Breakout Session 6:

Evaluation and Optimization of NWB Neurophysiology Software and Data in the Cloud

Dr. Cody Baker
Sr. Neurodata Scientist, Lawrence Berkeley National Laboratory
Ms. Urjoshi Sinha
Computer Engineer, Lawrence Berkeley National Laboratory

Evaluation and optimization of NWB neurophysiology software and data in the cloud

Cody Baker & Urjoshi Sinha

BERKELEY LAB

CATALYST NEURO

NEURODATA WITHOUT BORDERS

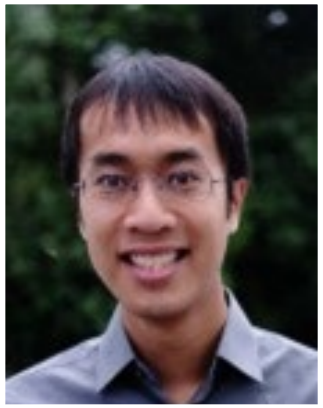
Multidisciplinary team science at work

Technology Team for the Supplement Award:

BERKELEY LAB

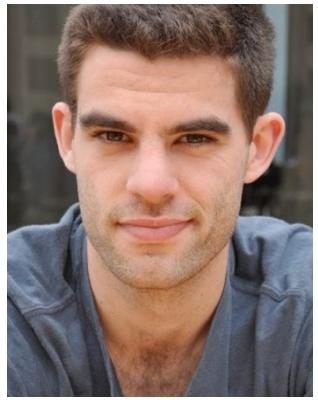
O. Rübel

U. Sinha



R. Ly

CATALYST NEURO



B. Dichter

C. Baker

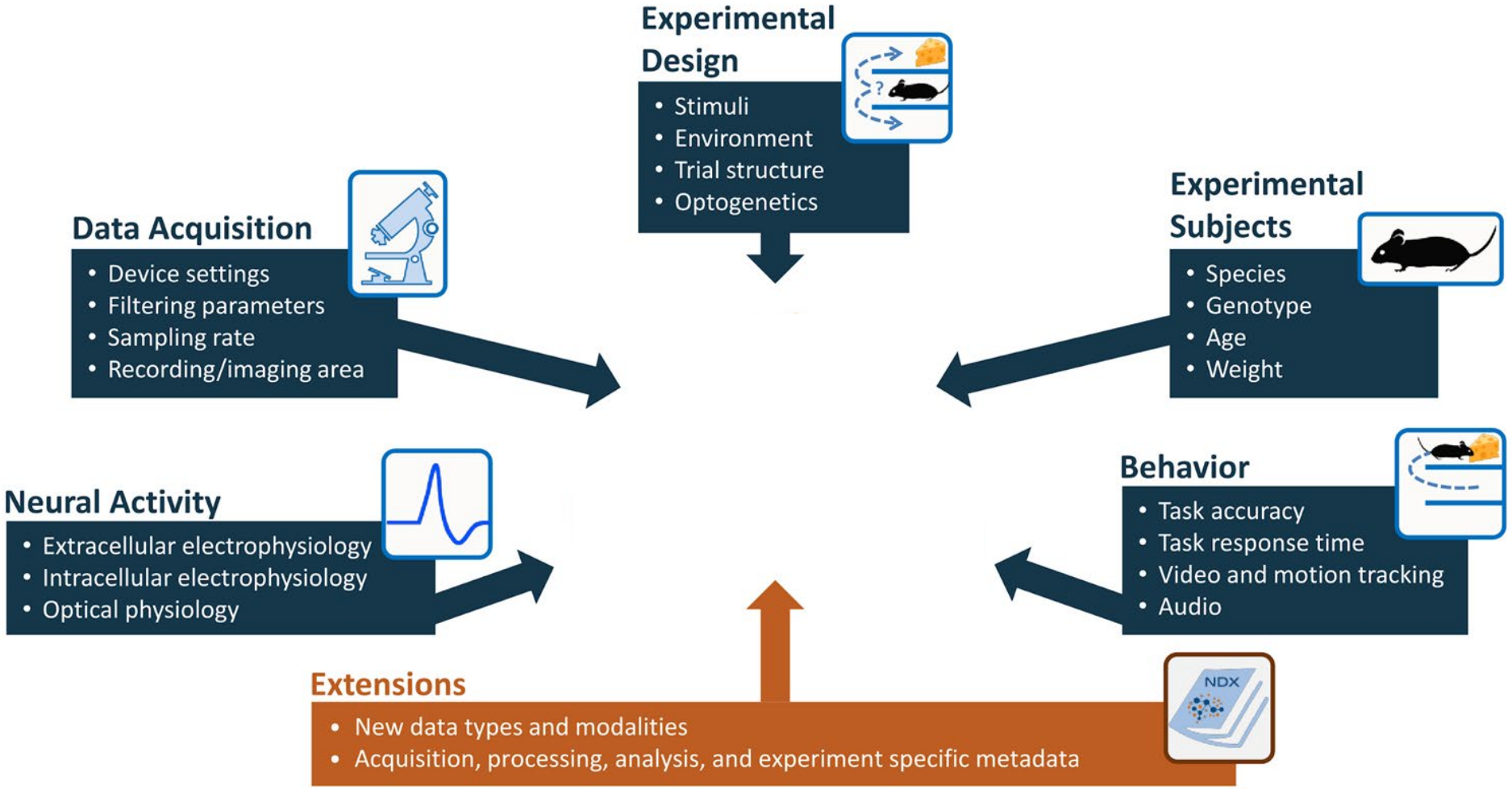
A. Weigl

Parent Award:

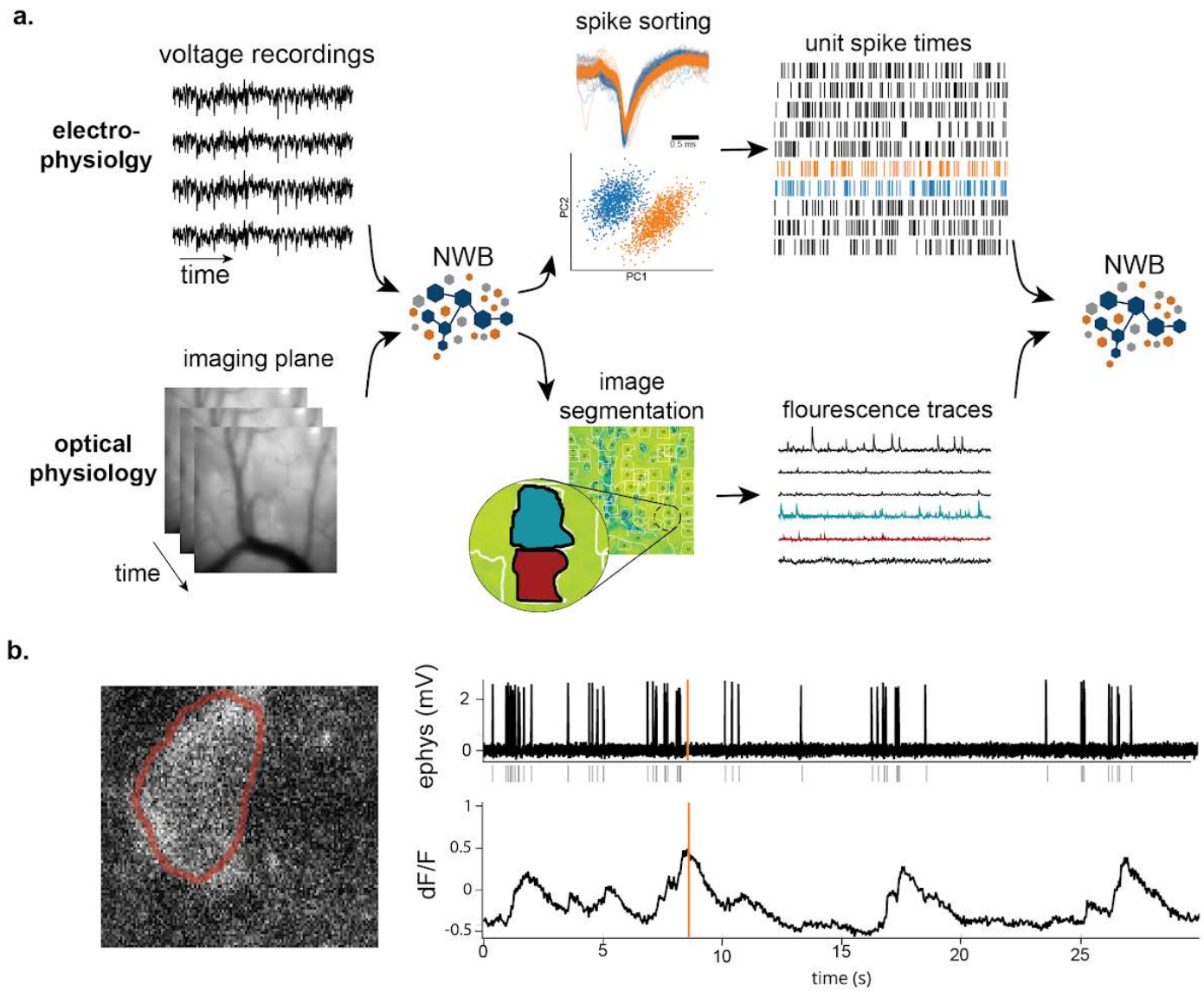
- **Title:** Advancing Standardization of Neurophysiology Data Through Dissemination of NWB
- **Award Number**: U24NS120057
- **Sponsor:** National Institutes of Health, National Institute of Neurological Disorders and Stroke

