Project ID: jh220050（国際課題）

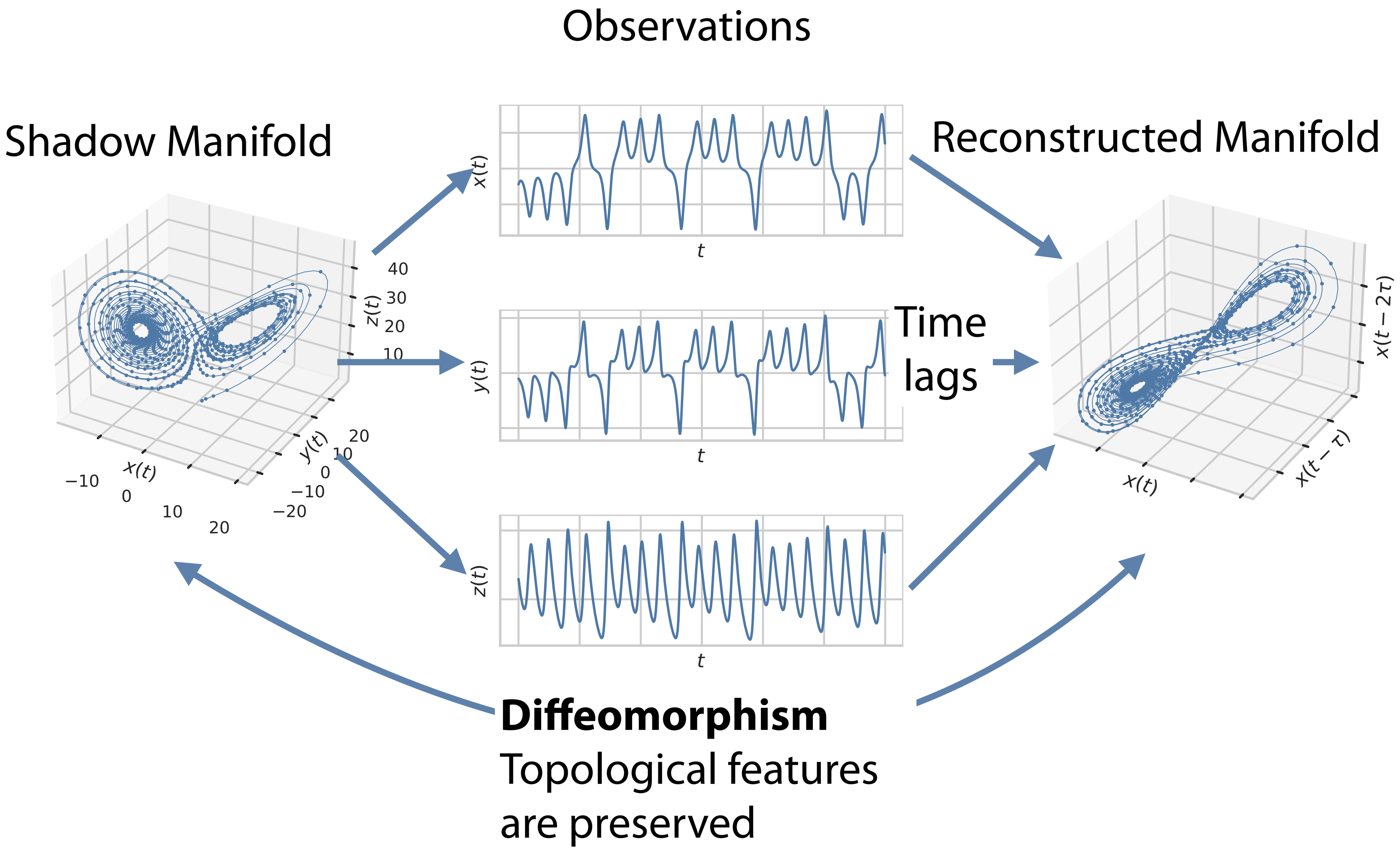
Implementation and Application of High-Performance Empirical Dynamic Modeling

