Annual Report for Blue Waters Allocation

- ## Project Information
  - Title: Image Uncertainty Quantification and Radio Astronomical Imaging.
  - PI: Athol J. Kemball, University of Illinois at Urbana-Champaign.
  - Names and affiliations of students or collaborators: Michael Katolik (Graduate Student; UIUC); Di Wen (Graduate Student; UIUC); Yashar Hezaveh (Stanford); Neal Dalal (Faculty; UIUC)
  - Contact: akemball@illinois.edu

- ## Executive summary (150 words)

This project concerns the use of Blue Waters to address key challenges in interferometric astronomical imaging by using new algorithmic approaches uniquely enabled by petascale computing. Specific areas of focus include direction-dependent (or pixel-level) image fidelity assessment, novel approaches to interferometric calibration and data reduction that are vital to current and future telescope arrays, and numerical simulations in support of interferometric studies of gravitational lensing. In the current reporting period we have used 225/240 KNH of our allocation and have made significant advances in our continuing work in these research areas. Two graduate students advised by the PI are engaged in these projects.

- ## Description of research activities and results
  - ***Key Challenges:*** Interferometric imaging is an ill-posed inverse problem described by an integral equation that connects the measured spatial coherence of the incident radiation as a function of separation orthogonal to the propagation direction, the unknown source brightness distribution on the sky, atmospheric propagation effects, and the instrumental response model. In contrast to optical interferometry, radio interferometry at frequencies below ~1 THz allows heterodyne down-conversion of the incident radiation before digital sampling, thus preserving phase information. This preservation of information describing the incident radiation has historically allowed significant innovation in post-detection calibration and imaging in this domain using advanced signal processing and computational methods. The petascale computational era provides new opportunities to extend this innovation in important directions. The imaging integral equation is a complex mathematical system with a large number of unknowns and cannot be solved directly. Traditional approaches, codified as best-practice methods of solving the inverse imaging problem, include iteration, separation of loosely-coupled or nearly orthogonal terms, regularization, linearization, or the use of auxiliary constraints from independent measurements, amongst other techniques. In this environment, uncertainty quantification (pixel-level imaging fidelity assessment) remains a key challenge, as do more direct solution methods for calibration terms or image formation. Our work on Blue Waters has focused on novel approaches to these problems, uniquely enabled by

petascale computing. In the past year this has also included a limited simulation study related to dark matter sub-structure measurement using radio interferometry.

- ***Why it Matters*:** This work benefits directly several large investments in telescope construction (Atacama Large Millimeter Array (ALMA); $1.4b) and future large-scale projects planned in astronomy such as the Square Kilometer Array (SKA). The new data analysis methods will enable better scientific use of existing facilities (e.g. ALMA). An improved understanding of pixel-level imaging fidelity both assists effective design of future arrays (SKA) but also lowers barriers to the use of existing facilities by the broadest possible user community by enabling much more automated methods of reduction that capture best practice and convey uncertainties accurately. This removes the requirement that individual users be expert in detailed interferometric data reduction methods that can be time-consuming given the nature of the inverse imaging problem described above.

- ***Why Blue Waters:*** This work requires Blue Waters because of the computational scope of the new algorithms being explored as well as the exponentially increasing data rates in this domain. In astronomical interferometry the data rate is proportional to the number of telescopes in the array squared. Current arrays have ≤ 50 telescopes but future arrays have science goals that require sensitivities demanding possibly thousands of telescopes. In addition the data rate is further increased as digital data acquisition electronics increase geometrically in capacity with Moore's Law.

- ***Accomplishments:*** Our summary accomplishments on Blue Waters in the current reporting period include:
  - **Novel interferometric calibration algorithms:** In this area, we have continued to work on highly computationally-intensive metaheuristic and stochastic methods for interferometric calibration. In our previous report we described an assessment of various novel fitness functions that allow more direct solution methods to the interferometric calibration problem. In the current period we have implemented and tested these algorithms in detail on simulated interferometric data and have results on effectiveness and convergence. In Figure 1 we show the evolution in the distribution of a swarm of 100,000 particles representing individual interferometric solutions. The metric distance $\left\|\mathbf{G}-\hat{\mathbf{G}}\right\|_2$ of each particle from the true solution (which is known in this instance as this is a simulation study) is

shown on the x-axis, where zero metric distance represents the true solution. The objective fitness function is plotted on the y-axis. The best solution in the swarm is shown as a red point. The sub-figure at left depicts the initial random distribution in multi-dimensional parameter space; the sub-figure at right shows the distribution after ten iterations. The global optimization problem is complex and has many secondary minima, as is visible in the figures, however the true solution is found efficiently in this initial result. The convergence appears to be geometric and is shown in Figure 2. This work is continuing in production mode at present, evaluating both the fitness functions and the metaheuristic update algorithm details. The current reporting period has included significant code development and testing but our results are promising and allow a novel approach to this domain, now at a production level. This work is in collaboration with Michael Katolik (Graduate Student, UIUC).