NWB – A unified data standard for neurophysiology

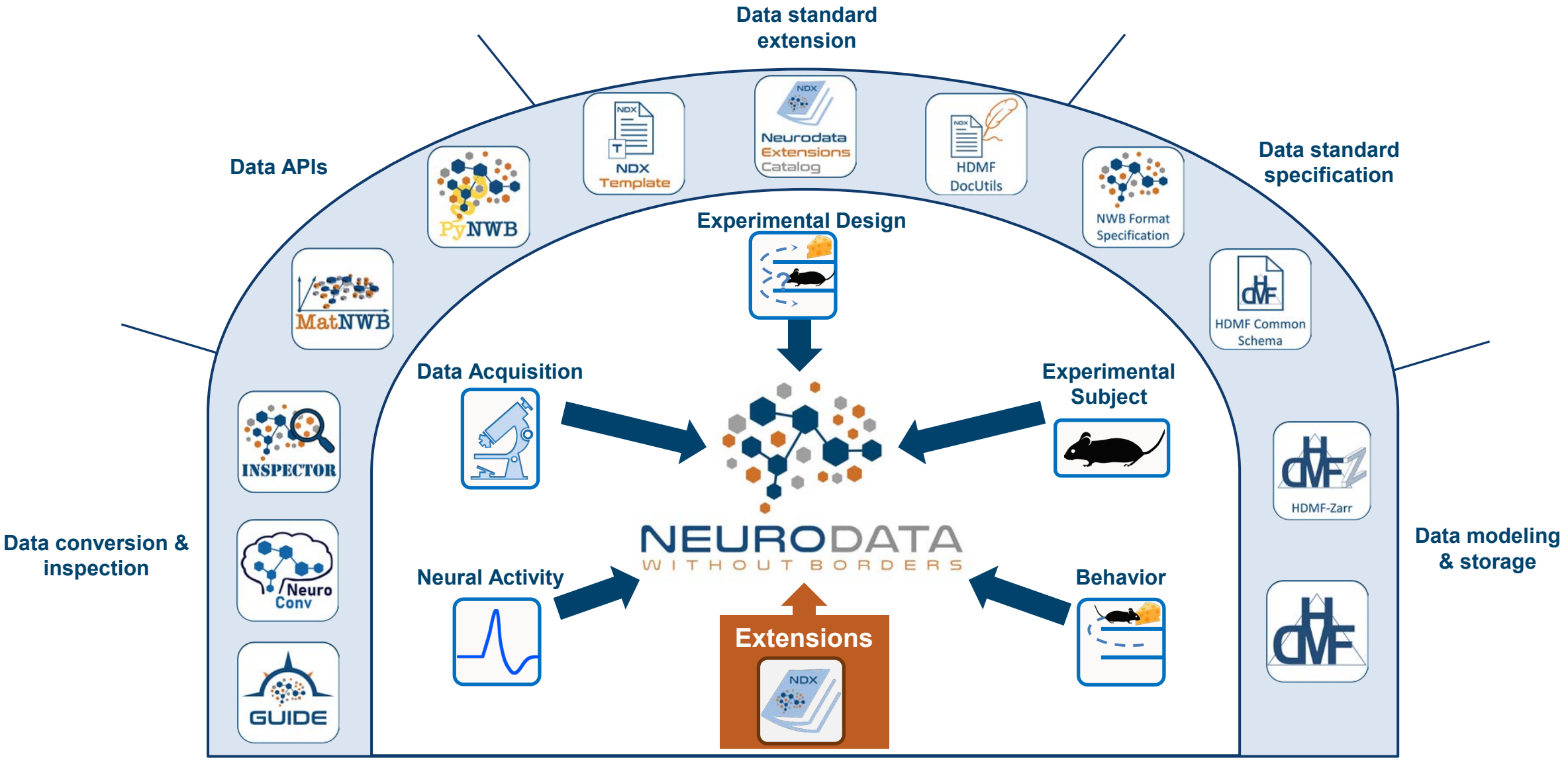
NWB defines a unified data standard for neurophysiology data, focused on the dynamics of groups of neurons measured under a large range of experimental conditi



