

Neurodatascience: Past, Present, and Future

Keiland W. Cooper^{a,b} , Babak Shahbaba^c , and Norbert J. Fortin^{a,b} 

^aDepartment of Neurobiology and Behavior, University of California, Irvine, Irvine, California, USA; ^bCenter for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, California, USA; ^cDepartment of Statistics, University of California, Irvine, California, USA

ABSTRACT

The study of the brain is a compelling example of the power of convergent science. Over the last few decades, advances in neuroscience techniques and experimentation, as well as in data science tools to analyze the resulting data, have dramatically furthered our understanding of fundamental brain functions. Historically, it has been common for analytical approaches to have a considerable lag in development following the availability of new neuroscience techniques. However, this relationship has not simply been unidirectional, as there have been examples in which analytical developments have directly led to new scientific questions and experiments. Here we review how this interplay between neuroscience and data science advances has unfolded in the past and into the present, with a focus on electrophysiology and calcium imaging. Applying lessons learned from the past and present, we then discuss expected developments, challenges, and opportunities in the future. We end by providing recommendations on how to foster the necessary team science approach to continue the advancement of research at the intersection of neuroscience and data science, which we call neurodatascience, toward a sustainable future.

ARTICLE HISTORY

Received 28 April 2025
Revised 22 December 2025
Accepted 3 January 2026

KEYWORDS

Brain activity;
dimensionality reduction;
machine learning; deep
learning; neural decoding;
signal processing;
computational neurosci-
ence; big Data

1. Introduction

“Computational biologists will be the ‘breakthrough’ artists of the future.”

—National Institutes of Health (NIH) Director Dr. Francis Collins, Charlie Rose Interview (2010)

The human brain consists of over 86 billion neurons, which communicate with each other at the level of milliseconds using an estimated 100 trillion synapses. Despite this daunting scale and complexity, our understanding of brain functions has dramatically increased in recent years. Technical developments are generally viewed as the driving force behind such progress since it is now possible to investigate brain activity across levels of analysis, from ion channels, to single neurons, to networks of neurons, to that of the whole brain. The importance of the corresponding analytical developments to handle these increasingly large and complex datasets, however, tends to be overlooked. In fact, for as long as neural data have been collected, methods to visualize and preprocess the data, extract patterns, quantify the results, and assess the reliability of findings have also been essential for new discoveries.

In the following sections, we begin with a case study illustrating the interplay between technical and analytical advances, we then review key technical and analytical advances of the past, followed by the data collection and analysis methods of the present. In those sections, we highlight how progress in neuroscience has been enabled by both technical and analytical advances and the feedback loop between them. We also

shed light on factors that limited the rate of progress, including the considerable lag between technical developments and the corresponding analytical developments needed to analyze the new data. In the final section, we provide an outlook on the future of neurodatascience, that is research at the intersection of neuroscience and data science, including expected developments, challenges, and opportunities. As we enter an era in which the magnitude of the datasets will allow us to investigate emergent properties of large neuronal ensembles, information dynamics between multiple brain regions, and perhaps bridge information processing across anatomical scales and species, these datasets will consistently bring new challenges for data science. By incorporating lessons learned from past neuroscience research and from other big data fields, we end by outlining recommendations on how to maintain the long-term alignment and synergy between neuroscience and data science needed to ensure that the rate of discovery keeps up with the rate of data generated. These recommendations focus on the increased importance of team science and collaborative platforms, unifying and scaling data infrastructure, leveraging the open source movement to ensure best practices for interpretable and shareable code, preserving the need for theory and proper experimental design, and interdisciplinary training programs.

It is worth noting that this article is intended to supplement prior reviews that have offered in-depth resources on more targeted topics, including computational neuroscience and

CONTACT Norbert J. Fortin  norbert.fortin@uci.edu  Department of Neurobiology and Behavior, University of California, Irvine, CA, USA

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1.1. The Interplay between Experimental and Analytical Advances: A Case Study of the Electroencephalogram and the Fourier Transform

Let us consider an example from the early days of neuroscience, specifically the development of the electroencephalogram

(EEG), to illustrate how the interplay between experimental and analytical advances can accelerate scientific progress. A key milestone toward modern neurophysiological recordings, Hans Berger’s invention of EEG in the 1920s provided a non-invasive method for recording the electrical activity of the human brain (Berger 1929). Berger placed electrodes on the scalp which recorded the millivolt-scale signals that are detectable outside of the skull (Millet 2002; Tudor et al. 2005).

The first few reports were largely qualitative observations, but generated considerable discussion on how best to quantify the patterns of brain activity. As Berger's EEG techniques reached wider adoption, many researchers qualitatively observed that the recorded voltages were periodic and appeared to oscillate. To analyze these data, the Fourier transform was identified as one of the first approaches to implement. The transform relies on the fact that a complex wave can be constructed through the summation of multiple, simpler constituent waves. Effectively, the transform

