

DATA SCIENCE IN HIGH RESOLUTION IMAGING OF SAMPLE HETEROGENEITY

Duane Loh, Colin Teo, Abhik Datta, Zhou Shen and Deepan Balakrishnan

Abstract: The advent of brighter sources and fast direct-detection detectors in x-ray and electron microscopy are rapidly changing high resolution, dynamic imaging. The most direct impact is the orders of magnitude increase in the number of measurements compared to what was previously possible. With vastly more measurements experimenters can access not just the average signal produced by their samples, but also the statistically significant fluctuations that reveal meaningful sample heterogeneity. Since these large number of measurements on Nature are unavoidably noisy and incomplete, the classifications and descriptions of her heterogeneity are also largely probabilistic. As a consequence, the analyses of such measurements is highly data-centric, where each measurement has to be considered many times before the analyses converges. In this talk, I will address some of the challenges in acquiring, analyzing, filtering, and archiving noisy and incomplete data from these direct detectors at the Centre for Bio-imaging Sciences at the National University of Singapore. I will also share our limited experience in using high performance computing to analyze such data, validate its interpretations, and help drive experiment design.

Author keywords: data science bayesian classification x-ray diffraction imaging electron microscopy heterogeneity x-ray free-electron lasers direct electron detectors single particle imaging