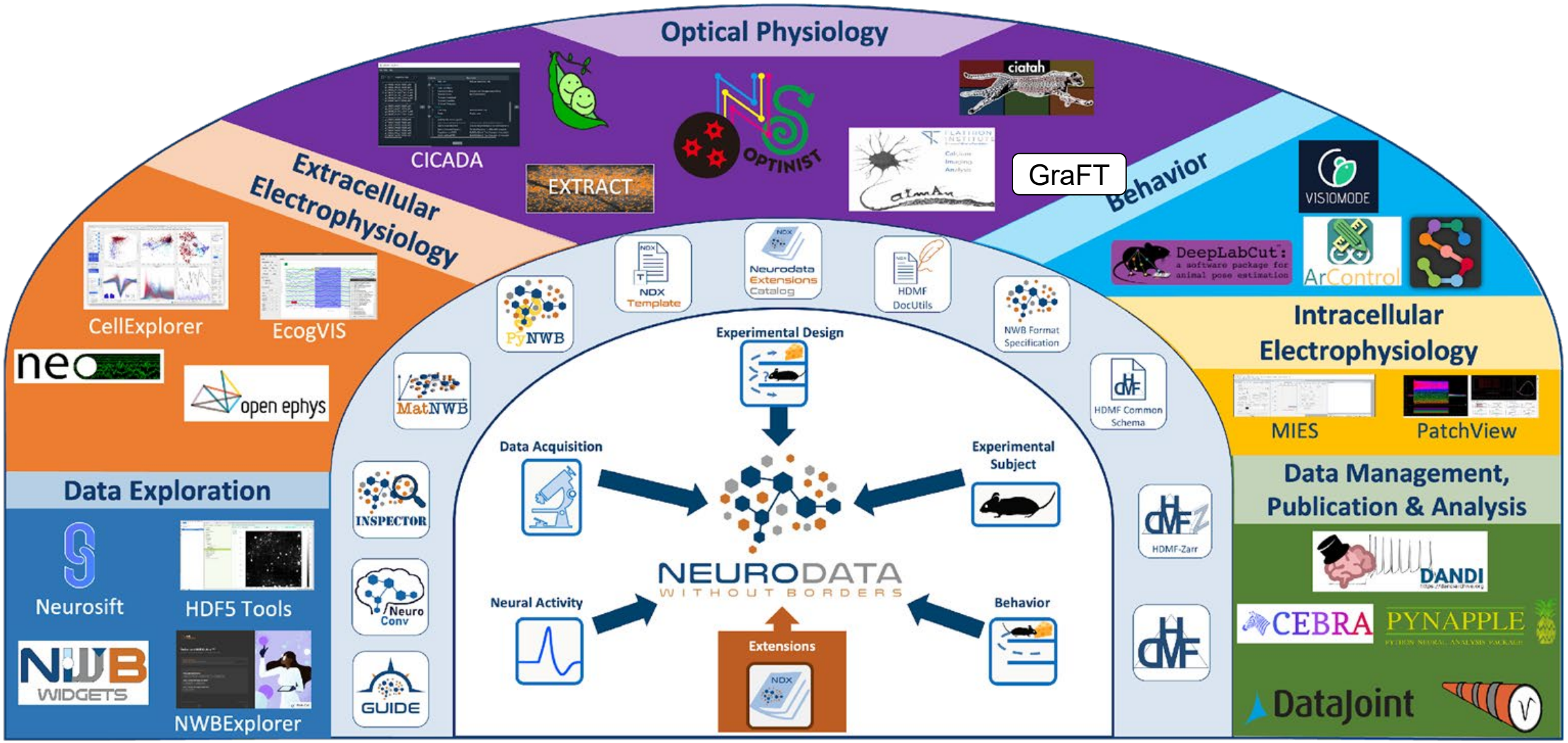
NWB enables unified storage of multimodal raw and processed data