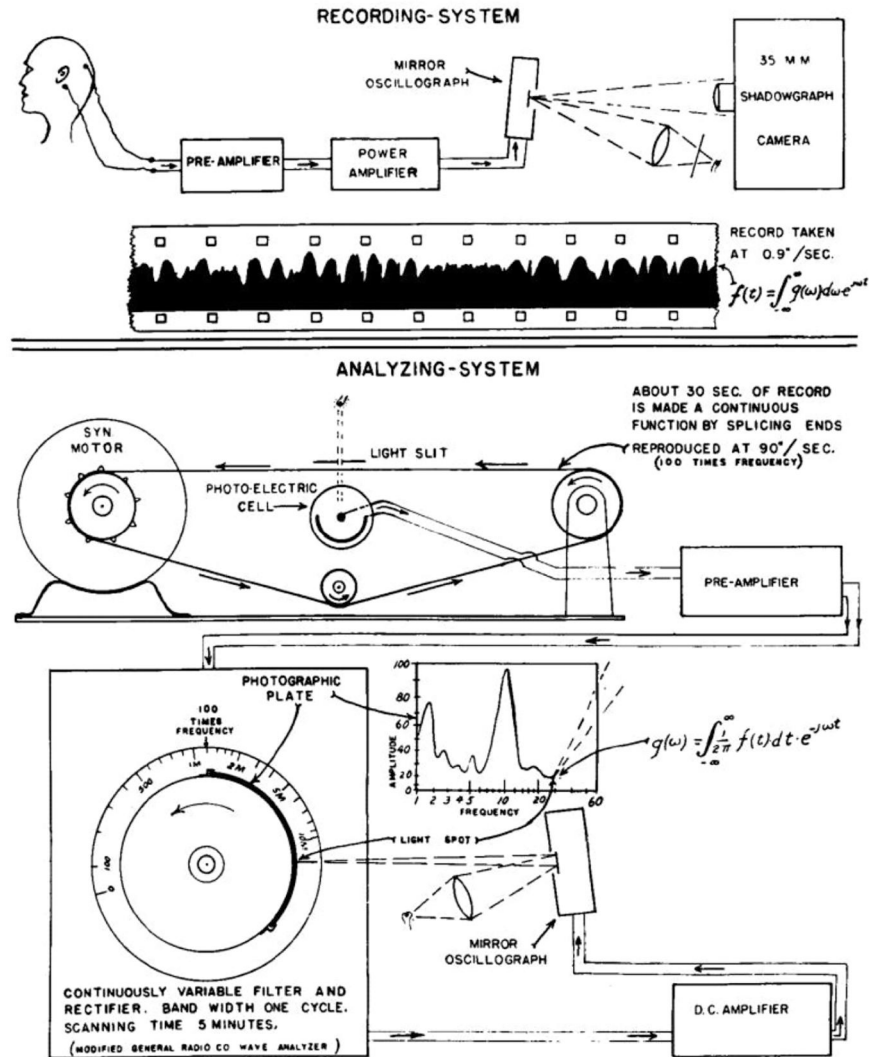


Figure 1. An early “automated” approach to calculate the Fourier transform on brain activity data. After the development of the electroencephalogram (EEG) in the 1930s, scientists could record brain waves from outside the skull. For years, they analyzed the resulting EEG data by calculating the Fourier transform by hand, which severely limited the rate at which experiments could be performed. Grass and Gibbs (1938) developed the mechanical device shown here which allowed the data to be analyzed much more rapidly, effectively lowering the barrier for analyzing EEG data. In their approach, an electroencephalogram (EEG) trace was first transcribed onto a long sheet of shadowgraph film which was then manually spliced into a loop. The film would then pass through their analyzing system. By rolling the film at different speeds, a photoelectric cell in the recording system could detect the EEG trace and measure the amplitude, or power, of each constituent frequency. Such automated devices, in turn, enabled new scientific questions to be generated and tested, which subsequently led to major discoveries in neuroscience. Figure from Grass and Gibbs (1938).

splits an EEG trace into its constituent frequencies (more technically, it decomposes the signal into frequency components), which results in a map of each frequency's influence on the brain activity. For two decades following EEG's discovery, the Fourier transform and other early analysis techniques were painstakingly calculated by hand, which severely limited the rate of progress.

Overcoming this analytical bottleneck required the development of an automated approach. However, in this era predating computers, the automation was required to be *mechanical* in nature instead of digital. Following development of a three-channel EEG device, Grass and Gibbs (1938) invented perhaps the first of such automated devices to perform Fourier analysis (Figure 1). By lowering the barrier for analyzing EEG data, this tool led to major discoveries in neuroscience, including the identification of distinct frequency bands and sleep stages, and set the stage for the study of brain activity disorders like epilepsy. Notably, almost a century after Grass and Gibbs' "simple" method, the Fourier transform is now a standard approach in an electrophysiologist's toolbox, though it can now be performed in seconds and, in some programming languages, by using a single command (e.g., 'Y = FFT(X)').

2. The Past

"... small research groups predominate, where each experimenter is [their] own theorist and the complexity of an investigation is limited, among other things, by the analyzing power embodied in the experimenter and [their] equipment"

—J. Krüger, *Simultaneous individual recordings from many cerebral neurons: Techniques and results* (1983)

Early electrophysiology experiments involved little statistical analysis and instead focused on visualizing snapshots of EEG traces or patterns of spiking activity from individual neurons. The importance of statistical and computational methods—an early version of what is known as data science today—increased considerably toward the end of the twentieth century. This increase was due to a number of factors, including increased yields of simultaneously recorded neurons in electrophysiological data, new neuroimaging techniques, and the increasing availability of high-performance computing resources. In this section, we review key developments in neuroscience and data science of the past that have catalyzed the transition to the approaches in use today. For more detailed accounts of the historical development of analytical approaches in neuroscience, we direct the reader to other reviews (Moore et al. 1966; Krüger 1983; Verkhratsky and Parpura 2014).

2.1. Recording Techniques: Early Developments

2.1.1. Recording Activity from Single Neurons

While Berger's EEG technique allowed for the recording of neural voltages outside the skull in the 1920s, methods to directly record the activity of individual neurons *inside* the brain were only developed in the 1960s. Such intracranial recordings initially involved manufacturing glass micropipettes by hand and slowly lowering these fragile electrodes

into the brain of anesthetized animals until the extracellular voltage traces of spiking neurons could be recorded (see Mountcastle et al. 1957). A subsequent breakthrough was the introduction of metal electrodes, which were robust enough to chronically record from unanesthetized and unrestrained animals. This allowed the systematic investigation of the relationship between patterns of brain activity and the processing of external stimuli, such as the classic studies of Hubel and Wiesel linking the activity of specific neurons in the visual cortex with specific patterns of light in the visual field (Hubel 1959). During the second half of the twentieth century, such intracranial recordings in increasingly richer experimental paradigms have significantly advanced our understanding of the relationship between neural activity and behavior.

2.1.2. Simultaneously Recording from Small Neuronal Ensembles

Early recordings focused on the activity of a single neuron per electrode, typically the neuron closest to the electrode as it would have the largest spike amplitude. While electrodes could also record smaller spikes from more distant cells (a few more microns away), the inability to reliably differentiate those smaller spikes and assign them to their corresponding neuron prevented any analysis focused on the activity of simultaneously recorded ensembles (groups) of neurons. Early analytical attempts to resolve this problem, using cluster analysis or template matching (Gerstein and Clark 1964; Glaser 1971; Abeles and Goldstein 1977; Pinsker and Eberly 1982), could infer variations in spike profiles but remained limited on their own. This analytical barrier led to technical development aimed at improving the amount of information given as input to the sorting algorithms. Specifically, given that the amplitude of a neuron's spike declines with distance from the electrode tip, the inclusion of multiple electrode tips in close proximity allowed for the triangulation of each neuron's spikes across those electrode tips (McNaughton et al. 1983; O'Keefe and Recce 1993; Wilson and McNaughton 1993). This ability to assign individual spikes to the corresponding neurons, known today as spike sorting, signaled the beginning of a race to maximize yield (the number of simultaneously recorded neurons).

2.2. Main Analytical Challenges and Solutions

2.2.1. Decomposing Signals into Individual Frequencies

The arrival of integrated electrical circuits, vacuum tubes, and transistors in the 1940s gradually supplanted the use of mechanical data analysis machines to perform computations such as the Fourier transform. As computers became available, the calculation of simple descriptive statistics of the EEG spectrum could be performed by transcribing data onto punch cards and producing printouts of the results (Grass and Gibbs 1938). As computer technology and availability increased over the subsequent two decades, so did the types of analysis that could be performed (Griffith et al. 1965). In fact, many analytical methods for EEG data analysis commonly used today were

developed during this period, such as cross-correlations and coherence analyses (Dumermuth and Flühlér 1967). Perhaps the most influential analytical advance of the time was the introduction of the *fast* Fourier transform. The algorithm overcame the computational limits of Fourier analysis, which at times could be computationally intractable, and allowed it to be performed more broadly and efficiently (Cooley and Tukey 1965). Notably, many of these methods would later be extended to spike data but computational constraints limited their widespread use until the early 1980s.

