rEDM Code Acceleration with ABCI Supercomputer

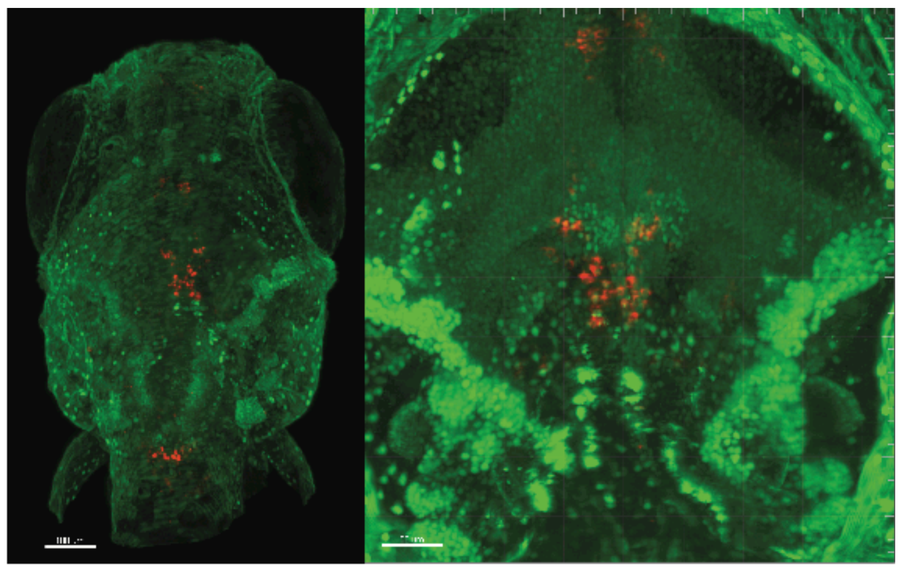
Wassapon Watanakeesuntorn, Kohei Ichikawa, Jason Haga, Gerald Pao, Erik Saberski

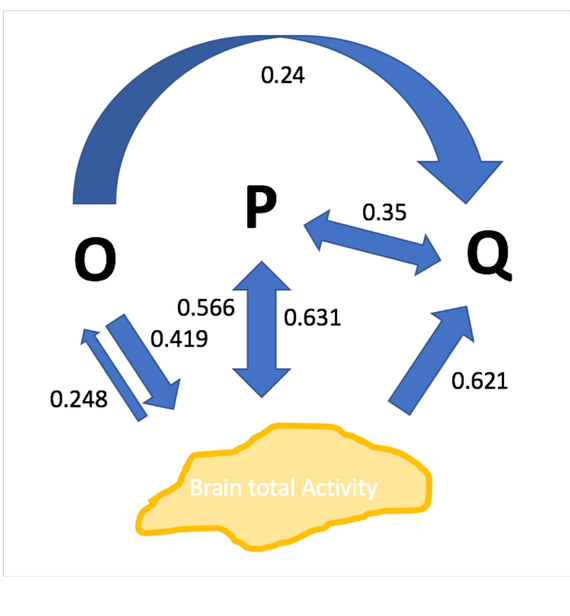
⚓ Introduction

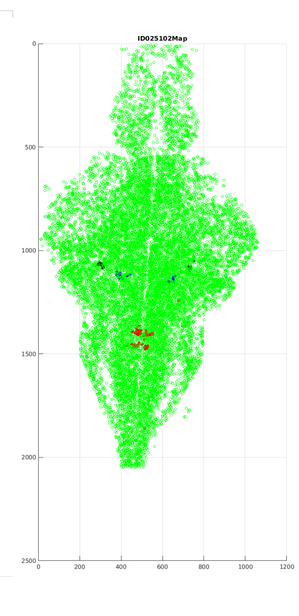
rEDM is an R package mainly written in C++ with R programming language wrappers. The rEDM package is a collection of methods for Empirical Dynamic Modeling (EDM), a suite of mathematical tools based on the generalized Takens embedding theorem for the analysis of nonlinear time series. Among these, Convergent Cross Mapping (CCM) allows the inference of causation from nonlinear time series even with substantial noise and complete absence of correlation

Zebrafish neural activity maps for novel Neuromorphic Deep Learning Architectures is a project from Scripps Institution of Oceanography, UC San Diego, and the Salk Institute in the United States which collects and analyzes Zebrafish neural activity for use with CCM and other tools from empirical dynamical modeling to find the relationships within the neural activity network of the transparent larval fish brain. Multiple datasets of neural activity were collected. Each dataset is around 1,600 timesteps in length sampled at 2 hertz containing 70,000-80,000 active neurons in each case. As such this becomes a big data problem for the rEDM package.

We aim to accelerate the rEDM code by using GPUs and CUDA because rEDM vector calculations should be compatible with GPU parallelization and computation. The goal is to adapt these calculations with the AI Bridging Infrastructure (ABCI) architecture which is the most powerful supercomputer in Japan. ABCI is operated by National Institute of Advanced Industrial Science and Technology (AIST).







⚓ Approaches

Our goal is to run the CCM calculations of the complete zebrafish brain datasets using the ABCI supercomputer. Our proposed development milestones are:

1. CUDA enabling the rEDM code
2. Containerizing with singularity and testing the code on ABCI
3. Execute one complete CCM run on a zebrafish brain dataset
4. Run every zebrafish brain dataset on ABCI

⚓ Development and Progress

We have to investigate the rEDM source code. The main computation part of rEDM code is on the “forecast_machine.cpp” which contains the vector calculation part. We try to put some function into CUDA. However, some of the functions are memory intensive tasks which are not compatible with the GPU architecture. The computation intensive tasks on this source code are compute vector distance functions and vector sort functions.

The truncated version of the dataset, which contains 154 neurons with 1,600 timesteps, is used to evaluate the performance. Our development machine has 2 Intel Xeon E5-2670v2 CPUs and 3 NVIDIA GTX Titan GPUs. We run the dataset with original code which it takes 1 hour and 30 minutes to finish the computation. But, it takes more than 2 hours after we apply CUDA on vector sort function which it comes from data transferring overhead between RAM and GPU memory. This means GPU calculation is not worth to tradeoff with data transferring overhead for sorting vector, when the size of vector is 1,600, on this dataset. After that, we tried to optimize the code to eliminate and reduce the loop in the code which it makes the code run in about 3 minutes to complete the calculation of CCM with the truncated dataset using only the CPU. However, the CPU usage is not efficient when we run the code with the real dataset which it contains 1,600 timesteps with more than 50,000 neurons. Currently, we are trying to enable CUDA on the vector distance computation function. Additionally, we will continue to investigate the rEDM source code to accelerate the computation.