Project 7B: Autocorrelation Using MPI

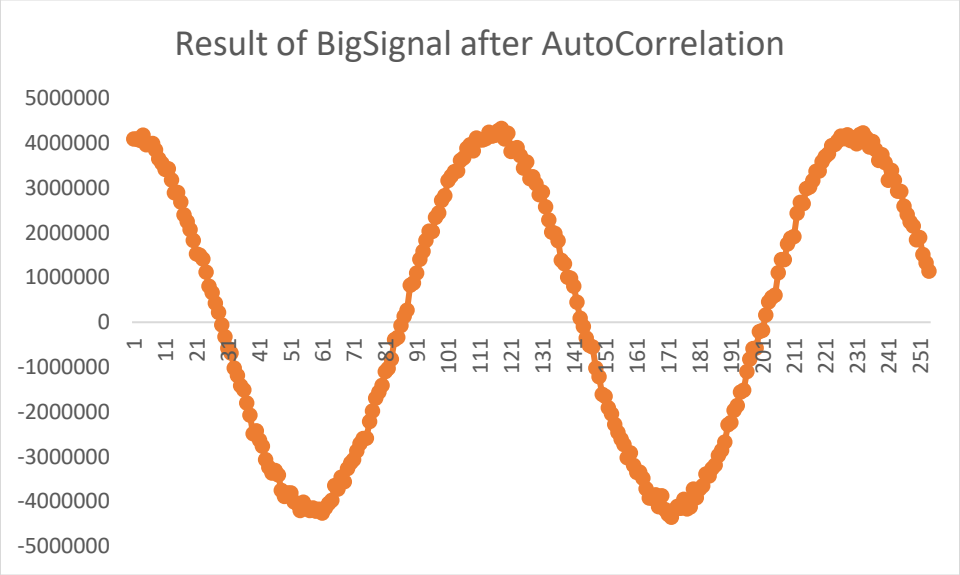
Andre Ortega

ortegaan@oregonstate.edu

Introduction

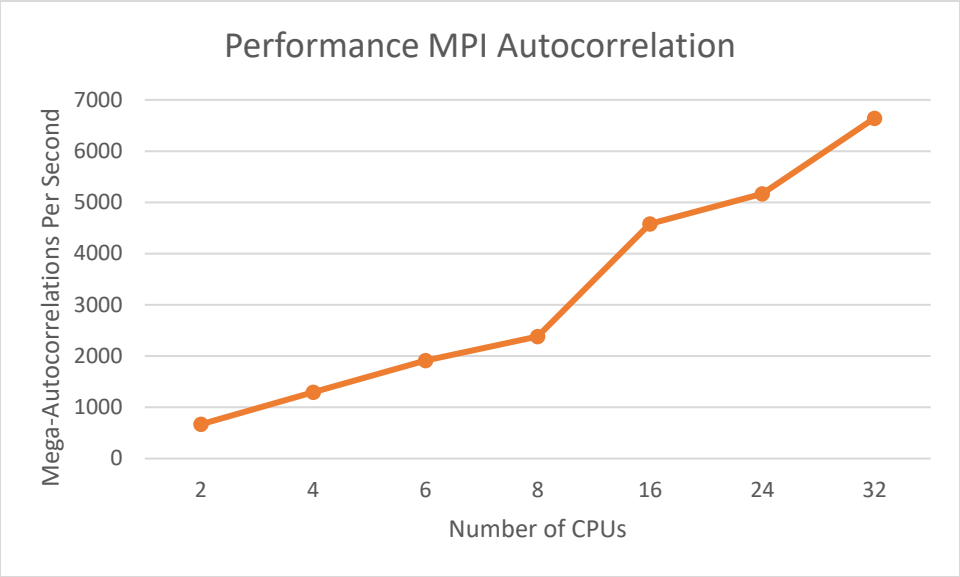
A program was constructed to find the hidden correlation of a noisy signal using a provided method of autocorrelation with an implementation of MPI to increase performance. An increasing number between 2 and 32 CPUs were used to graph the increases of this benchmarking. The results of the main algorithm itself - the sums[*] array - shows the clear result of the noisy signal after finding the pattern using autocorrelation.

Chart: Sine Wave



The period of a sine wave is the amount of units it takes to complete one wave cycle. This graph is the result of the sums of the noisy signal when added to a delayed value of itself, by how much the signal was shifted before the addition. As we can see, this produces a very noticeable pattern that would otherwise be difficult to decipher. The period of this resulting sine wave is 114.

Chart: Performance



Here we can observe the performance of the autocorrelation program very quickly reach computation speeds in the giga range when using multiple CPUs in parallel for processing parts of the same large data set. The biggest increases in performance can be observed the moment we begin utilizing either 16 or 32 CPUs in parallel, which handedly correspond to our idealized data set size of 8Mb, which is easily distributed for equal use among multiple processors that will each receive an optimal local array size for computation on a binary system.

Conclusion

The Message Passing Interface is a fantastic uppermost method of data parallelism using distributed CPUs synchronized in a cluster, which can be paired or partnered with multiple other types of data parallelism for even greater performance boosts within each distributed system. For this use case, a binary file containing an apparently random and noisy signal was read and processed using autocorrelation in order to find a hidden pattern that would otherwise not be easily apparent. After processing, an easily visible wave was produced in a CSV data file and graphed as a chart of the sum of the shifted values vs how much the data was shifted. When patterned in this way, the result was a clear wave that had a period of 114. The performance of this autocorrelation program increased with each added cluster of CPUs, with the greatest performance benefits shown when using 16 or 32 CPUs.