NWB provides an integrated ecosystem of software tools for neuroscience data standardization



A growing ecosystem of 30+ community tools supports NWB



DANDI Archive

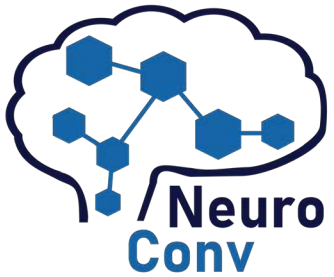
- The BRAIN Initiative archive for publishing and sharing neurophysiology data
  - Primary cloud storage of NWB files
  - S3 bucket, part of the Open Data program at AWS

- DANDI Hub enables **FREE** cloud-based data analysis (but fixed size instances)
- Dendro allows users to run specific analyses on custom compute resources

Challenges

Problems:

1) Growth in volume of data from experiments combined with inadequacy of current conversion tools for cloud integration…



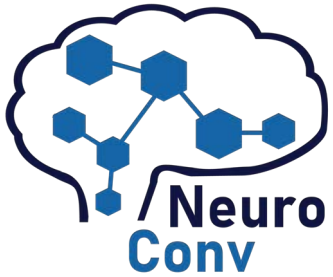
2) Inefficiency in the accessibility and storage of cloud assets…

Consequences:

…impedes use of modern cloud-based data processing and analysis applied to large-scale neurophysiology data.

… hinders web-based exploration of large neurophysiology datasets stored in the cloud

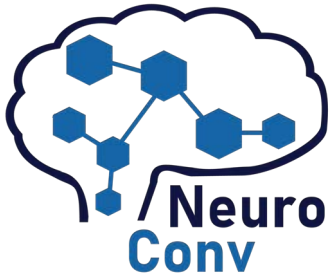
Aim 1: Evaluate and optimize containerization and cloud-resource-utilization strategies to enable efficient, cost-effective cloud-based conversion of neurophysiology data to NWB



NeuroConv is an open-source software package developed to simplify the conversion process for neurophysiology data to NWB

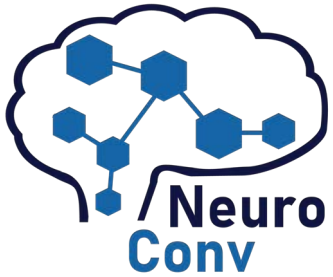
- Currently supports 40+ common neurophysiology data formats

Aim 1: Evaluate and optimize containerization and cloud-resource-utilization strategies to enable efficient, cost-effective cloud-based conversion of neurophysiology data to NWB