2.2.2. Quantifying Patterns in Individual Channels

Early methods to visualize the relationship between single-channel activity and experimental stimuli or responses were developed in the 1960s (see Kass et al. 2014 for a review; Figure 2a-c). One example is the peri-stimulus time histogram (PSTH; Gerstein and Perkel 1969) which is still commonly used today. The idea is quite simple: align the raster plots showing the spikes observed on each trial by time-locking them to a key task event (e.g., a stimulus or a response), then produce a histogram by averaging spikes across corresponding time bins and across trials. Another example is the use of autocorrelations to explore the temporal structure of activity of individual neurons (Gerstein and Kiang 1960). By calculating how likely a spike is to occur in each time window following a previous spike in the same neuron, this approach can identify potential rhythms governing spiking activity within a single neuron. While it is easy to take these methods for granted today because of their simplicity and omnipresence, they played (and still play) an essential role capturing and visualizing complex activity patterns that would otherwise be difficult to describe.

2.2.3. Capturing Relationships among Channels and Variables

The 1970s and 1980s marked a period when electrophysiological studies moved beyond the analysis of single cells using single electrodes, to that of small groups of neurons (typically < 10 neurons at the time). The transition from single to multiple spike train analysis led to a paradigm shift in the application of neural data analysis techniques (Brown et al. 2004). The dynamic and stochastic nature of largely sparse spike trains meant even single spike trains were challenging to model on their own. Neural activity could be modeled as a point process, a stochastic sequence of event times, with each event corresponding to a spike. This methodology contrasted with the standard approaches used previously, which were designed to handle continuous-valued data. Further, increasing neural yields experimentally and analytically meant increasing the number of dimensions for a given analysis, which further stressed computing resources. Extending the preceding methods to include more than three spike trains proved difficult due to the exponentially increasing number of dimensions and time bins required to compute them. Likelihood-based models offered a solution to this problem (Borisjuk et al. 1985; Brillinger 1988; Okatan et al. 2005; Truccolo et al. 2005) because, rather

than predicting exact spiking patterns (as would a deterministic approach), they estimate the probability of observed data arising from a given model. This probability (i.e., the assumed joint probability density) describes how likely the observed data are for different values of the model's free parameters. These parameters can be estimated by maximizing the corresponding likelihood function within a frequentist framework or by obtaining their posterior distribution using Bayesian inference.

This period also saw the advancement of many of the most common neuroscience analyses and visualizations used at present to capture relationships among channels or variables Kass et al. 2014; see examples in Figure 2d-f). For example, cross-correlograms were first popularized in that period (Griffith and Horn 1963; Perkel et al. 1967). The cross-correlogram is a time-domain measurement of the association between a set of neurons, where given a pair of neural spike trains and a parameterized bin width, the correlation between the two binned spike trains is calculated at a series of lags over time. Another example is the Joint PSTH (JPSTH) which, if the dependence of two neurons is assumed (Aertsen et al. 1989), allows for the visualization of the joint activity of two neurons via a two-dimensional histogram (displaying the joint spike count per unit time at each time u and v for neuron 1 and 2, respectively). As above, these methods may be taken for granted today but they play(ed) an important role capturing and visualizing complex relationships.

2.3. Scaling up Visualization, Computational, and Analytical Tools for Big Data

One may argue that the analysis of brain activity began with data visualization, specifically with Berger's raw EEG traces and their qualitative relationship with behavior in the 1930s. Interestingly, around the same time, Santiago Ramon Y Cajal's visualization of stained neurons through a microscope, which he elegantly captured with hand drawings, sparked a revolution of hypothesis testing at the cellular level. For decades, research was thus bottlenecked by the rate at which a picture could be taken or the hand could draw. That is, until the 1980s, when many of the building blocks for modern systems neuroscience grew into standard practice. Thereafter, the use of computers in neuroscience research accelerated, with the developments of new computer hardware and software, such as the IBM 360 mainframe computer or the FORTRAN programming language. These user-friendly and increasingly available computational advances facilitated data storage and subsequent offline data processing through computer code rather than by hand.

These computing advances allowed existing visualization and analytical methods to be scaled up in response to the growth of datasets (Figure 2g). Advances in quantifying animal behavior also emerged, largely driven by video recording technology. This drastically improved the precision at which animal behaviors could be tracked, quantified, and ultimately related back to brain activity (McNaughton et al. 1983; see example in Figure 2h), in addition to allowing a

