

# 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 next-generation mock catalogs for DESI, Euclid, Rubin/LSST, and Simons Observatory. The AbacusAurora simulations will study growth of structure through RSD and weak lensing, galaxy population statistics, and cross-correlations with CMB datasets. To achieve these objectives, the simulations require (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 galaxy clustering. 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 (Hahn et al., 2023) to improve RSD measurements with 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, which uses the primordial density power spectrum to compute the initial particle displacements and velocities. This past semester, I have been building this code infrastructure and testing it on the Aurora cluster. I now have a clear pipeline for the matrix generation, grid decomposition, and inter-node communication. The goal is to complete this work by the end of January 2026 to align with the AbacusAurora timeline. Second, I will make use of MAP-defined galaxy populations to design density split statistics for the currently underutilized BGS sample to improve RSD measurements. The dense BGS samples are ideal for multitracer analyses because they minimize shot noise (McDonald & Seljak, 2009). However, color-based splits do not provide large enough bias differences for effective multitracer cosmic variance cancellation. To address this limitation, I will develop a density split method that tags each BGS galaxy by its local environment—specifically, counting faint photometric neighbors around each galaxy within  $\sim 1$  Mpc scales. 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 in mock catalogs.

## 2 Advising Committee Suggestions

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

- Lars Hernquist
- Cora Dvorkin (physics department)
- John Kovac
- Charlie Conroy
- Doug Finkbeiner

### 3 Parallel Research Projects

- I am very interested in pursuing a parallel project with John Kovac on the problem of missing modes in CMB E/B decomposition. The idea is to remove lensing effects before computing the purification matrix so that E/B separation happens on the source plane (unlensed) rather than the observed plane (lensed). We have discussed this approach in detail over several meetings. I am currently familiarizing myself with the LensPyx package and will begin implementing initial tests using BICEP2/Keck maps.
- I am also finishing up a project with my Master's advisors (Blake Sherwin, Fiona McCarthy, and Miles Cranmer) on CMB B-mode polarization reconstruction with inter-scale Galactic dust correlations. I am at the stage of drafting the paper and preparing for submission.

### References

- McDonald, P., & Seljak, U. 2009, Journal of Cosmology and Astroparticle Physics, 10, 007  
Hahn, C., Wilson, M. J., Ruiz-Macias, O., et al. 2023, AJ, 165, 253