- **Objective 1.1: Containerize and optimize NeuroConv for cloud deployment.**
  - Most popular way to convert source data to NWB, but requires local file access
  - Many labs have TBs of data stored in the cloud already (Drive, Dropbox, Box, etc.)
  - Could horizontally scale the conversion process via AWS EC2 Batch
    - Requires Docker image of NeuroConv package
    - Offer price estimates before user gives OK to spin up instances
    - Over time, improve the EC2 configuration to make it as cheap as possible

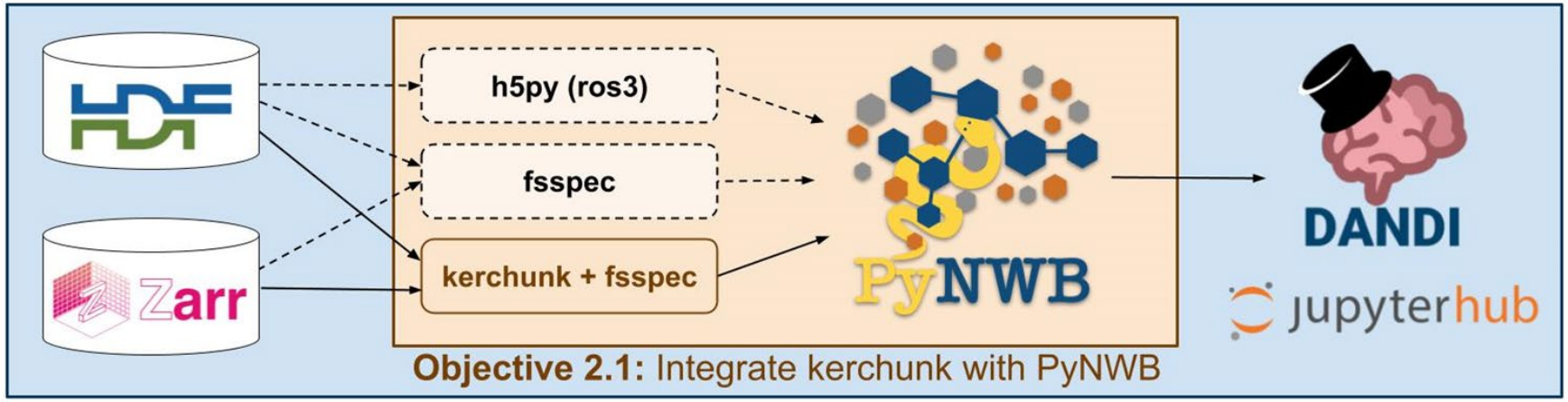
Aim 1: Evaluate and optimize containerization and cloud-resource-utilization strategies to enable efficient, cost-effective cloud-based conversion of neurophysiology data to NWB



Objective 1.1: Containerize and optimize NeuroConv for cloud deployment.

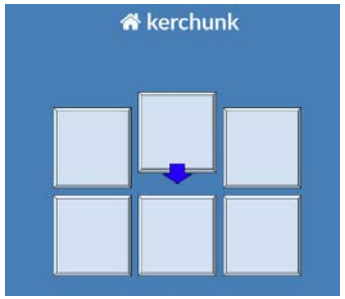
- **Objective 1.2: Integrate existing cloud resources for input and output of conversion processes.**
  - Running the NeuroConv container is independent of the data transfer however, so need an easy way for users to specify where and how to execute the data transfer onto EC2 instances
  - Integrate with data transfer tools such as Rclone to allow for a simple syntax

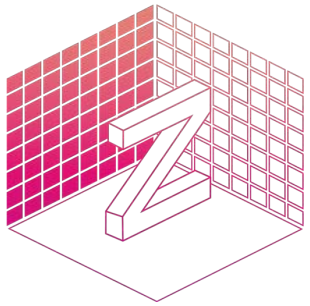
Aim 2: Evaluate and optimize reading of NWB neurophysiology data from cloud storage to enhance cloud-based analysis



- **Objective 2.1: Integrate Kerchunk into PyNWB to read data efficiently from the cloud.**
  - Overcome of challenges of existing methods in handling metadata and web-based data access
  - Accelerate data access by caching metadata in separate locations

Aim 2: Evaluate and optimize reading of NWB neurophysiology data from cloud storage to enhance cloud-based analysis