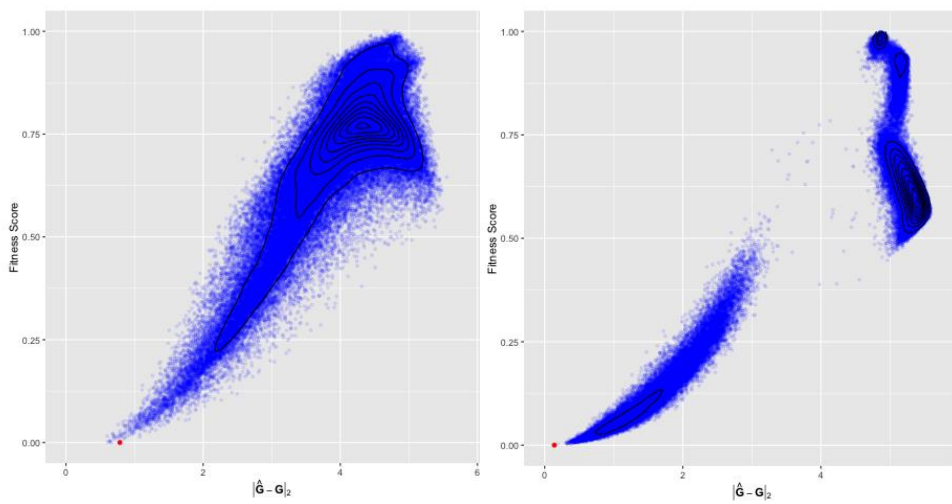

Figure 1. Distribution of particles in a metaheuristic optimization study concerning a novel interferometric calibration method. The x-axis is the metric distance from the true solution, so truth is at zero metric distance. The y-axis shows the fitness objective function value for each particle. Contours are drawn as Gaussian kernel density estimates. The figure at left shows the initial random distribution of 100,000 particles and the figure at right the distribution after ten iterations of the metaheuristic algorithm update. The red dot in each figure depicts the lowest objective fitness function value at a given iteration.

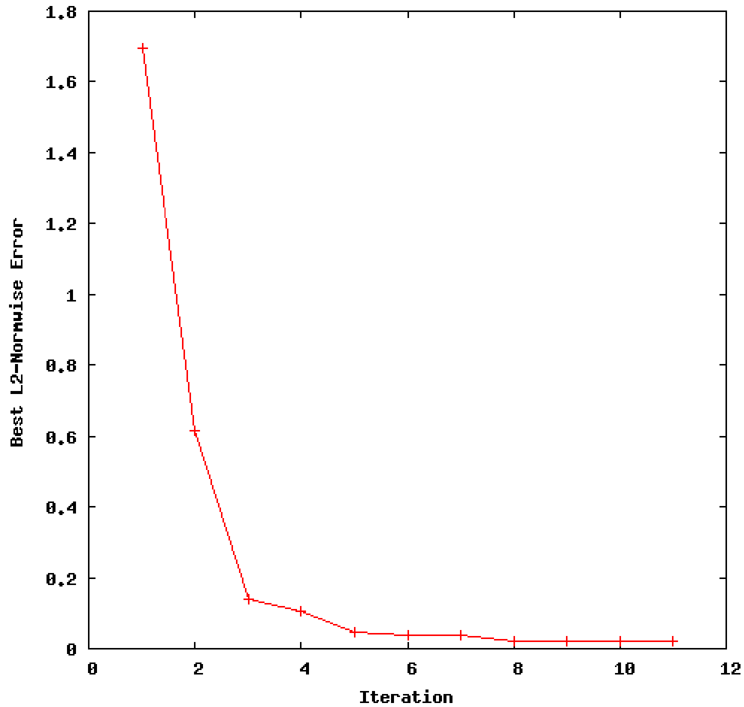

Figure 2. L2 norm of best swarm particle as a function of iteration, for the simulation study shown in Figure 1.

- **Pixel-level imaging fidelity assessment:** Work on large-scale image fidelity simulations have continued to explore the accuracy of approximate semi-analytic relations that are widely accepted in the field to describe these effects, but often significantly inaccurate due to the analytic intractability of the problem. In the current reporting period the focus has been on a more complete sampling of parameter space as this earlier work is finalized for publication.

- **Gravitational lensing and dark-matter substructure:** In the current reporting period I have initiated a research collaboration involving a graduate student (Di Wen; UIUC), fellow faculty at UIUC (Neal Dalal), and external collaborators (Yashar Hezaveh (Stanford)) to participate in simulation study of the impact of dark matter substructure on gravitational lensing measured using the ALMA array. The large-scale structure of the universe is well described by the Lambda Cold Dark Matter (ΛCDM) model. The observed structures, for example, galaxies and clusters of galaxies, form through a hierarchical process from the gravitational instabilities of matter density seeded by the primordial quantum fluctuations in density. At scales larger than the intergalactic distances, the predictions made by the ΛCDM model are verified observationally with multiple approaches. At sub-galactic scales, the ΛCDM model predicts a large number of low-mass subhalos orbiting around the galaxies, but observational evidence is lacking. This is called the Missing Satellite Problem.

