DATA- AND COMPUTE-INTENSIVE CHALLENGES FOR OBSERVATIONAL ASTRONOMY IN THE GREAT SURVEY ERA

Allocation: Blue Waters Professor/ 300 Knh
PI: Athol J. Kemball[1]
Collaborators: Di Wen[1], Michael Katolik[1], Samantha Thrush[1], Jamila Serena Taaki[1]

[1]University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

This Blue Waters project is motivated by the data- and compute-intensive challenges posed by the large-survey telescopes of the coming decade, including the Large Synoptic Survey Telescope (LSST) and the Square Kilometer Array (SKA), among others. The science goals of these telescopes require high data acquisition rates and concomitant innovations in data analysis methods, some very compute-intensive and requiring resources on the scale of Blue Waters. The specific subprojects included in this parent project report include: (1) exoplanet detectability in large-scale surveys and optimal new approaches to exoplanet detection and parameter estimation in general, and (2) development of new statistical estimators for dark matter distribution and clustering in large-scale cosmological simulations and surveys in anticipation of LSST and related surveys.

RESEARCH CHALLENGE

This project is focused on data- and compute-intensive problems posed by the next generation of telescopes in observational astronomy, specifically those forming part of the imminent Great Survey Era, such as the LSST [1] and the Square Kilometer Array (SKA) [2]. Their science goals require high data acquisition rates with concomitant requirements for novel compute-intensive analysis approaches given their considerably more complex instrumental data models. Specific areas of focus include: (1) development and characterization of new algorithmic approaches for transiting exoplanet detection and parameter estimation, particularly for future large surveys; (2) development of new statistical estimators for the distribution and properties of dark matter for use as discriminants of cosmological models and in preparation for future large-scale surveys such as LSST; and (3) algorithmic challenges that are foundational to radio-interferometric imaging for highly data-intensive future telescopes such as the SKA.

The relevance of this research with Blue Waters centers on the fact that the telescopes of the Great Survey Era such as LSST and SKA represent multibillion-dollar investments in the observational astronomy research infrastructure for the next decade. This research is critical to this key transition in observational astronomy.

METHODS & CODES

For radio-interferometry projects, the research team has used the generic software framework eM [6], which encapsulates community codes to allow efficient prototyping in the initial stages of algorithm exploration. Prototyping for other projects has been performed in parallelized Python. However, the team has developed optimized MPI codes when needed, such as in the cosmological analysis problem described below. When using community-based codes, the team carefully considers I/O and memory-access patterns in order to match the Blue Waters architecture and ensure resiliency.

RESULTS & IMPACT

The research subproject into exoplanet detection and estimation (with Jamila Serena Taaki) is strongly informed by the contemporary state-of-practice in spectral estimation and detection. A code framework has been developed and deployed on Blue Waters to evaluate the statistical performance of new transiting exoplanet estimators against data from the Kepler satellite [3], including algorithms based on marginalization, Monte Carlo methods, and new Bayesian approaches.

In a research subproject (with Di Wen), the team has explored the cosmological analysis of large-survey data with application for future telescopes such as LSST, specifically the development and characterization of estimators for the statistical properties of the distribution of dark matter in large-scale N-body cosmological simulations. This specific research project focuses on the use of the estimator for the counts-in-cells probability distribution function (PDF) $f(N,V)$, which defines the probability of finding N galaxies in a cosmological volume V. Although computationally intensive to estimate, this PDF contains significant embedded information on the properties of dark matter. The team's specific research interest is on the use of this distribution function as a discriminator of cosmological models and also as an assessment of its application to future survey data from LSST.

In the past year, the research team has applied the counts-in-cells PDF estimator for dark matter halos to the Dark Energy Universe Simulations (DEUS) [4] on Blue Waters. They have obtained measurements between redshifts $z = 0$ to $z = 4$ at both linear and nonlinear scales and have compared the best fits of four analytical models to the measured counts-in-cells distributions. The resulting counts-in-cells distributions $f(N,V)$ as a function of redshift z, cell-size, and cosmological model are shown in Fig. 1. The team has also used Blue Waters to compute particularly expensive fits to PDF models where variable precision arithmetic is required owing to very large numerical coefficients. The percentage difference between $f(N,V)$ for different cosmological dark energy models is a few percent at small z but can be significantly higher at high z, and therefore provide a useful manner to separate different dark energy models; here, the researchers consider quintessence and phantom dark energy against standard cold dark matter cosmology.