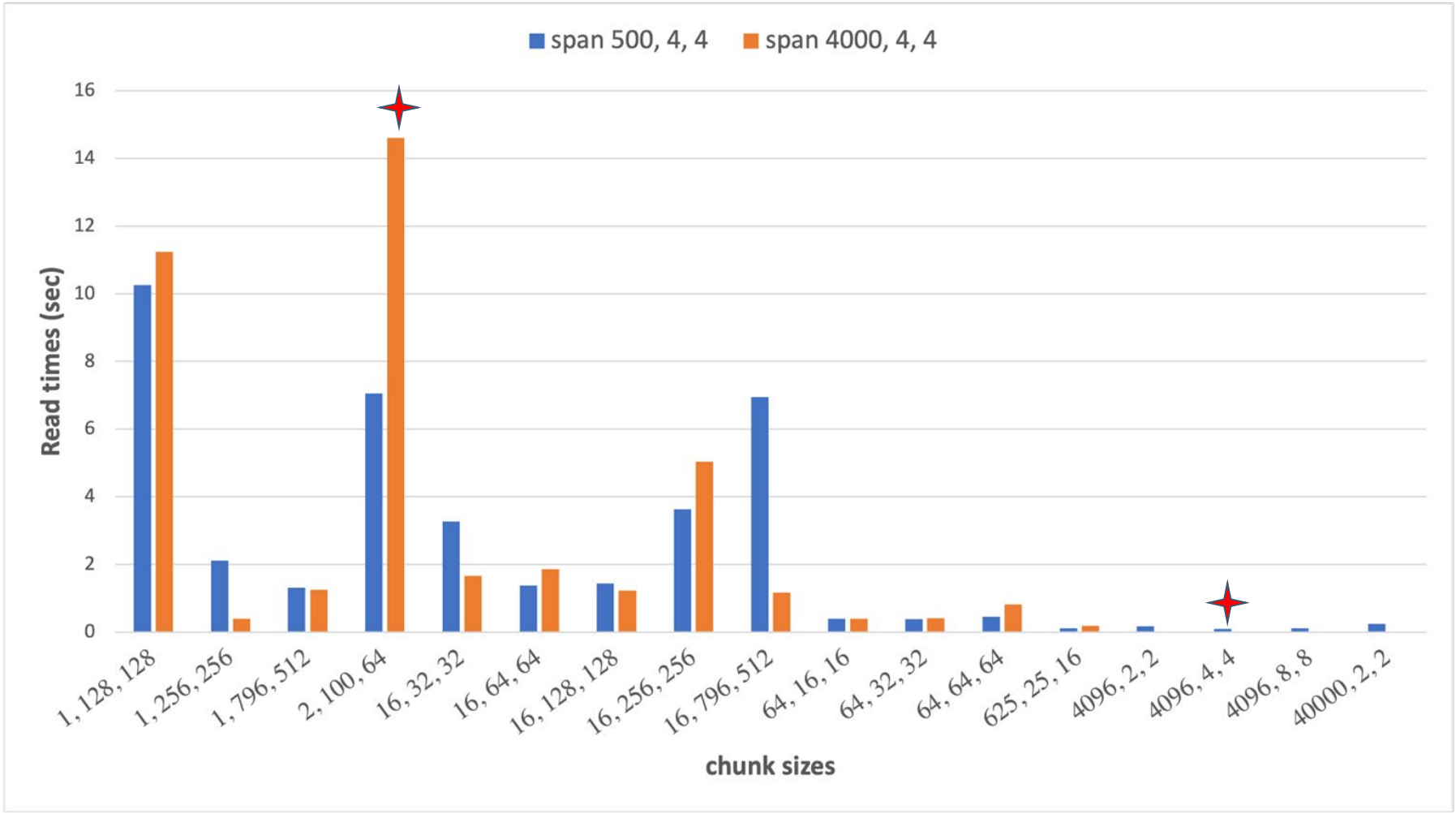




- **Objective 2.2: Evaluate the performance of data layout strategies for reading data from cloud storage.**
  - Determine storage configurations (compression algorithms, chunk shapes, etc.) based on use cases
  - Conduct performance analysis and provide guidance on best practices

