Pre-proposal

Thesis: Hybrid MPI + OpenMP Programming for Pre-Sequencing Simulation

Problem Description: The problem is encountered when I justified my results for a theoretical research on statistical properties of pre-sequencing behavior. A highly simplified version of the simulation (enough for parallelism) can be briefly described as following: Experiment begins with N molecules evenly distributed into M wells. Then each molecule separately duplicates C_1 times, and a dilution will be done in each well to sample K molecules per well. After first dilution, each molecule again duplicates separately C_2 time and then mixed all wells together to sample S molecules. In every duplication, mutation may happen in certain probability. We shall produce simulation for final S molecules, each with information about wells it comes from and mutation counts it has. Then main problem here is that S and K*M are typically very large, which might cost hundreds of gigabytes for memory. Therefore, I suggest using a Hybrid MPI + OpenMP solution. Using MPI to distribute work to separate computing elements to reduce memory requirement and within each element do OpenMP to improve performance compared to pure MPI code.

To Do: (1) Derive and test Amdahl's law for this Hybrid model. (2) Compare Hybrid solution to Pure MPI and sequential version to study the performance. (3) Some discussion on the result and implementation.

Members of Team:

Bohao Tang	btang11@jhmi.edu
Zhengzhong Jin	zjin12@jhu.edu

Reference:

Best Practice Guide to Hybrid MPI + OpenMP Programming, Version1.1, 30th March 2017