Contents

Intention to submit document for the Work with US researchers BBSRC-NSF/BIO lead agency 2024 funding opportunity

Enabling Naturalistic, Long-Duration and Continual Neuroscience Experimentation with Advanced Machine Learning

October 28, 2024

1 Summary

Word limit: 2 A summary is not required for this section, please write 'N/A' in the textbox. Please still include a title for your project.

N/A

2 Core team

List the key members of your team and assign them roles from the following:

- project lead (PL)
- project co-lead (UK) (PcL)
- specialist
- professional enabling staff
- research and innovation associate
- technician
- researcher co-lead (RcL)

Only list one individual as project lead.
The core team section must only contain details of the UK applicants. The US applicant information should be listed in the 'US applicants' section.
Find out more about UKRI's core team roles in funding applications.

project lead (PL) Prof. Maneesh Sahani

project co-lead (UK) (PcL) Prof. Tiago Branco, Prof. Thomas Mrsic-Flogel

researcher co-lead (UK) (RcL) Dr. Joaquin Rapela, Dr. Dario Campagner

professional enabling staff Dr. Adam Tyson

3 Application questions

3.1 Research theme

Word limit: 5 Please state the research theme you are applying under. Choose one of the following research themes:

1. biological informatics
2. understanding host-microbe interactions
3. synthetic cells and cellular systems
4. synthetic microbial communities

biological informatics

3.2 Vision

Word limit: 500

What are you hoping to achieve with your proposed work?
What the assessors are looking for in your response
Your vision should clearly address:

- one of the opportunity research themes (biological informatics, understanding host-microbe interactions, synthetic cells and cellular systems or synthetic microbial communities)
- the remit of the BBSRC and the NSF/BIO division associated with your chosen research theme

References may be included within this section, but this will count towards your word count.
Images are not required for this section.

3.2.1 Context

Conventional systems neuroscience experiments are typically short in duration and often place significant constraints on subject behaviour to simplify data analysis. However, these restrictions may limit our ability to observe critical aspects of brain function and behaviour that only manifest in more naturalistic and extended conditions.

At the Sainsbury Wellcome Centre (SWC) and Gatsby Computational Neuroscience Unit (GCNU) we are pioneering Naturalistic, Long-Duration, and Continual (NaLoDuCo) foraging experiments in mice that span weeks to months. During these experiments, we collect high-resolution behavioural and neural recordings in naturalistic settings.

This novel experimental approach will enable researchers to explore neural mechanisms underlying naturalistic behaviour over extended periods for the first time, offering the possibility of uncovering insights across a wide range of phenomena, including long-term behavioural adaptation, neural plasticity, and learning. The data generated from NaLoDuCo experiments represent an entirely new resource in neuroscience, with the potential to drive breakthroughs and discoveries that are beyond the reach of traditional experiments.

While experiments in neuroscience that are naturalistic, long-duration, or continuous have been conducted in the past (e.g., [1]), to the best of our knowledge, we are the first to integrate all three of these features in a single experimental paradigm.

Our US collaborator, the Allen Institute for Neural Dynamics (AIND) is also investigating foraging, but using head-fixed mice. Key to their mission is distributing very large Neuroscience datasets, and providing functionality to process them on the cloud.

Since the project started in 2021, our UK business partner, NeuroGEARS Ltd. has been contracted by the SWC to lead the implementation of the NaLoDuCo experimental framework. It also provides services to the AIND.

The extremely large datasets–on the order of hundreds of terabytes– gathered from experiments spanning weeks to months pose significant challenges in data acquisition, visualisation, and analysis. Together, the GCNU, SWC, AIND and NeuroGEARS will address these challenges, co-develop this new type of experimentation, share expertise and build software infrastructure to help scientists around the world perform NaLoDuCo experiments.

3.2.2 Focus areas

The focus areas of the proposed project are:

Data Collection & Management Efficiently gathering and organising massive datasets over extended periods.

Data Sharing Providing global access to large-scale datasets.

Data Visualisation Developing efficient web-based tools to visualise very large behavioural and neural datasets.

Spike Sorting Assigning spikes to neurons reliably and tracking individual neurons across long-periods of time in real time.

