**Julianne Chung**
Emory University

High Performance 3-D Image Reconstruction for Molecular Structure Determination

DOE CSGF

MOTIVATION: Accurate volume reconstructions from cryo-electron microscopy (cryo-EM) allow scientists to study protein structures, viruses and other complex molecular structures. Given a collection of 2-D projection images from various angular orientations, the goal of the reconstruction process is to recover a 3-D volume representation.

THE COMPUTATIONAL CHALLENGE: In cryo-EM microscopy imaging, high resolution is achieved by using a low electron dose, resulting in low signal to noise ratio in the projection images. Thus, a large number of projection images are needed for good reconstruction. Furthermore, reconstruction of large-volume structures can be burdensome, due to massive data and memory requirements. Current parallel implementations of reconstruction algorithms are inadequate for computing large-volume macromolecule structures, even on today's state-of-the-art high performance supercomputers.

SUMMARY OF RESEARCH: A more efficient MPI parallel implementation allows both the image and volume data to be distributed among processors of a parallel computer, thereby overcoming the current per-processor memory limitations. Use of the built-in MPI virtual Cartesian topology functionality simplifies coding and allows the choice of processor decomposition to depend on the details of the underlying hardware. In addition, we study a variety of algorithms for volume reconstruction, proposing a Lanczos-based reconstruction algorithm for cryo-EM data. Compared to standard methods such as SIRT (Simultaneous Iterative Reconstruction Technique) or LSQR (Least Squares QR), this algorithm is less sensitive to noise in the problem and computes better reconstructions in fewer iterations.
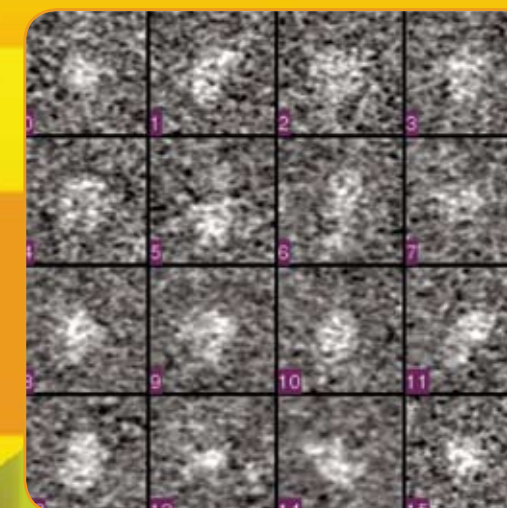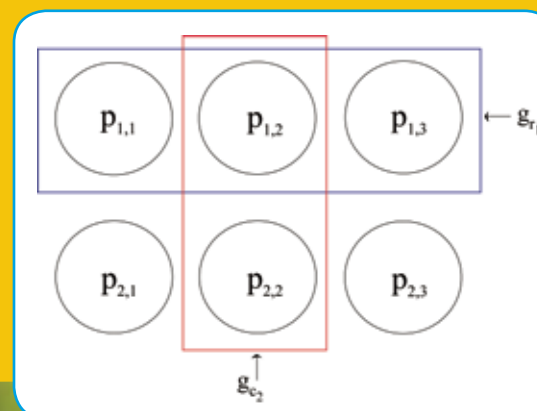
Sample projection images from real cryo-EM data. Usually we need thousands of these images to reconstruct a good volume.

Isosurface rendering of a reconstructed TFIID protein using the new Lanczos-based iterative algorithm.

Processor layout for volume data distribution. Each processor is identified by its row and column group numbers.