Massively Parallel Empirical Dynamic Cross Mapping

Joseph Park[1], Gerald M Pao[2], Erik Saberski[3], Cameron Smith[3], Jason Haga[4], Ryousei Takano[4], Chen Min Yeh[2], Sreekanth Chalasani[2], George Sugihara[3]
1.U.S. Department of Interior, JosephPark@IEEE.org 2. Salk Institute for Biological Studies, pao@Salk.edu
3. Scripps Institution of Oceanography, UC San Diego, gsugihara@ucsd.edu 4. National Institute for Advanced Industrial Science and Technology (AIST), abci-sp@aist.go.jp

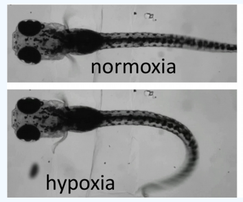
Synopsis

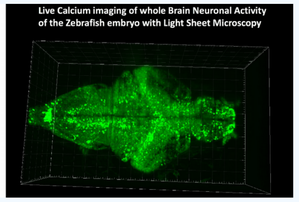
Empirical dynamic modeling (EDM) is an emerging data-driven framework for modeling nonlinear dynamic systems based on the mathematical theory of reconstructing system attractors from time series data (Takens 1981). Cross mapping is an EDM technique to identify dynamical similarities, and can be used to indicate coherent and casual interactions with no model assumptions.

We use a whole brain recording of approximately 13.5 minutes of larval zebrafish (Danio rerio) performing a low oxygen avoidance task imaged by SPIM light sheet microscopy where neural activity is obtained from a transgenic fluorescent calcium indicator GCAMP6 at a sample rate of 2 Hz. Activities of ~53K individual neurons were recorded and converted into time series for EDM causal inference analysis.

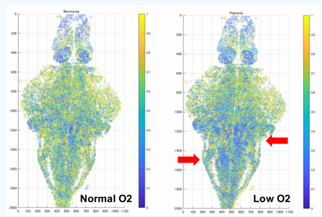
The analysis of 53K timeseries requires O(3E9) cross map computations, each of which entails a 20 dimensional embedding and simplex algorithm forecast at a rate of roughly 70 cross maps per second for a total computational time of 465 days. The numerical approach is to leverage the massively parallel architecture of the AI Bridging Cloud Infrastructure (ABCI) supercomputer. The application is a multithreaded C++ code based on the cppEDM library developed at UCSD/SIO Sugihara Lab. Each ABCI node supports 80 threads, and the data are decomposed across 510 nodes using the Univa Grid Engine. A full brain dynamical dimension estimate and cross-map was achieved in 8.5 hours.

Larval zebrafish hypoxia escape response





Transgenic larval zebrafish expressing histone GCAMP6-Histone H2A were head fixed and imaged in custom built microfluidic chambers by light sheet microscopy. 120K neurons. 50-80K active in most experiments.



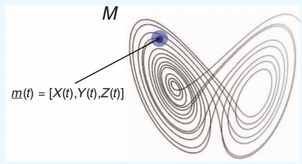
Whole brain dimensionality at single neuron resolution

Emergence of low-dimensional order in response to environmental stress

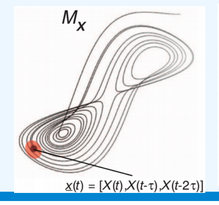
Blue: low dimensional
Yellow: high dimensional

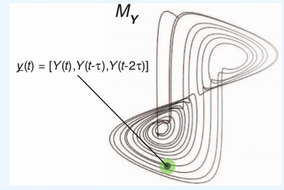
Cross Mapping : Coherent Dynamics & Causality

Convergent cross mapping (CCM) tests for correspondence between projected manifolds. Consider an attractor manifold of the observed system M = F(X,Y,Z) and two projected manifolds, M_X and M_Y constructed using lagged-coordinate embeddings of X and Y.



Since X and Y are dynamically coupled points that are nearby on M_X (red ellipse) will correspond temporally to points that are nearby on M_Y (green circle). This enables us to infer states across manifolds using Y to estimate the state of X and vice versa using manifold nearest neighbors.

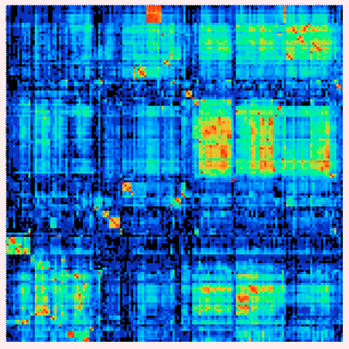




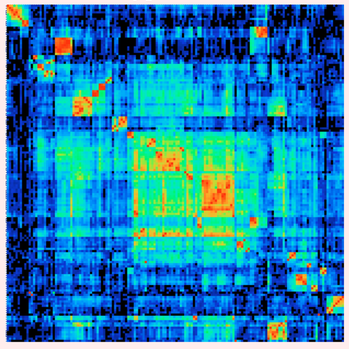
Granger causality assessed through predictive correleation across states provides X:Y cross map interaction amplitude.