Data Analysis Characterising behavioural and neural recordings.

Inference-Driven Experimentation Creating a new type of experimentation driven by real-time behavioural and neural inferences.

3.2.3 Cross fertilisation

The foraging experiments at the AIND are different from those at the SWC. They do not probe freely moving and naturalistic behaviour, but are able to perform electrophysiological recordings more densely than those at the SWC. These experimental approaches to foraging are complementary and this collaboration will greatly benefit both of them.

Currently, both GCNU and AIND are independently developing methods to address the previouus focus areas outlined. We will joint forces to co-develop these areas and our foraging research programs, leveraging our combined expertise for greater impact.

3.3 Approach

Word limit: 500

How are you going to deliver your proposed work?

What the assessors are looking for in your response

Your approach should give an overview highlighting:

- a clear description of the objectives and methodology for the proposed work, including the contributions of the UK and US teams
- the potential outputs and outcomes of the proposed work

References may be included within this section, but this will count towards your word count.

Images are not required for this section.

3.3.1 Data collection & management

We have developed an innovative platform for housing of mice in large arenas (>2m diameter) enabling precise behavioural manipulation and high-resolution monitoring (online figure, [2]). We have openly shared software for supporting data acquisition [3] and management [4] in this arena. Additionally, the platforms supports continuous, long term monitoring of neural activity with Neuropixels probes, capable of recording from thousands of neurons simultaneously spanning the entire brain depth. This setup has allowed us to collect several week-long datasets with single and multiple mice per arena.

To facilitate the replication of our experimental setup by other groups, we will share instructions for building foraging arenas, as well as specifications of hardware used in them, and we will improve the documentation of the software repositories for data acquisition and management.

3.3.2 Sharing data and methods

The very large datasets produced by NaLoDuCo experiments make traditional methods of data distribution impractical. Instead, users will interact with the data directly where it is stored. The maturation of cloud technologies now makes this possible.

We will leverage DANDI, which utilises Amazon S3 storage, for hosting the data. Additionally, we will provide software to visualise and analyse data using Amazon EC2 instances, thereby minimising the need for time-consuming data transfers.

Handling and sharing continuous behavioural and neural recordings of this scale presents unique challenges. Runtime performance is one of them. If we encounter unacceptable delays, we will explore advanced optimisation strategies, such as parallel processing and resource-efficient cloud configurations.

3.3.3 Data visualisation

Our visualisation tools need to display very large datasets at different temporal scales, from milliseconds to weeks and months, and they need to be web based. We will use multi-resolution visualisation techniques, which store data at various resolutions, and use the appropriate resolution for each zoom level. Web-based visualisation will be optimised using web workers.

3.3.4 Spike sorting

Spike sorting is specially challenging in NaLoDuCo experimentation since we want to track individual neurons of freely moving mice for weeks to months. In addition, we need online spike sorting, to allow experiments driven by real-time machine learning inference, as described below. We will evaluate methods for tracking neurons over long periods of time (e.g., [5]) and for online sorting (e.g., [6]). If needed, we will develop new methods, as we are experienced on the subject.

3.3.5 Data analysis

The very large size of NaLoDuCo experimental data, the fact that the statistics of these data change across time, and the requirement for real-time and close-loop inference create new challenges to conventional machine learning data analysis methods. We will evaluate how existing methods targeting the focus areas described above cope with these challenges and, if necessary, create new ones.

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 the 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]).

3.3.6 Inference-driven experimentation

We call inference-driven experimentation to a type of experimentation driven by machine learning inferences on neural or behavioural data, where the result of these inferences can change the experiment in real time.

We will apply inference-driven experimentation to test if patterns of neural activity are causally related to foraging behaviours. We would first check that a pattern of neural activity always precedes a given foraging behaviour. We would then detect the occurrence of the pattern and in real time optogenetically inactivate the neurons responsible for the pattern. If the behaviour disappears the causality argument would be supported.

For this we will use the Bonsai ecosystem for experimental control and online machine learning functionality that we are adding to Bonsai [15], funded by a BBSRC award.

