Annual Report for Blue Waters Professor 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 co-PIs and collaborators: Local research group.
  - Contact: akemball@illinois.edu

- **Executive summary (150 words)**

Contemporary astronomical interferometry for modern arrays such at the Atacama Large Millimeter Array (ALMA) and future instruments such as the Square Kilometer Array (SKA) present unsolved grand-challenge problems, now uniquely amenable to solution using extreme-scale computing. My research with Blue Waters has focused on two such problems, namely new techniques for pixel-level fidelity assessment in interferometric imaging, and new instrumental calibration techniques for modern arrays. Over the past year we have completed significant work in interferometric fidelity assessment. In addition, we have identified and evaluated completely novel methods of interferometric calibration at petascale, of broad potential scientific application to new interferometer arrays.

- **Description of research activities and results**

a) ***Key Challenges:*** Interferometry is a key technique that enables astronomical imaging at the highest angular resolutions. It is an inverse imaging method possible in wavelength ranges (radio, sub-millimeter, and some far-infrared wavelength windows) in which heterodyne (phase-preserving) signal conversion is possible. In interferometry the detected phase-coherent signal is cross-correlated to measure the spatial coherence (fringe visibility) of the incident radiation at all spatial separations sampled by the array configuration; these projected separations vary over time due to Earth rotation. Interferometry creates effective telescope apertures far exceeding any monolithic telescope diameters that could be physically constructed. The measured visibilities are connected to the astronomical source brightness distribution on the sky by an integral equation that reduces to a Fourier transform in the limit. However, the synthesized aperture is inherently sparsely sampled in interferometry, so the inverse problem is ill-posed and requires mathematical regularization for solution. In addition, the imaging integral equation contains terms representing atmospheric propagation and instrumental calibration errors at each telescope. These terms need to be jointly solved for along with the unknown image brightness distribution during inversion from the measured visibility data. A variety of best-practice heuristics have been developed in the community over a long period for this joint solution, typically using specialized incremental and iterative solution methods. Our project on Blue Waters has concerned several key related grand challenge problems in this domain. For this reporting period these have included: a) a means of interferometric inverse imaging and joint calibration that provides estimates of the image brightness uncertainty, such as bias and variance to first-order, at each pixel of the derived image; and, b) the development of uniquely new data reduction algorithms enabled for the first time by petascale computing.

b) ***Why it Matters*****:** Data rates in interferometry scale approximately as the square of the number of telescopes in the array multiplied collectively by exponential advances in data acquisition rates enabled by Moore's Law. Data rates for future arrays will render custom data reduction and imaging by individual principal investigators impractical. In addition, the niche nature of heuristics and custom techniques for PI-driven reduction present a barrier to access to users who are not expert in interferometry or close members of this community. As a result, automated pipeline data reduction workflows are increasingly critical. However, the scientific use and interpretation of their image products is only fully possible by non-expert users if accompanied by pixel-level uncertainty quantification. In addition, the new arrays, such as the Atacama Large Millimeter Array (ALMA) and the future Square Kilometer Array (SKA) require significant advances in current data calibration algorithms as they need to achieve much higher sensitivities.

c) ***Why Blue Waters:*** Both research problems (a) and (b) described under *Key Challenges* above require petascale computing resources. Our work on pixel-level fidelity assessment in interferometric calibration and imaging has utilized contemporary advances in computational statistics concerning resampling of data with long-range statistical dependence. In addition, we have augmented this work on Blue Waters with large-scale Monte-Carlo simulations of imaging fidelity for a range of instrumental errors with various statistical distributions. Given the data set sizes and large-dimension parameter space, these techniques are not easily pursued on systems smaller than Blue Waters. This is especially true for the new algorithms we are exploring for interferometric calibration on Blue Waters – these are novel methods completely impractical on earlier systems.

d) ***Accomplishments:*** Our accomplishments on Blue Waters to this reporting date are summarized below:

   i) **Code migration and development:** Our work utilizes an internal code base (eM) that combines key community codes for interferometric calibration and imaging in a framework that allows parallel deployment and custom research development. This code, which comprises C++, C, Fortran, and python, was previously developed using the Intel software development compilers and tools and has been used on a range of HPC systems. As part of our initial work on Blue Waters we migrated the code to the GNU compiler suite (as the Intel compiler was not available early) and adapted it for the Linux and software stack variations on Blue Waters. As a research customization of large community packages, the code base is large and of order 1-2 MSLOC in C++. Porting the C++ code component was a moderate effort given the usual differences with C++ template instantiation, Koenig lookup, non-circular header file partition, namespaces, and C++ standard library compliance encountered with legacy C++ code bases and evolving compilers. We also migrated and updated our job control scripts to Blue Waters as part of this work. As several community codes are legacy packages and were written implicitly assuming a least one local fast disk, several adjustments were made where these packages were used to adjust their I/O access patterns to match the global file system on Blue Waters. In addition, substantial new features were added in the area of calibration algorithm research development during this period.

ii) **Pixel-level imaging fidelity assessment:** We have been able to conduct a series of large-scale image fidelity simulations to explore the accuracy of approximate semi-analytic relations that are widely accepted in the field, but often significantly inaccurate due to the analytic intractability of the problem. Fig. 1 and Fig. 2 show some results of full statistical simulations for different array sizes and instrumental phase and amplitude error distributions. Our work on Blue Waters has provided key new insights into the long-accepted models for imaging fidelity, and how they should be revised. This work will be described more completely in an upcoming paper.