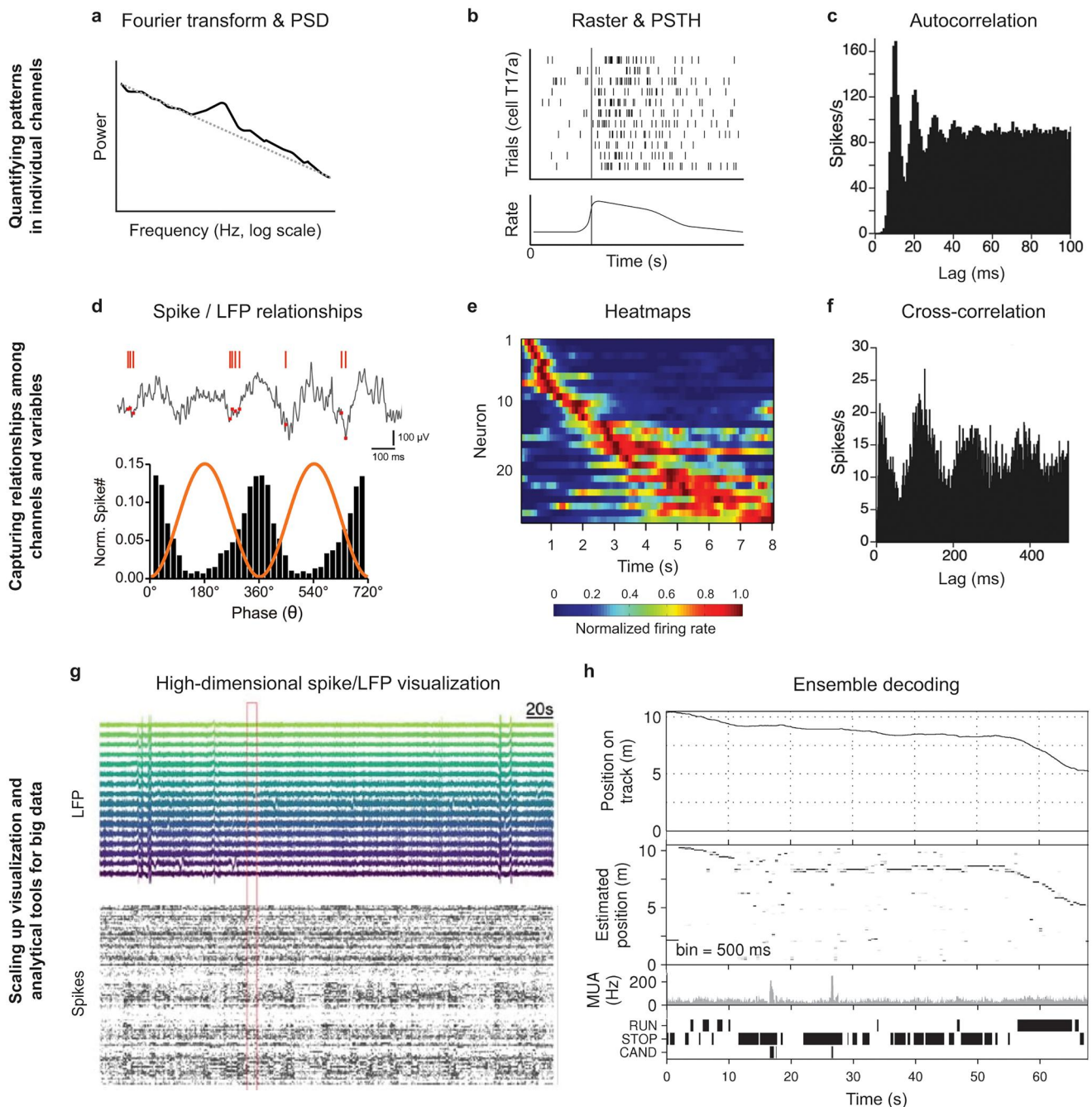


Figure 2. Common methods to visualize and analyze neural activity data across scales and levels of analysis. (a-c) Example methods to visualize and quantify activity patterns in individual channels. The plot on the left shows a Fourier transform and power spectral density for local field potential (LFP) activity (a). The middle plot shows spikes over a number of trials aligned to a stimulus, with the corresponding peri-stimulus histogram (PSTH) averaging the activity across trials below (b). The plot on the right shows an autocorrelation function reflecting the temporal structure in the firing of a given neuron (c). (d-f) Methods to capture relationships among multiple channels and variables. The plot on the left shows the timing of spikes aligned to an LFP trace to capture spike-phase relationships (d). The plot in the middle shows a heatmap converting firing rate to colors to facilitate the visualization of complex patterns across neurons (e). The plot on the right shows a cross-correlation function reflecting the relationship in spike timing between two neurons (f). (g-h) Example methods scaling up the visualization and analysis to larger datasets. The plot on the left shows spiking and LFP activity across a large number of channels (g). The plot on the right shows decoding (prediction) of the animal's position in a maze based on the ensemble activity alone (h). Figures adapted from Li et al. 2013 (panel c, f), Ahn et al. 2019 (panel d), MacDonald et al. 2011 (panel e), Guardamagna et al. 2022 (panel g), and Davidson et al. 2009 (panel h).

broader repertoire of stimuli and behaviors to be examined. Multi-modal analyses also considerably expanded with computer advances, specifically those examining relationships between discrete spike times and a continuous voltage signal (EEG or LFP; Figure 2d). While such early multi-modal analyses focused on qualitatively aligning the spiking of individual neurons to specific features of the EEG/LFP signal

(e.g., peak firing of cortical neurons to sleep spindles in the EEG trace; Creutzfeldt and Jung 1961), the ability to more rigorously visualize and quantify such relationships uncovered new types of information processing that could not have been discovered through the examination of the spiking or LFP activity alone. An example is the phenomenon of theta phase precession in the hippocampus, whereby

knowing the specific phase of theta oscillations at which neurons fire provides additional information (i.e., adds information about past and future states to that of the present state; O'Keefe and Recce 1993; Skaggs et al. 1996). Collectively, these analytical and computational advances, in turn, led to new types of scientific questions as well as new experiments to test them.

2.4. Lessons to Remember

Progress in neuroscience has been driven by developments in both experimental and analytical methodology. Those developments, however, did not proceed in parallel. In fact, there was often a considerable lag between technical breakthroughs producing new data and the development of the proper analytical tools to adequately analyze and interpret them. Importantly, while advances often occurred in that order (from technical to analytical), the process was more cyclical in nature in that analytical developments could also lead to new scientific questions and experimentation. Finally, analytical methods which at the time were considered significant advances are now routine steps in extended preprocessing pipelines.

3. The Present

“... the real complexity of brain function does not reside in the neurons themselves, but in the great manifold of their interconnections and in the spatiotemporal activity patterns to which these connections give rise”

—J. Krüger, *Simultaneous individual recordings from many cerebral neurons: Techniques and results* (1983)

In recent years, there have been considerable advances across neural recording methods (Figure 3a), ranging from electrophysiological (e.g., EEG, extracellular, intracellular), to optical (e.g., calcium imaging), to magnetic approaches (e.g., fMRI, MEG). In this section, we primarily focus on extracellular electrophysiology (Figure 3b,c) and optical imaging approaches (Figure 3d) as their high temporal resolution and recent dramatic scaling makes the successes of these approaches most dependent on the corresponding advances in data science. We refer the reader to other reviews for a more detailed treatment of advances in intracellular recordings (Lee and Brecht 2018), EEG (Buzsáki et al. 2012), or fMRI (Heeger and Ress 2002; Glover 2011).

3.1. Recording Techniques: Modern Methods

3.1.1. High-Density Single-Cell Electrophysiology (e.g., Tetrode Hyperdrives, Silicon Probes)

The fixed electrode arrays of the past have evolved into today's hyperdrives, which allow the depth of individual tetrodes (groups of 4 electrodes) to be adjusted manually to improve the targeting of specific layers and neuronal yield (Figure 3b). While the yield of tetrode hyperdrives can be high, especially in densely packed regions like the CA1 region of the hippocampus, building such drives requires considerable training and time. The size and weight of the drives can also limit behavior. For those reasons, silicon probes are gradually replacing hyperdrives as the

standard approach for single-cell recordings. Silicon probes are neural interfaces that enable high-density, multisite extracellular electrophysiology with exceptional spatiotemporal precision (Figure 3b). Advancements in photolithography added increased precision and geometrical control to the design and manufacture of printed circuits (Wise et al. 1970), laying the foundation for the present era of silicon-based recording probes. Such probes can contain up to hundreds of recording sites on a single shank, facilitating the collection of data from very large neuronal populations (Csicsvari et al. 2003; Buzsáki et al. 2004; Kelly et al. 2007; Urai et al. 2022; Figure 2g). These advances also improve signal fidelity and their miniaturization facilitates chronic implantation.