3.4 US applicants

Word limit: 200

Please provide the following details of the US applicants on this application:

1. name
2. institute
3. job title
4. role in project (for example, project lead or project co-lead)
5. email address

Please also indicate who the lead US applicant will be.
NSF will use this information to confirm applicant eligibility.
Please do not include details of US applicants in the 'Core team' section.

1. Saskia de Vries

   institute Allen Institute for Neural Dynamics
   job title Associate Director, Data and Outreach
   role in the project project lead
   email saskiad@alleninstitute.org

2. David Feng

   institute Allen Institute for Neural Dynamics
   job title Sr. Director, Scientific Computing
   role in the project project co-lead
   email david.feng@alleninstitute.org

3.5 Resources

Word limit: 200

Please provide the following:

- overall estimates for costings and staffing full time equivalent (FTE) for both the UK and US components
- clear separation of UK and US costings, in pounds sterling and US dollars (USD) respectively

The overall budget should be below the maximum £2 million combined funder contribution

If there is more than one UK or US team associated with the application, please combine their estimates together.

A detailed calculation and breakdown of resources is not required at this stage, nor is a justification of costs.

The following is an example of how this might look.

UK Resources:

Total cost estimate: £600,000

Research council contribution: £480,000

0.2 FTE time, 1.0 FTE PDRA, 0.5 FTE technician

US Resources:

Total cost estimate: $300,000

1.0 FTE PDRA or 1.0 FTE doctoral researcher

Total funder contribution estimate:

£716,475 (£480,000 + £236,475 ($300,000 at exchange rate 0.79))

UK Resources:

Total cost estimate: £xxx,xxx

Research council contribution: £yyy,yyy (i.e., 0.8 * £xxx,xxx)

```
# 3 PI at 0.1 FTE
# 1 experimental postdoc at 0.25 FTE
# 2 RSE at 1.0 FTE
```

3 x 0.1 FTE PI, 1 x 0.25 FTE PDRA, 2 x 1.0 FTE RSE

US Resources:

Total cost estimate: $www,www

```
# 1 AIND scientist1 at 0.5 FTE
```

1 x 0.5 FTE scientist1

Total funder contribution estimate:

£716,475 (£yyy,yyy + £ppp,ppp ($www,www at exchange rate 0.79))

References

[1] B Voloh, DJN Maisson, RL Cervera, et al. *Cell reports*, 42(9), 2023.

[2] D Campagner, J Bhagat, G Lopes, et al. In *Society for Neuroscience Abstracts*, p. PSTR033.03 / I26, 2024. SfN online abstract.

[3] SFBW Group. https://github.com/SainsburyWellcomeCentre/aeon_acquisition, 2024.

[4] SFBW Group. https://github.com/SainsburyWellcomeCentre/aeon_mecha, 2024.

[5] AX Yuan, J Colonell, A Lebedeva, et al. *Elife*, 12:RP92495, 2024.

[6] U Rutishauser, EM Schuman, and AN Mamelak. *Journal of neuroscience methods*, 154(1-2):204–224, 2006.

[7] A Mathis, P Mamidanna, KM Cury, et al. *Nature neuroscience*, 21(9):1281–1289, 2018.

[8] J Rapela. https://github.com/joacorapela/lds_python, 2024.

[9] S Challa, MR Morelande, D Mušicki, and RJ Evans. *Fundamentals of Object Tracking*. Cambridge University Press, 2011.

[10] AB Wiltschko, MJ Johnson, G Iurilli, et al. *Neuron*, 88(6):1121–1135, 2015.

[11] H Zhu, B De La Crompe, G Kalweit, et al. *arXiv preprint arXiv:2311.13870*, 2023.

[12] JH Macke, L Buesing, JP Cunningham, et al. *Advances in neural information processing systems*, 24, 2011.

[13] S Escola, A Fontanini, D Katz, and L Paninski. *Neural computation*, 23(5):1071–1132, 2011.

[14] X Deng, DF Liu, K Kay, et al. *Neural computation*, 27(7):1438–1460, 2015.

[15] J Rapela, N Guilbeault, and G Lopes. https://bonsai-rx.org/machinelearning/, 2024.