Hi I am a PhD student in Seismology at Berkeley Seismological Laboratory. My work involves processing data recorded by seismometers, both earthquake shaking and background noise for different objectives. Most of the methods involve time-series analysis tools, processing in spectral domain (FFT, smoothing), matrix manipulation, etc. In the last decade, there has been a substantial growth of seismic networks and seismic data collected, therefore, many previously used methods and codes need to be parallelized now or have been parallelized recently. Simulating wave propagation in 3D domain by numerical methods also needs clusters and supercomputers, where each node solves the wave equations in a specific part of domain. In my field, knowledge of parallel computing is a very important skill to have and in this course, I hope to learn the necessary details and parameters that go into writing simple parallel codes.

For example, this paper describes methodology for making cross-correlation of background seismic noise parallel.


For N (~100 to 800) number of time-series, we cross-correlate (a signal processing operation) each time-series (let’s say 36000 floating point samples in an hour) with all other time-series. Then, this process has to be repeated over all hours in a day and all cross-correlations in a day are then averaged. Then, this is repeated over months/years of data. Since there is a lot of data even for an hour, it has to be divided into groups and shuffled around different nodes (distributed memory platform). While there is a C++ code already available as described in the paper, we have a serial C code that works on different data formats, diverse quality of data and uses slightly different methodology than what is described in the paper and I would like to modify it for parallel processing on a local cluster in our lab. I am open to other problems for projects as well. Thanks.