The main advantage of these electrophysiological techniques is that their high temporal resolution allows for the investigation of the fastest neural computations. By directly recording the extracellular voltage change produced by spiking, they contrast with indirect measures of neural activity, such as calcium influx or changes in blood oxygenation, that are noisier and of lower temporal resolution. Disadvantages of silicon probes include their fragility and inherent rigidity, with the latter thought to lead to gradual tissue and cell damage resulting in decreasing yields over time for chronic recordings. While these probes exhibit a high cost per unit, where each device can cost up to a few thousand dollars, they offer significant time savings owing to their commercialization, manufacturing automation, and batch production that can improve the hand-built workflow of traditional wire electrodes. Handling the resulting increase in yield can also prove challenging. In fact, analyzing and interpreting the large volumes of data from hundreds of simultaneously recorded neurons presents a significant computational burden, placing new constraints on data collection systems, storage solutions, and processing pipelines and analyses.

3.1.2. Optical imaging (e.g., Two-Photon Microscopy)

Optical imaging approaches record single-cell activity *in vivo* via functional microscopy (for review see Yang and Yuste 2017; Figure 3d). Optical imaging relies on genetically-modified indicators that transduce light in response to intracellular signals (typically calcium influx) which represent an indirect and temporally filtered proxy for spiking activity. More specifically, when neurons depolarize during spiking, voltage-gated calcium channels open which transiently increases the neuron's intracellular calcium levels. These intracellular calcium levels can be detected in real-time using fluorescent calcium-sensitive proteins (Chen et al. 2013), which can either be virally transfected (through surgical infusions of non-replicating viruses) or bred into neurons of transgenic animals (Luo et al. 2018). Optical techniques offer several advantages. First, these approaches have led to a massive increase in the yield of recordings, such that neural activity across populations ranging from hundreds to thousands of cells can now be simultaneously captured (Ohki 2005; Smetters et al. 1999; Stosiek 2003). Even larger yields can be obtained in non-mammalian species, where whole-brain recordings have been achieved in invertebrates (Nguyen et al. 2016) and up to 52,000 neurons in zebrafish

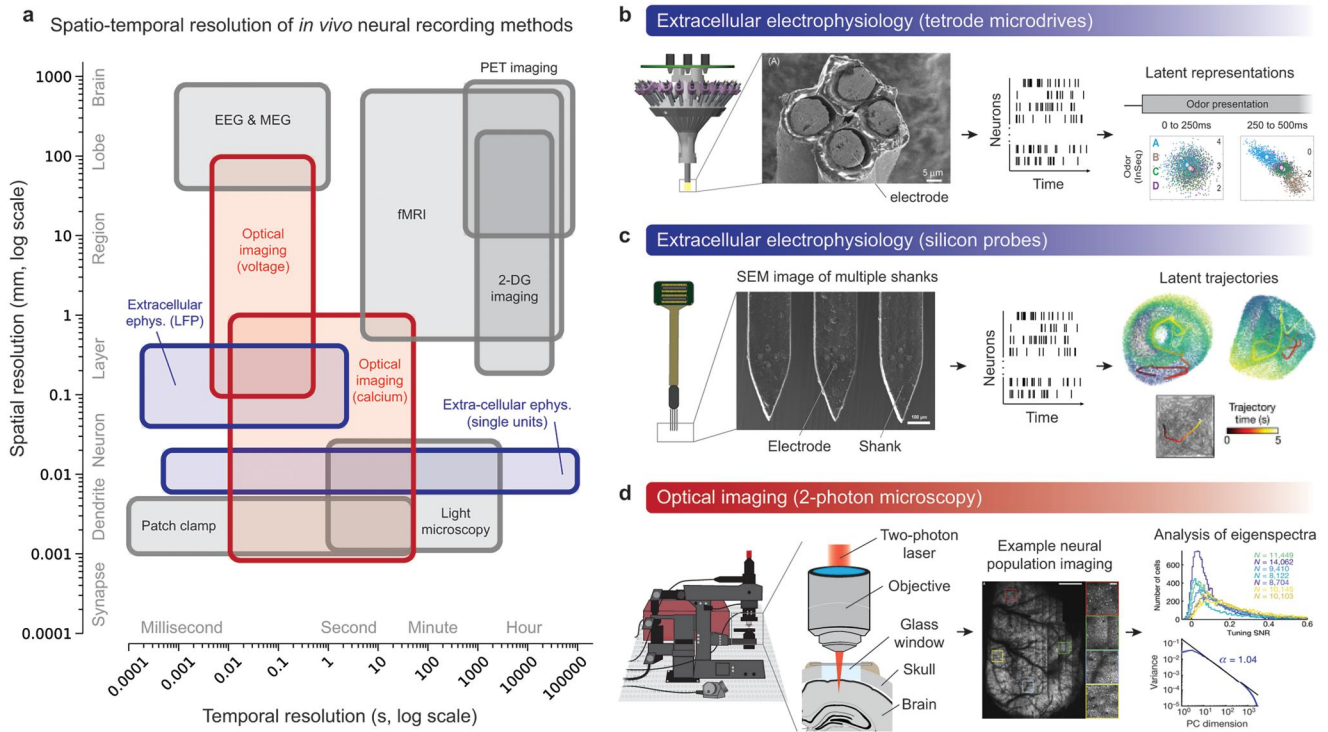


Figure 3. Modern methods for recording and analyzing activity from simultaneously recorded neuronal ensembles. (a) Each box represents the spatial and temporal resolution (the smallest physical or temporal scale at which a technique can reliably distinguish two separate units or time points, respectively) of methods commonly used to record neural activity. Lower values indicate higher resolution. Red and blue boxes indicate recording methods highlighted within the scope of this review. (b) Illustration of a 32-tetrode microdrive, with zoomed-in view of a single tetrode showing the four wires used to triangulate the spike waveforms from nearby neurons (left). Example analysis of tetrode spiking data depicting the latent representation of non-spatial stimuli extracted from the ensemble activity using an autoencoder (right). (c) Illustration of a silicon probe, with a zoomed-in micrograph of three probe shanks and their recording sites (left). Example decoding from a larger population of neurons showing the ability to recover the geometric properties of the manifold and temporal trajectories through the manifold (right). (d) Illustration of a two-photon calcium imaging microscope and glass window implantation over the rodent brain (left). Example *in vivo* optical imaging of mouse cortex showing activity (calcium transients) in cortical neurons (center). Example analysis showing that recording from several hundreds of neurons simultaneously (right, top) can enable proper quantification of manifold geometry and of the dimensionality of task-relevant representations, such as the total number of variables encoded during complex behavior (right, bottom). EEG, electroencephalography; MEG, magnetoencephalography; PET, positron emission tomography; 2-DG, 2-deoxyglucose. Figures adapted from Sejnowski et al. (2014) (panel a), Ferguson et al. (2009) (panel b, micrograph on left), Shahbaba et al. (2022) (panel b right), Han et al. (2024) (panel c, micrograph on left), Gardner et al. (2022) (panel c, right), and Kim et al. (2016) and Stringer et al. (2019) (panel d, right).

larvae (~50% of the total neurons; Bruzzone et al. 2021). Second, it allows for the recording of specific cell populations (e.g., pyramidal neurons, interneuron subtypes, astrocytes). Third, three-dimensional functional volumes of tissue (up to a certain depth) can also be reconstructed. To do so, a thin sheet of light is used to excite only a two-dimensional section of the sample. Moving this sheet vertically through the tissue captures multiple frames which can later allow for 3D reconstruction (Ahrens et al. 2012).