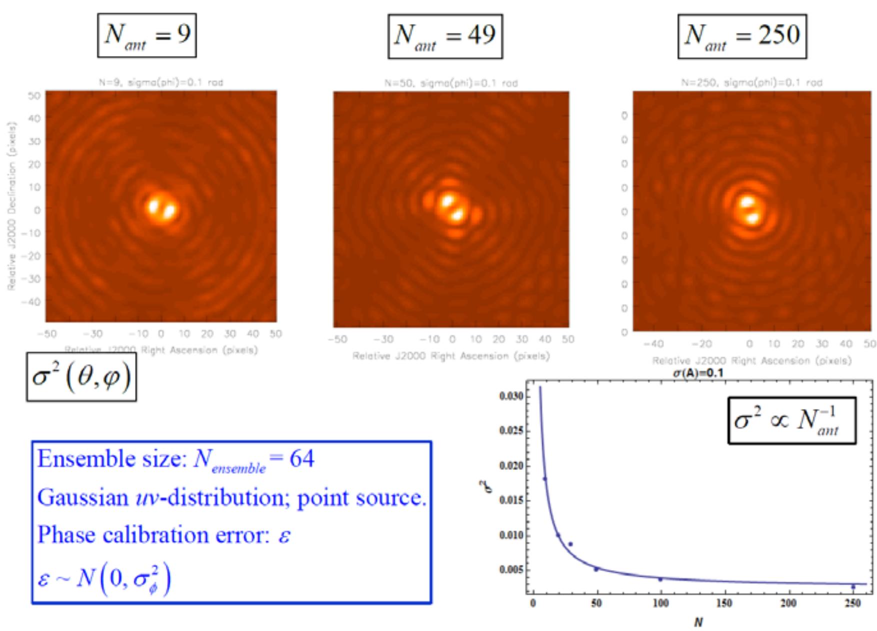

Figure 1. Statistical imaging simulation of interferometric phase errors.

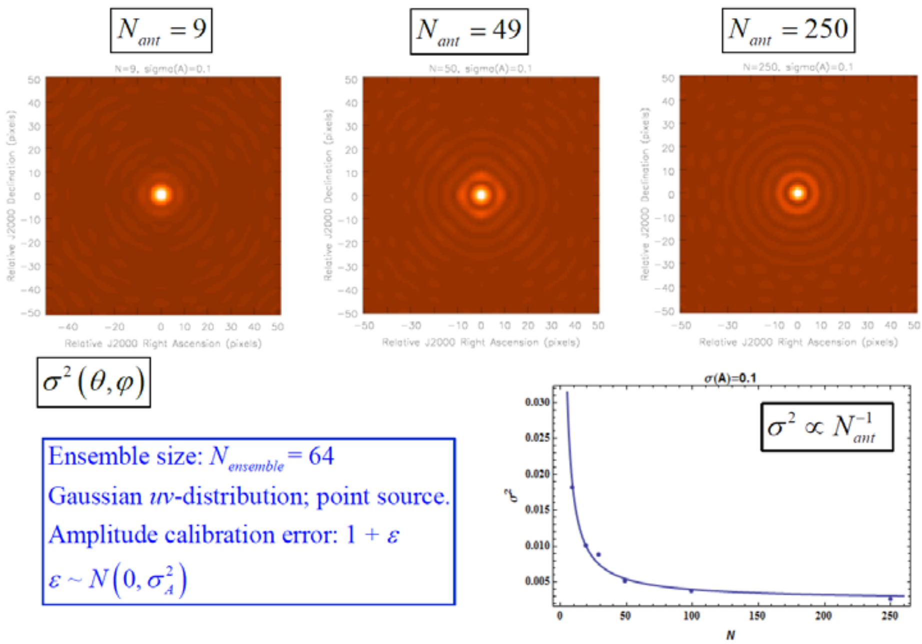

Figure 2. Statistical imaging simulation of amplitude errors.

iii) **Novel interferometric calibration algorithms:** In this area, we have worked intensively on the application of highly computationally-intensive calibration techniques in interferometry. These have included metaheuristic and stochastic methods that have unique scientific advantages in this regime but have not been possible until the advent of systems on the scale of Blue Waters. Part of this work has included the identification of new objective metric functions for these optimization methods that are required for these new approaches (two examples out of many are shown in Fig. 3). Blue Waters has allowed this work for the first time.

