Objectives and Specific Aims

Joaquín Rapela and Maneesh Sahani

Gatsby Computational Neuroscience Unit

March 11, 2025

Contents

Objectives

1. **Co-develop a new type of neuroscience experimentation** that enables animals to express ethologically relevant behaviours over extended periods in naturalistic environments, while capturing high-precision behavioural and neural activity measurements.
2. **Support research groups worldwide in adopting this new experimental framework** by providing comprehensive documentation on the hardware and software necessary to implement these experiments.
3. **Enable global access to the data generated by these experiments** by developing web-based platforms for data access, visualisation, and analysis, while also allowing users to conduct custom analysis.

Contents

Specific Aims

1. **Extend the AEON hardware and software infrastructure** to support new long-duration experiments, starting with novel olfactory learning experiments at AIND. Fields project?.
2. **Develop software for neural and behavioural time-series visual exploration**, enabling continuous and event-aligned views. (Neurosift, DENDRO)
3. **Design advanced statistical data analysis methods** for very long-duration and non-stationary time-series.
4. **Build real-time machine learning algorithms** to enable intelligent neural manipulations.
5. **Develop batch and online spike sorting methods** to process weeks- to month-long continual electrophysiology recordings and support real-time machine learning inferences.

Advanced statistical data analysis methods

Challenges:

long-duration : batch $\rightarrow$ online

non-stationarity : alternatives

1. detect performance degradation $\rightarrow$ retrain
2. use non-stationary methods

Advanced statistical data analysis methods

For behavioural data, we will investigate methods to:

track multiple body parts of animals (e.g., [7] and a switching-linear-dynamical method using RFIDs that we will develop),

infer kinematics of foraging mice (e.g., [8,9]),

segment behaviour into discrete states (e.g., [10] and a hierarchical HMM that we will develop),

infer rules that govern mice behaviour from behavioural observations only (i.e., policy inference) (e.g., [11]).

For neural data, we will investigate methods to:

estimate low-dimensional continual representations of neural activity (i.e., latents inference) (e.g., [12]),

segment neural activity into discrete states (e.g., [13]),

decode environment variables from neural activity (e.g., [14]).

Contents

Use Case

A user begins by **visualising continuous behavioural measurements**, such as the kinematics (speed and acceleration) of a mouse during a months-long experiment measured by the IMU of an ONIX probe.
Next, she **examines the results of a machine learning analysis**, such as behavioural states inferred by a Switching Hidden Markov Model (SHMM) using kinematic data.
The SHMM was initially trained on the first two hours of the experiment and was periodically retrained to **adapt to non-stationarities**, such as sensor fluctuations, changes in motivation, fatigue, or learning.

Use Case

She then **visualises epoched data**, such as SHMM states aligned to key events—e.g., the onset of a foraging bout in the richer patch at a specific time of day.
Curious about the neural basis of these SHMM states, she checks the neural recordings but realises they have not yet been spike-sorted. She runs our **offline spike sorting method**, developed for very long continuous recordings, and performs quality control on its results.
Based on quality metrics, she detects drift in the recorded signal, adjusts the drift correction parameters, and reruns the sorting algorithm.

Use Case

After examining the sorted spikes from a large neural population, she finds it challenging to interpret activity across so many neurons. So she decides to summarise the population spiking activity by **estimating continuous latent variables**. She then returns to the behavioural visualisation software, integrates machine learning indices corresponding to these latent variables, and **visualises behavioural data aligned to the newly estimated latents**.

Use Case

Through these explorations, she hypothesises that a peak in a neural latent variable from the prefrontal cortex signals the moment when mice decide to begin a foraging bout.
To test this, she **runs an online machine learning model** to estimate latent variables from prefrontal cortex activity, predicting when this peak will occur. She then **optogenetically inactivates the neural population** at the forecasted time. Because inactivation prevented the mouse from initiating a foraging bout, her hypothesis was supported.

Use Case

Notes:

1. for a week-long experiment, the size of the behavioural and neural recordings exceeds 200 terabytes. Due to the large datasets sizes, we will **bring users to data**, instead of data to users. Data will be stored in the cloud, computation will run on the cloud and user computers will only display small data and analysis results.
2. we will **share data, hardware specifications and open-source software openly**.