Keichi Takahashi[1], Gerald M. Pao[2], Wassapon Watanakeesuntorn[3]

[1] Cyberscience Center, Tohoku University, Japan keichi@tohoku.ac.jp
[2] Salk Institute for Biological Studies, USA
[3] Nara Institute of Science and Technology, Japan

Background

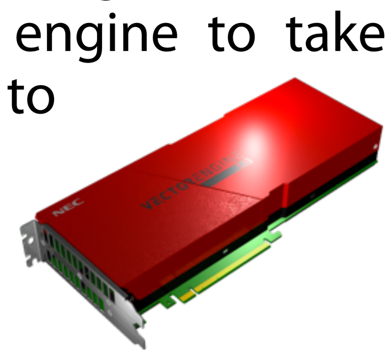
We have been developing a high-performance implementation of Empirical Dynamic Modeling (EDM), **an emerging framework for non-linear time series analysis**, and applying it to large-scale datasets. EDM enables a variety of analyses such as short-term forecasts, quantification of non-linearity, and causal inference. These analyses are achieved by reconstructing the latent dynamics behind the data without assuming a parametric model or using prior knowledge (thus the name *empirical*).



Objective

Although EDM is a generic method for modeling non-linear time series data, it was originally developed in the field of ecology. **Existing libraries for EDM analysis are thus designed to target a small number of short time series**. While recent studies have applied EDM to datasets including neural activity, gene expression and meteorology, the lack of a high-performance implementation limits the scale of datasets that can be analyzed. To handle large-scale datasets, we have been developing a high-performance EDM libraries [1, 2]. In this research, we continue this effort by tackling the following challenges:

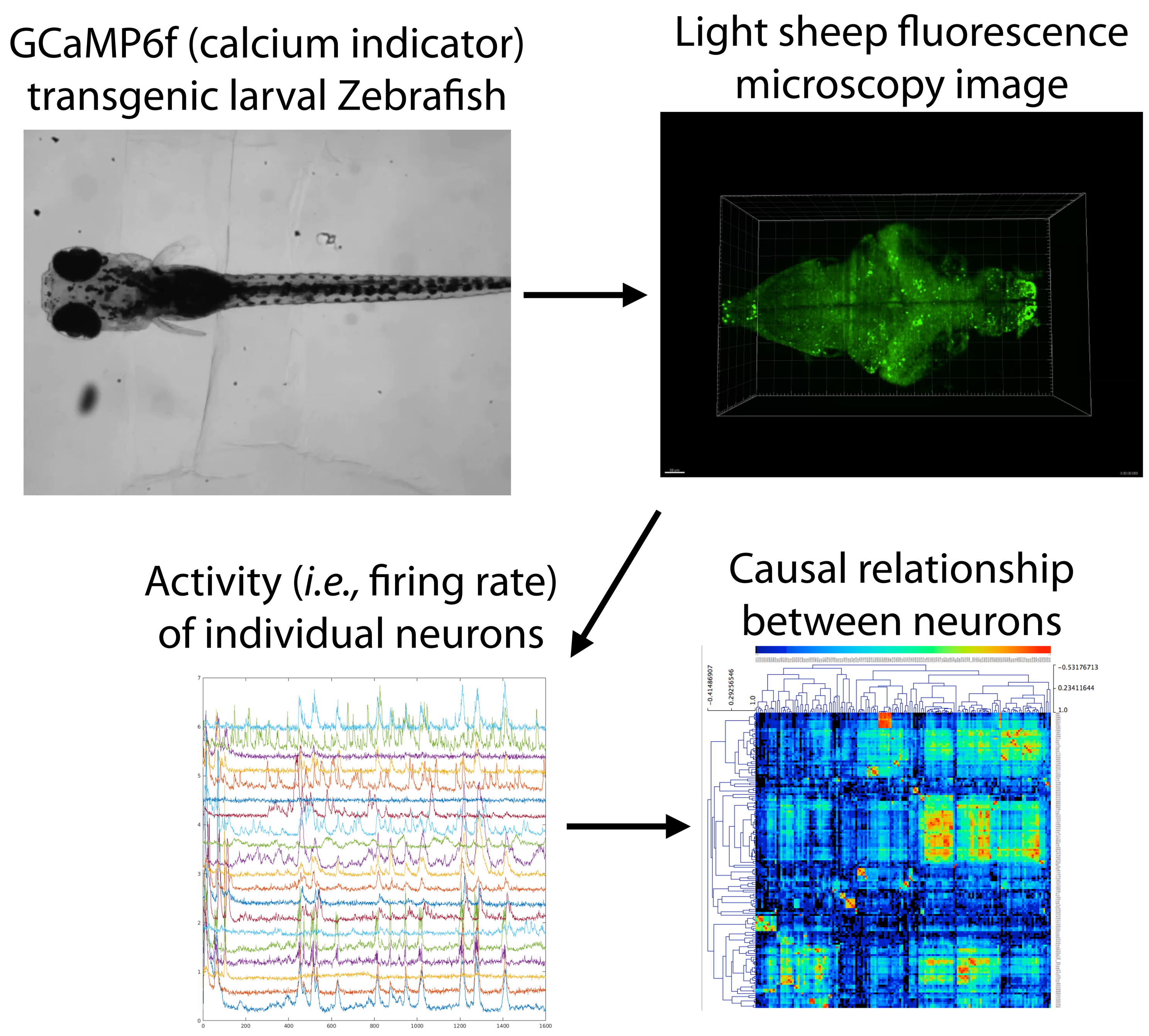
1. **Porting our EDM implementation to the SX-Aurora TSUBASA vector engine:** Since EDM is a highly memory-bound algorithm, we will port EDM to the SX-Aurora TSUBASA vector engine to take advantage of its massive memory bandwidth. We plan to use our implementation of EDM named kEDM (https://github.com/keichi/kEDM) based on the *Kokkos* performance portability framework.