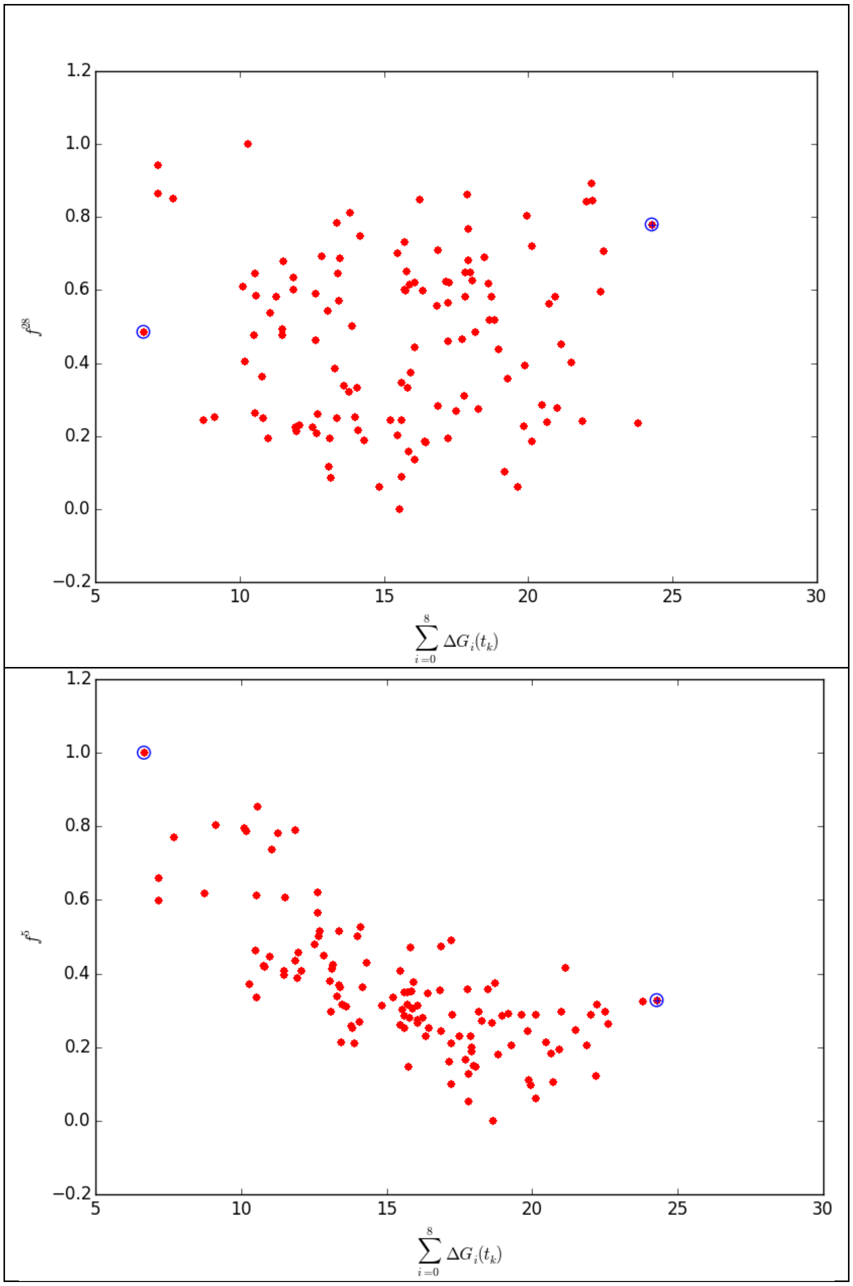

Figure 3. Objective function evaluation for novel interferometric calibration: poor objective function (*above)*, and good objective function (*below*). These can only be evaluated by direct numerical simulation given the analytic intractability of the problem.

- # List of publications and presentations associated with this work

Seidel, E. and Allen, G. and Freemon, M. and Gammie, C.F. and Kemball, A.J. and Pennington, R. and Petravick, D., 2014, Computing and Data Challenges for Multi-Messenger Astronomy, Exascale Radio Astronomy Conference, 20001, (http://adsabs.harvard.edu/abs/2014era..conf20001S).

Several papers are in preparation based on our current work with Blue Waters.

- **Plan for next year:**

Our work to date has used time under an exploratory allocation (ILL_jog; fully used) and a BWP allocation (ILL_jq8) for a total of approximately 100 KNH. We have carefully structured our research to use community implementations within the eM framework in initial stages of algorithm exploration, augmenting with specialized and optimized implementations when proven. Our work thus far has used of order 600-1000 nodes XE per run, using 2 cores per decomposable application invocation, 32 GB per node, as tmpfs file systems are used where implicit local disk assumptions are too expensive to remove in current community code implementations.

Our work in the current year will require an expanded number of runs of the current type as our research and code development proceeds to the final stages of algorithm evaluation. All essential code development for this phase is now complete. We extrapolate this to a total of 200,000 NH on XE nodes, using 32 GB per node, and 2 cores per parallel application invocation. We expect to move to a higher node count usage by correcting some final I/O optimizations in legacy components. The calibration methods chosen are targeted at GPU deployment in Phase II in highly optimized form; for this phase of anticipated work in the current year we request 40,000 NH on XK nodes.

We request 500 TB of nearline project storage due to our anticipated future use of much large interferometric archives than currently used.

Our estimated use of the requested allocation is: Q1: 25%, Q2: 25%, Q3: 25%, Q4: 25%, but the XK nodes will be used later in the current year.