The subhalos are too faint to be observed directly either because they are primarily made of dark matter, or because few stars forms in them. Gravitational lensing is particular suitable for detecting subhalos because any massive subhalo could leave a signature on the mass model of the lensing galaxy, including dark matter substructure. The searches for subhalos require very high resolution imaging of the lensing system, as well as accurate and precise models for the matter density profile of the lensing galaxy.

In this collaboration Di Wen has worked with Y. Hezaveh (Stanford) to run a Markov Chain Monte Carlo (MCMC) algorithm developed by Y.H to sample the high-dimensional parameter space describing the gravitational potential produced by the main lensing galaxy and a range of subhalos. Results for ALMA data on the strong lensing system SDP.81 in which dark matter substructure have been detected have been submitted to the Astrophysical Journal (Hezaveh et al. 2016). We are both co-authors on this paper. Di Wen will continue work in this area as part of his thesis research with me.

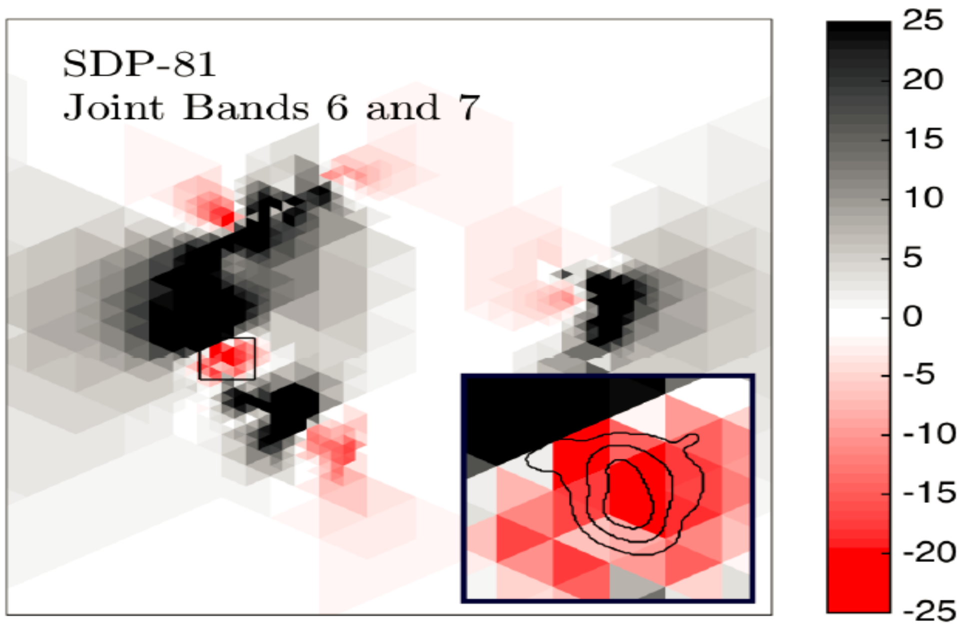

Figure 3. Figure 5 in Hezaveh et al. (2016), showing the subhalo isolation in observations of SDP.81 using ALMA. Values are proportional to the difference in log marginalized posterior probability density between a smooth lensing model with no sub-structure and a smooth model with a subhalo of mass $10^{8.6} M_{\odot}$ at the position indicated.

- **List of publications and presentations associated with this work**

Yashar D. Hezaveh, Neal Dalal, Daniel P. Marrone, Yao-Yuan Mao, Warren Morningstar, Di Wen, Roger D. Blandford, John E. Carlstrom, Christopher D. Fassnacht, Gilbert P. Holder, Athol Kemball, Philip J. Marshall, Norman Murray, Laurence Perreault Levasseur, Joaquin D. Vieira, Risa H. Wechsler **2016**, *Detection of lensing substructure using ALMA observations of the dusty galaxy SDP.81*, Astrophysical Journal (*submitted*) (http://arxiv.org/abs/1601.01388).

Several other papers in preparation.

- **Plan for next year**

In the current year, we have used approximately 94% of the 240 KNH allocated to project jq8. We have continued our strategy of using the generic software framework (eM) that encapsulates community codes to allow efficient prototyping in the initial stages of algorithm exploration. However, we have developed optimized MPI codes when needed (metaheuristic update step for novel interferometric calibration methods described above). When using community codes for prototype testing, we continue to rely on tmpfs file systems where implicit local disk assumptions are too expensive to remove from the current community code implementations. Similarly, we carefully consider I/O patterns when using these codes in order to match the Blue Waters architecture and ensure resiliency.

In the coming year, I would like to request 200,000 XE NH and 40,000 XK NH. This is based on current production measurements of both the fidelity assessment and metaheuristic calibration work. Runs using the eM community code framework extend to 800-1000 XE nodes at present, using 2 cores per decomposable application invocation and 32 GB per node. Larger scaling is planned for this coming year.

I would also like to request 500 PB of nearline project storage to allow a near research initiative involving re-analysis of large archival datasets using new algorithmic approaches. To allow efficient staging and processing of the nearline data I would also like to request 50 TB of Lustre project disk space.

The estimated utilization schedule for the requested allocation is uniform: Q1: 25%, Q2: 25%, Q3: 25%, Q4: 25%).