Optical imaging also has limitations that warrant consideration. First, these microscopes tend to be expensive, quite technically demanding, and require head fixation of the animal to facilitate control and stability of the imaging window. This unnatural restriction of movement can induce stress, disrupt natural behaviors like exploration, and can restrict proprioceptive feedback (Juczewski et al. 2020). Technological advancements such as virtual-reality navigation setups, miniature lower-resolution microscopes that are mounted to the animal's head (Aharoni et al. 2019) or mobile two-photon devices using optical fibers, may bridge the gap between controlled environments and natural behaviors (Helmchen et al. 2001; Zong et al. 2022). Second, the indirect relationship between calcium indicator transients and spikes can pose analytical challenges. The non-linear transformation, limited dynamic range, and inherent

low-pass filtering can distort spike timing and frequency and hinder accurate decoding of ensemble activity. Some of these limitations may be mitigated by voltage-sensitive indicators, which emit light as the voltage changes across a neuronal membrane (Ferezou et al. 2006). Like calcium indicators, however, voltage-sensitive indicators are potentially phototoxic and may alter neural dynamics, while genetically-encoded indicators exhibit expression variability and potentially disrupt cellular processes. Similarly, photobleaching, the photochemical destruction of a fluorescent calcium indicator caused by prolonged exposure to excitation light, can result in a permanent loss of signal intensity. For deeper brain regions, calcium imaging faces light scattering and tissue penetration limitations. This restricts recordings to neuronal populations closer to the cortical surface, or requires removal of large volumes of neural tissue to allow space for optical lenses. Finally, optical imaging faces the same computational and data storage burden as silicon probes.

3.1.3. Other Notable Recording Techniques

Beyond silicon probes and optical imaging, there are a number of important recording methods to consider. On one end of the spectrum (highest spatial and temporal

resolution; [Figure 3a](#)) is intracellular recording, which provides direct access to the electrical potential across a cell's membrane. In addition to recording spiking activity, this approach allows the recording of subthreshold activity, synaptic potentials, and ion channel kinetics that are not possible with other recording techniques (Lee and Brecht 2018). However, such high-fidelity recording of individual neurons comes with tradeoffs: intracellular recordings are technically challenging as they require the precise interface of an electrode directly onto a neuron. Unfortunately, the nature of the technique—whereby even movements of a few microns can disrupt this electrode-cell interface—limits its ability to scale beyond a handful of simultaneously recorded neurons.

On the other end of the spectrum are recording techniques available in humans, such as EEG and fMRI (Heeger and Ress 2002; Glover 2011; Buzsáki et al. 2012). Key advantages of these methods are their noninvasiveness and their ability to sample activity from the whole brain simultaneously. The main disadvantages of EEG are poor source localization (though this can be improved with MEG) and the fact that the signal is primarily limited to oscillations from the cortex just below the skull (i.e., little information about spikes or about deep brain structures). The main disadvantage of fMRI is that it relies on an indirect measure of neural activity, specifically the blood oxygenation level-dependent (BOLD) signal. While the BOLD signal correlates with neural activity, it is inherently a noisy marker and thought to more strongly reflect local field potentials than spiking activity. Particularly problematic is the delayed hemodynamic response (the increase in blood flow to activated brain regions) which means the recorded signal lags local information processing by approximately 5 s.

It is important to note that a growing number of studies have overcome limitations from one recording method by combining multiple methods within the same study. For example, extracellular and intracellular recordings can be combined for ground truth datasets (Henze et al. 2000). Anatomical methods such as structural or diffusion MRI can be combined with fMRI to determine how white matter pathways influence function. EEG has been successfully combined with structural imaging and fMRI as well (Warbrick 2022). Recording approaches can also be combined with active manipulations to examine causal relationships, such as drug manipulations, transcranial magnetic stimulation (TMS), or optogenetics in animals. The challenge of such studies is that they often depend on expertise from multiple domains and on analytical approaches integrating heterogeneous types of data.

3.2. Main Analytical Challenges and Their Emerging Solutions

3.2.1. Implementing Complex Preprocessing Pipelines for Electrophysiology

The quality of neural activity data depends upon the signal-to-noise ratio of the recording itself, as well as the quality of the preprocessing steps. Our focus here is on spiking activity but note that other types of electrophysiological data, like

EEG, also rely on preprocessing pipelines for artifact rejection and noise reduction (see Bigdely-Shamlo et al. 2015 for a detailed account of standardized preprocessing protocols in EEG). The critical preprocessing step for spiking activity is spike sorting, which often follows a standardized pipeline (Lewicki 1998). First, the activity of single spikes is extracted from the wideband signal through thresholding or filtering. Spikes are then differentiated based on standard waveform parameters (e.g., peak-to-valley amplitude or time) or template matching. These data are then passed to a clustering algorithm to separate them into clusters based on their features, with each cluster representing spikes from the same putative neuron. Lastly, the clustered neural activity data are curated and corrected based on clustering quality and other metrics, typically a laborious manual process that is often more of an art than a science and variable across experimenters (Wood et al. 2004).

Given the recent and projected increases in cell yield, the development of reliable automated spike sorting algorithms has become critical. In fact, preprocessing will soon become intractable using the manual approach alone and it is hoped that automated sorting algorithms will help minimize user interaction and allow for swifter analysis of large numbers of channels. Many potential solutions to the automated spike sorting problem have been proposed over the years (Pachitariu et al. 2016; Rossant et al. 2016; Hilgen et al. 2017; Diggelmann et al. 2018; Yger et al. 2018; Magland et al. 2020), with some even challenging its necessity in some cases (Trautmann et al. 2019). As mentioned previously, ground truth data that combine extracellular recordings with intracellular recordings from individual neurons within the same population are an important component of spike sorting algorithm development. However, these datasets are rare due to their technical difficulty (Yger et al. 2018).

In an attempt to integrate this fragmented software ecosystem, SpikeForest (Magland et al. 2020) allows multiple sorting algorithms to be implemented, benchmarked, and compared across a consistent interface with a growing dataset of ground truth data and other experimental data. The key problem is that these proposed solutions continuously face new and more complex challenges. These include increases in the numbers of channels that make previous solutions intractable (as for some implementations, processing time scales exponentially with channel number), long-duration recordings that disallow entire recordings to fit into computer memory, or novel requirements for spikes to be sorted in real-time (Zhang et al. 2023). Thus, despite its critical importance to neural data analysis, automated spike sorting remains a key data science challenge.