2. **Enhancing the scalability of EDM by utilizing approximation algorithms:** A time-consuming kernel in EDM is the All *k*-Nearest Neighbor (A*k*NN) Search in the state space. Existing implementations use a brute-force algorithm, but this is clearly not scalable. We will thus integrate various approximate *k*-NN search algorithms (spatial trees, proximity graphs, product quantization, *etc.*) into EDM and evaluate which algorithm is best suited.
3. **Analyzing neural activity datasets to evaluate the performance of the ported implementation:** Our main goal is to perform EDM causal inference between neurons in an animal brain and elucidate how each individual neuron interacts with one another.

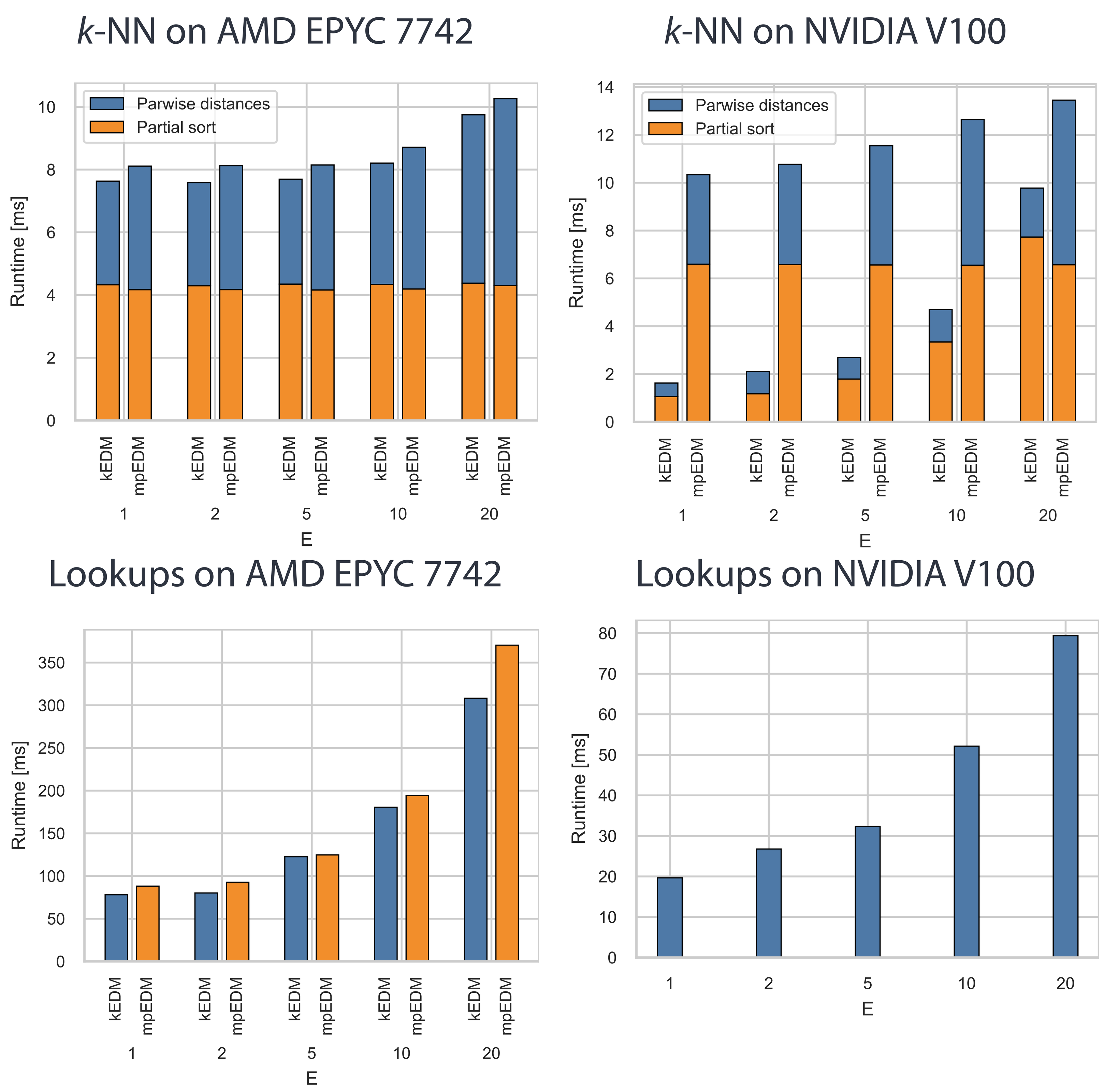
Target Application

We aim to analyze **the neural activity of an entire larval zebrafish brain that we recored at singe-neuron resolution** using light sheet fluorescence microscopy. So far, we identified a manifold that predicts turns of the fish at least 0.5 seconds ahead of time. Whenever the neural activity trajectory enters one of the loops of the manifold, the fish will turn.



Preliminary Performance Results

Using our EDM implementations mpEDM and kEDM, we have conducted preliminary performance measurements and roofline analysis on AMD EPYC 7742 and NVIDIA V100.



References

[1] K. Takahashi, W. Watanakeesuntorn, K. Ichikawa, J. Park, R. Takano, J. Haga, G. Sugihara, G. M. Pao, "kEDM: A Performance-portable Implementation of Empirical Dynamic Modeling using Kokkos", PEARC 2021, Jul. 2021.
[2] W. Watanakeesuntorn, K. Takahashi, K. Ichikawa, J. Park, G. Sugihara, R. Takano, J. Haga, G. M. Pao, "Massively Parallel Causal Inference of Whole Brain Dynamics at Single Neuron Resolution", ICPADS 2020, Dec. 2020.