Aim 2 – Example impact on optical physiology

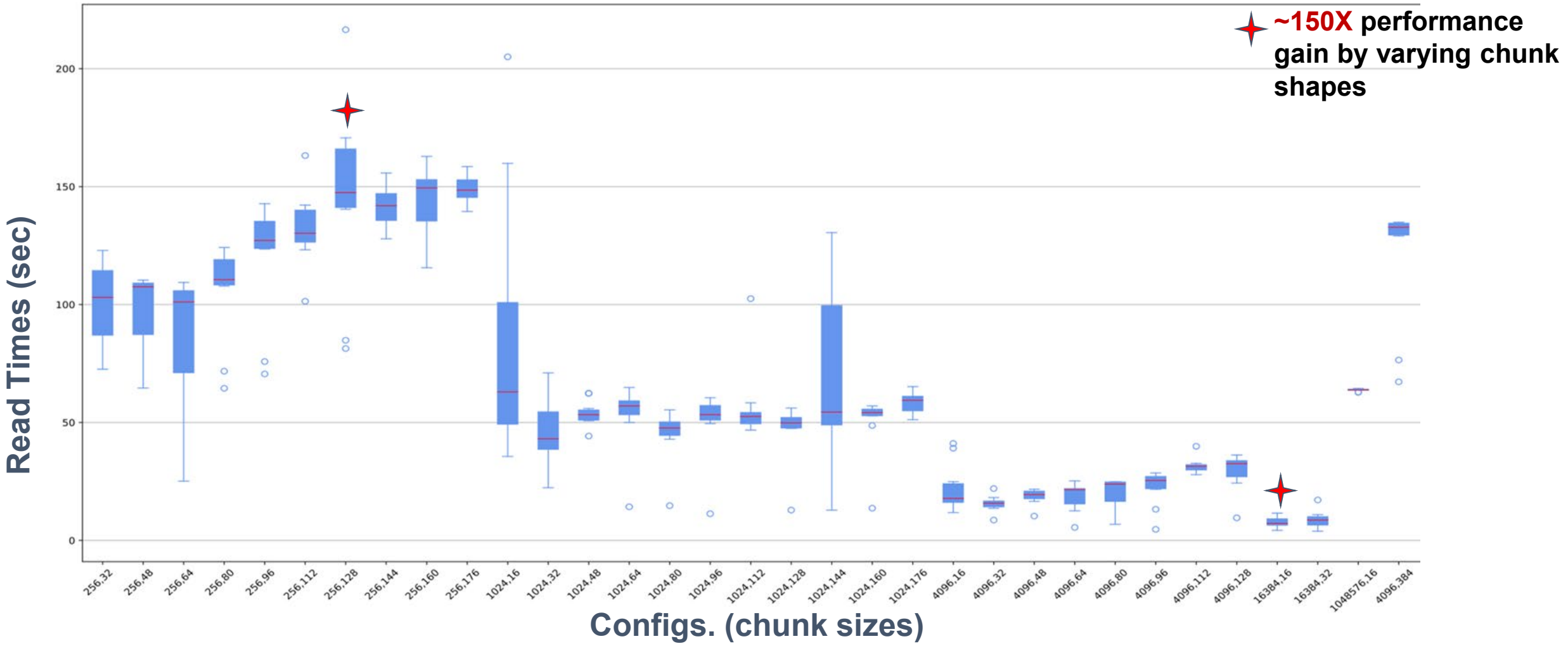
- **Objective 2.2:** Example ophys use case reading a small patch of an image over a block of time



~14X performance gain by varying chunk shapes

Aim 2 – Example impact on electrophysiology

- **Objective 2.2:** Example ecephys use case band pass filter raw data one electrode at a time



NEURODATA
WITHOUT BORDERS

Thanks for listening!

To learn more about NWB, visit us at
[NWB.org](NWB.org) and nwb-overview.readthedocs.io

Contact us at
github.com/NeurodataWithoutBorders/helpdesk