central to the first part of my project. The latter is crucial for more realistic modeling of assembly bias and clustering as a function of galaxy properties. This is key for the second part of my proposal, which will focus on developing new density split statistics for the DESI Bright Galaxy Survey (BGS) sample to improve RSD measurements with a currently underutilized dataset through multitracer analysis.

First, to accomodate the enormous simulation volumes, I am implementing MPI parallelization to generate the kinematic initial conditions for over 35 trillion particles. The basis for this is the Zeldovich Approximation (ZA), which uses the primordial density power spectrum to compute the initial particle displacements and velocities needed to track the formation of cosmic web structures. Solving the Poisson equation for the AbacusAurora simulations requires performing Fourier transforms of  $32768^3$  Hermitian matrices distributed across multiple compute nodes. This past semester, I have been building this code infrastructure and currently testing it on the Aurora cluster. I now have a clear pipeline for matrix generation, grid decomposition, and inter-node communication that can accommodate the memory constraints. The goal is to complete this work by the end of January 2026 to align with the AbacusAurora timeline.

Short term next steps include scaling tests, testing edge cases of node and cell grid dimensions, and ensuring accurate reproduction of previous versions of the code—specifically verifying that simulations run at different resolutions match up to their common Fourier modes. Second, I will make use of MAP-defined galaxy populations to design density split statistics for the BGS sample. The DESI BGS contains over 10 million galaxies ( $z < 0.6$ ), but currently only  $\sim 13\%$  of the sample is utilized in clustering analyses due to modeling challenges with denser samples. While simple color-based splits (red vs. blue galaxies) can create different galaxy populations, they do not provide large enough bias differences for effective multitracer cosmic variance cancellation. To address this limitation, I will develop a density split methodology that tags each BGS galaxy by its local environment—specifically, counting faint photometric neighbors (from deep imaging surveys like LSST and Euclid,  $< 26$  mag) around each BGS spectroscopic galaxy within  $\sim 1$  Mpc scales. Galaxies with many neighbors are tagged as residing in high-density environments (clusters, groups), while those with few neighbors are tagged as low-density (field). This environment tagging creates density-based subsamples with larger bias separation than color splits alone. The MAP method is well-suited for this because MAP properties such as local DM density, velocity dispersion, formation history naturally encode environmental information, enabling accurate environment tagging in mock catalogs. By cross-correlating these density-split samples, we can achieve improved multitracer analysis through better cosmic variance cancellation on large scales, while the abundant photometric neighbor counts also reduce shot noise at small scales, enabling precision RSD measurements across all scales. This work will maximize the scientific value of the dense BGS dataset and demonstrate new statistical techniques applicable to next-generation surveys.

## 2 Advising Committee Suggestions

In addition to my main advisor, Daniel Eisenstein, here are suggestions for my advising committee:

1. Lars Hernquist
2. Cora Dvorkin (physics department)
3. John Kovac
4. Charlie Conroy
5. Doug Finkbeiner

### **3 Parallel Research Projects**

- John Kovac
- Blake Sherwin

### **A Why Multitracer Is Especially Useful for RSD Measurements**