Emergence of cross neuronal coherent dynamics in 96 dominant mid and hindbrain neurons during hypoxia

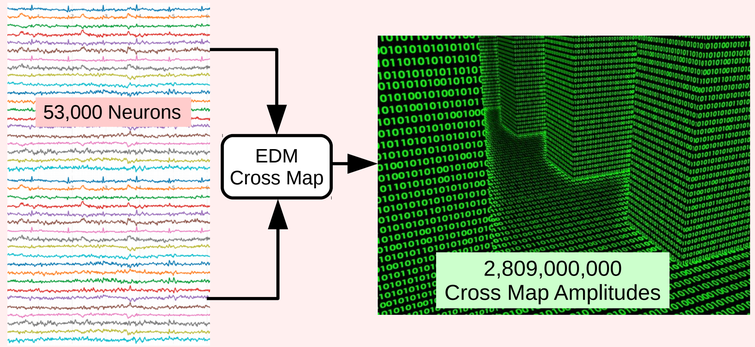
Normoxia



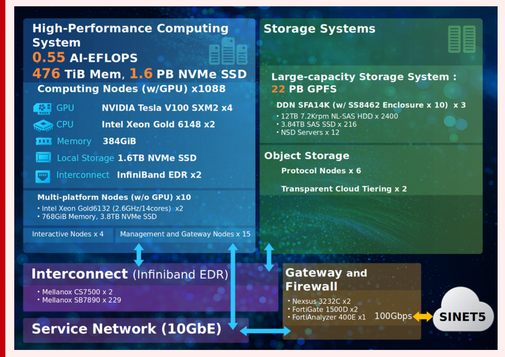
Hypoxia

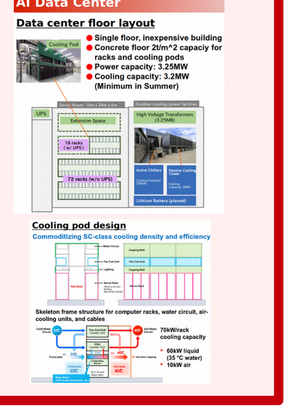


Can we scale the cross mapped coherent dynamics from 96 to 53,000 neurons?



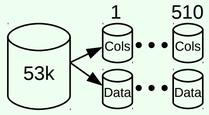
ABCI AI Bridging Cloud Infrastructure





Information: http://abci.ai/ Contact: abci-sp@aist.go.jp

Decompose 53K time series Across 510 compute nodes

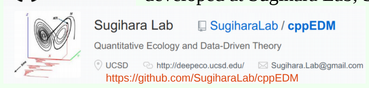


Each column file specifies data columns for 5,518,760 cross maps per compute node.

UNIVA grid engine

Job scheduler distributes and executes node application across 510 nodes

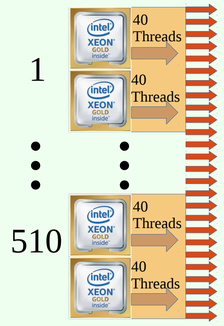
GitHub CppEDM C++ library of EDM tools developed at Sugihara Lab, SIO, UCSD




```
// BlockCM-Node.cc Psuedo Code
main() {
    df = Read.Data.Files(); // cppEDM DataFrame class

    sf = SyncFile(); // Synchronized file class

    block = Make.Block( df ); // Embed data to Emax = 20

    E_opt_WorkQueue( df.columns ); // E_opt Work Queue
    CM_WorkQueue   ( nodes );      // Cross Map Work Queue

    E_opt_Threads( block, E_opt_Map, WorkQ ); // 80 threads

    CrossMap_Threads( sf, E_opt_Map, WorkQ ); // 80 threads
}

void E_opt_Thread(block, E_opt_Map, WorkQ ) {
    while( workQ ) {
        E_opt_Map[ workQ ] = Compute_E_opt();
    }
}

void CrossMapThread( sf, E_opt_Map, WorkQ ) {
    while( workQ ) {
        CM = CrossMap( E_opt_Map[ workQ ];
        sf.BufferData( CM );
    }
}
```

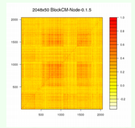
cppEDM Application: BlockCM-Node.cc

Full CPU Thread Utilisation



80 Threads x 510 Compute Nodes

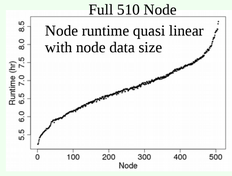
40,800 Concurrent Threads

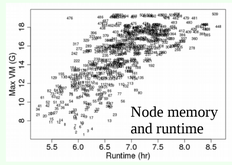


Test Case 2048x2048 Cross Map: 10 min

Multithreaded disk output is buffered with an asynchronous shared pointer file class to avoid I/O resource contention.

Full 510 Node





Results

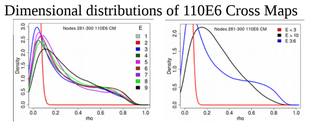
The first vertebrate full-brain Infonome.

53,052 20-Dimensional Embeddings with Dimension Estimate and 2.81E9 EDM Cross Maps in 8.5 hours.

The largest EDM Cross Map to date.

Big Data Analysis is underway.

Dimensional distributions of 110E6 Cross Maps



For strong interactions (ρ > 0.7) in low dimensional space (E 3:6):

30% of available connections used.
55% of source neurons coupled to 38% of targeted neurons.

Hypoxic infonome is sparse relative to all nueronal/interaction capacity.

Acknowledgements
Computational resources provided by National Institute of Advanced Industrial Science and Technology (AIST) AI Bridging Cloud Infrastructure (ABCI).