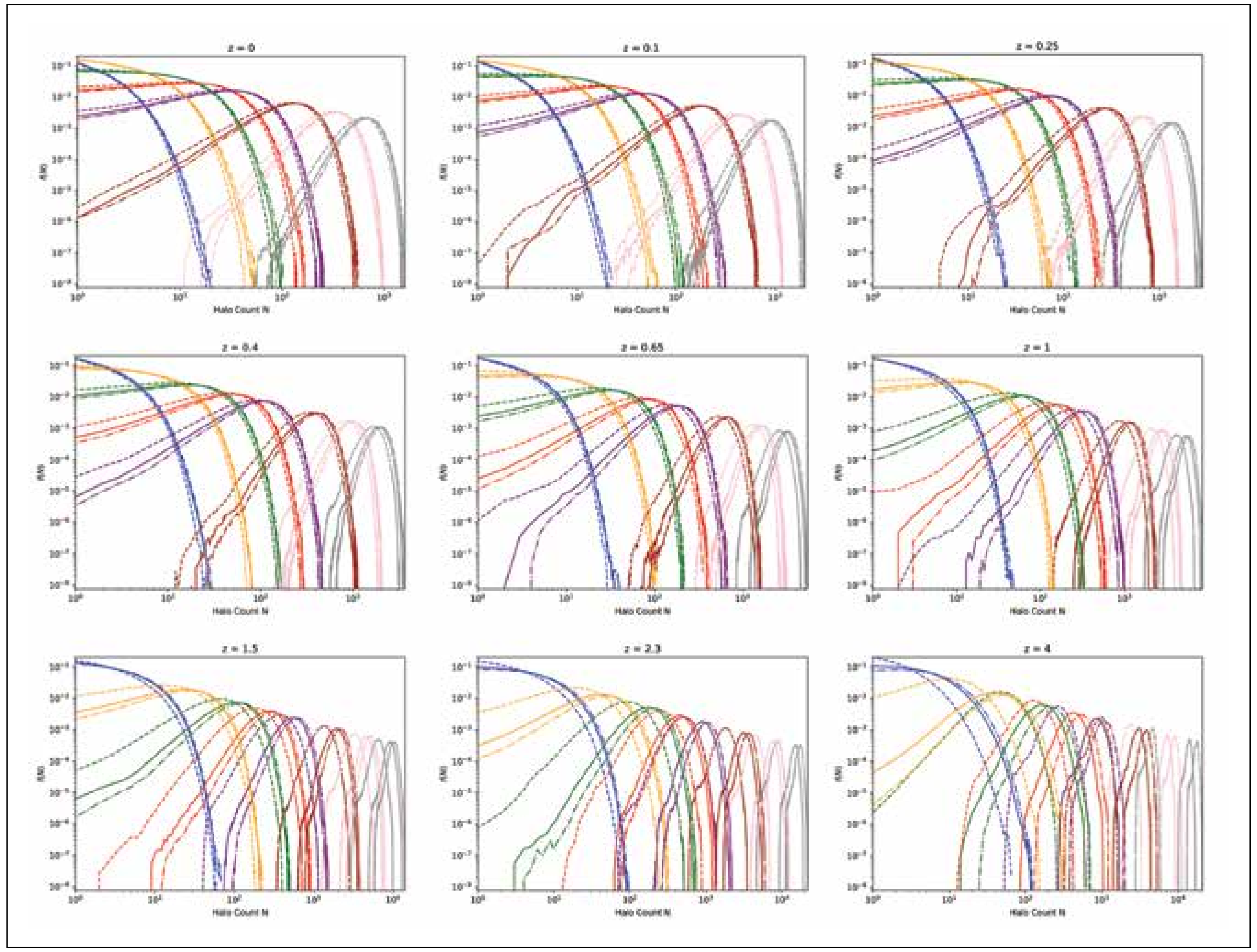

Figure 1: Counts-in-cells distributions $f(N)$ for the ΛCDM (solid line), RPCDM (Ratra–Peebles; dash line) and wCDM (phantom dark energy; dash dot line) models at various redshifts. Line series colors indicate spherical cell radii R = $2h^{-1}$ (blue), $4h^{-1}$ (orange), $6h^{-1}$ (green), $8h^{-1}$ (red), $10h^{-1}$ (purple), $15h^{-1}$ (brown), $20h^{-1}$ (pink) and $25h^{-1}$ (gray) [all Mpc].

The research team's work in radio-interferometric calibration and imaging complications concerns the most computationally challenging problems for future data-intensive telescopes in this field.

WHY BLUE WATERS

The specific research conducted under this allocation explores new computationally intensive approaches to data analysis in this discipline. Several key projects are not practical without petascale resources. In addition, the code bases used in a number of these research subprojects are based on serial community codes in a parallel framework and therefore benefit particularly from the balanced architecture of Blue Waters.

PUBLICATIONS & DATA SETS

D. Wen, A. J. Kemball, and W. C. Saslaw, "Halo counts-in-cells for cosmological models with different dark energy," *Astrophys. J.*, submitted, 2019.

D. Wen, "Counts-in-cells distribution of dark matter halos," 12th Great Lakes Cosmology Workshop, Rochester, NY, U.S.A., Aug 6–8, 2019.