3.2.2. Implementing Complex Preprocessing Pipelines for Optical Imaging

The preprocessing of imaging data is also complex and challenging. Note that our focus here is on optical imaging though some of these challenges may also apply to other types of imaging methods like fMRI (see Poldrack et al. 2011 for a detailed account of standardized preprocessing protocols in fMRI). In contrast to spike sorting in

electrophysiology, optical imaging benefits from observing the precise location of individual neurons such that spike isolation becomes an “image segmentation” problem. In fact, researchers are tasked with identifying the precise location of each neuron within an image frame, and developing techniques to deconvolve recorded traces. Open source tools, such as Suite2p, have facilitated this process by providing a rendering pipeline to detect neurons and infer spike times from their measured calcium traces (Pachitariu et al. 2017). However, while such solutions are available, reliably separating neuronal signals from noise and movement artifacts and extracting population activity remains a challenging task that cannot be fully automated at this time. In addition, since calcium fluorescence traces represent an indirect and nonlinear proxy for spiking activity, analyses that require individual spiking activity to be inferred from the calcium traces also pose a challenge. A great deal of progress has occurred in this area and spike deconvolution methods are now standard practice. Validation with simultaneous electrophysiological recordings, when possible, provides a reliable ground truth for further development of deconvolution algorithms (Pachitariu et al. 2018; Tübiana et al. 2020). Given advances in image segmentation using deep learning approaches, it will be interesting to see how these advances transfer to the preprocessing of optical imaging data.

3.2.3. Visualizing Complex Activity Patterns

Modern scientific data visualization, and particularly the visualization of large and multivariate datasets, is an independent science in its own right (Fuchs and Hauser 2009). In fact, effective visualization has become critical for proper interpretation of multi-dimensional data: a survey of 288 neuroscience articles showed that graphical visualizations became less informative as the dimensions and complexity of datasets increased (Allen et al. 2012). This is particularly problematic for neural activity datasets, as such visualization needs to capture the temporal structure of task-critical neural activity at different timescales as well as its modulation across task parameters and dimensions. Adhering to established principles for software design and data visualization is thus more important than ever, such as presenting an overview of the data first, allowing the user or reader to zoom and filter as needed, and providing details on demand (Shneiderman 2003).

A wide variety of methods have been proposed to visualize complex neural datasets, particularly in cellular morphology analysis or connectomics where the scale of individual data points can reach into the trillions (Pastor et al. 2021). For example, to overcome the vast complexity of visualizing nano-scale neuroscience data, NeuroLines presents an abstracted overview of the topology of 3D tissue using a multiscale “subway map” style visualization (Al-Awami et al. 2014). In fMRI, principles of data intermixing, combining the anatomical image with high-dimensional statistical maps of neural activity have proved critical (Fuchs and Hauser 2009). A separate approach includes adding depth-dependent halos to fiber tracts for the illustrative rendering of dense multi-line data (Everts et al. 2009). Freely available

tools are increasingly aiding the visualization process as well, such as the MATLAB toolbox iELVis, which can be utilized for localizing and visualizing human intracranial electrode data (Groppe et al. 2017), or brainrender, a python-based package for rendering anatomical data, electrodes, or cells (Claudi et al. 2021). Open source libraries leveraging GPU advances will become increasingly important to neuroscience data visualization by overcoming the shortcomings of their CPU-bound tasks (Campagnola et al. 2015; Rossant et al. 2016). These have already resulted in packages such as Visbrain (Combrisson et al. 2019), an open source, python-based comprehensive visualization suite for neuroimaging and electrophysiological brain data.

3.2.4. Reducing the Dimensionality of Neural Activity Data to Capture Its Underlying Structure

The logic behind the use of dimensionality reduction approaches is that high-dimensional data can be viewed as multiple indirect measurements of an underlying pattern or source, and that reducing the number of dimensions in the data can help identify this underlying structure. Neural activity datasets are a particularly good example for this, as dimensionality reduction approaches have been shown to extract fundamental patterns in the collective activity of populations of neurons (Ganguli and Sompolinsky 2012; Cunningham and Yu 2014; Sussillo 2014; Vyas et al. 2020; Chung and Abbott 2021; Stringer and Pachitariu 2024). Let us assume we record from N neurons in a population and that we represent the activity of all neurons using a N -dimensional firing rate space (i.e., each neuron as one axis in this space). In theory, the population activity could traverse all N dimensions of this space. In practice, observations have shown that neural dynamics are low-dimensional in that they only traverse a limited region of the entire space (Ganguli et al. 2008; Yu et al. 2009; Machens et al. 2010), as correlated activity across groups of neurons constrains the number of unique representations (Sadttler et al. 2014). In fact, the consensus emerging from a wide range of recording studies is that much of population activity lies upon a low-dimensional subspace, H , with dimension q such that $q \ll N$.

Principal component analysis (PCA) is a widely used dimensionality reduction technique that identifies a set of orthogonal directions (principal components) that successively maximize the variance of the projected data, subject to orthogonality constraints. The observed neural data are then represented in a lower-dimensional subspace spanned by the leading q components (Chapin and Nicolelis 1999; Ahrens et al. 2012; Churchland et al. 2012; Mante et al. 2013; Pfau et al. 2013; Kobak et al. 2016). Another approach is non-negative matrix factorization (NMF; Lee and Seung 1999) which decomposes neural activity into basis vectors and corresponding activation patterns. Both PCA and NMF reduce data dimensionality, yet with different goals. Whereas PCA maximizes explained variance using linear projections, NMF instead minimizes reconstruction error subject to non-negativity constraints, yielding additive, parts-based components. The orthogonal components of

PCA can be used to reconstruct the data, though the interpretation may be less intuitive. NMF, on the other hand, can improve interpretability by computing the non-negative basis that can be additively combined to reconstruct the original data. NMF has been leveraged to capture complex patterns such as sequential firing in the rodent hippocampus or in the premotor cortex of songbirds (Mackevicius et al. 2019). Within NMF, sparsity can be further promoted by adding an $L1$ penalty to the factor matrices, encouraging many coefficients to become exactly zero; this enhances interpretability by highlighting only the most relevant components. These properties make NMF particularly well suited for modeling neural firing rates, as it naturally respects the non-negative nature of the data while producing components that correspond to meaningful and interpretable patterns of neural activity.

Beyond these common approaches, increasingly sophisticated dimensionality reduction methods have been developed in recent years (Churchland et al. 2012; Cunningham and Yu 2014; Pang et al. 2016; Griffanti et al. 2017; Holbrook 2017; Pellegrino et al. 2024). Given that the demand for such analyses spans a broad range of scientific fields, an extensive suite of techniques has emerged, including (but not limited to) independent component analysis (ICA), t-distributed stochastic neighbor embedding (t-SNE), Laplacian eigenmaps, locally linear embedding, multidimensional scaling, Gaussian process factor analysis, dynamical-system modeling, linear discriminant analysis, regularized regression models, and recently, various deep learning architectures, such as autoencoders (for review, see Sarveniazi 2014).

Methods that can accurately and efficiently reveal the low-dimensional space, H , from a given neural population can then support the analysis of how those dynamics vary across stimuli, brain regions, and task conditions. Such approaches have successfully identified novel population dynamics across multiple species and brain regions (Ahrens et al. 2012; Churchland et al. 2012; Mante et al. 2013; Sadtler et al. 2014; Gallego et al. 2018; Chaudhuri et al. 2019; Dabagia et al. 2023; Gardner et al. 2022; Shahbaba et al. 2022). The success of these approaches provides another example in which analytical tool development can lead to new scientific questions and experiments. For example, neurons in the medial entorhinal cortex, a key region of the brain's navigation system, have been shown to reliably predict an animal's orientation (head-direction cells; Taube 2007) and position in space (grid cells; Hafting et al. 2005). Biologically plausible models of this circuit showed that the recurrent connectivity in the circuit should constrain neural activity to a low-dimensional subspace, specifically a torus-shaped attractor (Skaggs et al. 1995; Redish et al. 1996; Zhang 1996; Fuhs and Touretzky 2006; McNaughton et al. 2006; Couey et al. 2013). Experimentally testing this theory, however, would require recording from hundreds of grid cells, which was outside the capabilities of recording technology at the time. The question remained unanswered for years, that is until advances in silicon probe technology increased the yield of cells and topological data

analysis approaches were further developed. To test this long-standing theory, Gardner and colleagues (2022) recorded data from large ensembles of entorhinal cortex neurons and then used a dimensionality reduction data analysis pipeline (PCA to improve noise robustness, UMAP for 3D visualization, and persistent cohomology for quantification; Rybakken et al. 2019). Consistent with the models proposed decades prior, they found that entorhinal cortex activity spanned a low-dimensional subspace, which exhibited toroidal topology, and that neural activity moved across the torus in relation to an animal's movement across its environment (Gardner et al. 2022; Figure 3c).

3.2.5. Accurately and Efficiently Classifying Complex Patterns

The need to accurately classify massive numbers of complex patterns is not limited to the domain of internet search engines, it represents an increasingly common task in neuroscience data analysis as well. A brief history of deep learning, illustrated by the example of developments in convolutional neural networks (CNNs), may be illustrative here. CNNs were inspired by the "Neocognitron" (Fukushima 1980), an early connectionist model that aimed to incorporate Hubel and Wiesel's findings of "simple" and "complex" neurons, local receptive fields, and feature extraction observed in the visual cortex. The Neocognitron also introduced ideas of convolutional layers and shared weights. Later, LeCun et al.'s (1998) further development of backpropagation for training CNNs on handwritten digits reignited interest in this area of research. The brain was, again, a key source of inspiration as they noted "[The brain's] local receptive fields neurons can extract elementary visual features such as oriented edges, endpoints, corners [...]. These features are then combined by the subsequent layers in order to detect higher order features" (Lecun et al. 1998).

The contemporary resurgence of deep learning approaches may in part be traced to the success of AlexNet (Krizhevsky 2014), an image recognition model which marked a turning point after achieving unprecedented accuracy on ImageNet, a dataset of over 1.2 million images and over 1000 classes. CNNs then grew in popularity and became widely used in the visual modeling field and have been applied as decoding models to neural activity data as well, including classification based on patterns of EEG activity (Aellen et al. 2021) or fMRI activity (Zhang et al. 2022). CNNs have also been used for difficult tasks such as decoding the sparse and distributed representations of non-spatial stimuli in the hippocampus. For instance, by integrating local field potential activity and spiking activity, a CNN could successfully decode the series of discrete stimuli presented to an animal, as well as demonstrate the sequential reactivation of an extended sequence of non-spatial events (Shahbaba et al. 2022Figure 3b).

In parallel, network science, a data science and mathematical technique to quantify patterns in complex systems that often relies on graph-theoretic methods, has also developed a mutually beneficial relationship with neuroscience over the last few decades (Sporns 2011; Barabási et al. 2023).

Network science has found much of its success in neuroscience through analyzing structural and functional fMRI (Sporns and Betzel 2016). In recent years, however, sufficient data can be collected in animal studies for the methods to become more readily applied, such as in the analysis of LFP activity data across channels (El-Yaagoubi et al. 2025).

Along these lines, methods that analyze or reconstruct functional networks of multiple neurons given their spiking activity are also of growing interest. These methods required the development of multivariate extensions to models traditionally used for the analysis of single spike trains (Brown et al. 2004; Kass et al. 2011; Ramezan et al. 2022). For example, Hawkes processes, a class of self-exciting temporal point processes in which the conditional intensity function depends on the history of past events (Hawkes 1971) have been used to analyze neural populations (Pillow et al. 2008; Lambert et al. 2018; Chen et al. 2019; Tang and Li 2023), and have motivated a number of extensions (e.g., allowing for inhibition) to the more classical models.

3.2.6. Neural decoding

Neural decoding refers to the process by which, given a pattern of neural activity, one can reconstruct the stimulus or task variables coded in the neuronal population. Neural decoding often entails a two-step process. First, a given model is trained on the relationship between the neural activity and external stimuli or conditions (or learns the relationship intrinsically) using a subset of the data. Second, the model is tested using the other subset of the data to identify periods during which the neural activity resembles the trained patterns. For example, given the firing of a population of neurons in the human medial temporal lobe, it is possible to decode, and even predict ahead of time, which picture was presented to the subject at above-chance levels (Quian Quiroga and Panzeri 2009). The past few decades have seen remarkable progress in this area, with a range of methodologies that have been applied across brain regions and species (for reviews see Quian Quiroga and Panzeri 2009; Vu et al. 2018; Xu et al. 2019; Glaser et al. 2020; see examples in Figure 2h, Figure 3b-d). While neural decoding is inherently supervised (i.e., relies on label information), it is common to incorporate unsupervised methods (which ignore label information), such as the dimensionality reduction techniques discussed above, as an intermediate step to capture underlying structure in the data and thereby improve the model's generalizability and interpretability.

For supervised methods, Generalized Linear Models (GLM) as well as nonlinear regression and classification models are commonly used for neural decoding across stimulus modalities and species (e.g., Shahbaba et al. 2022; Weng et al. 2024). A classic example is that of human brain-computer interfaces (BCI), which allow for the control of robotic arms by decoding neural activity in the motor cortex of individuals with paralysis (e.g., Collinger et al. 2013). Bayesian models are another widely used approach, whereby a given trial, time period, or stimulus is assigned to the class with the highest posterior probability. For instance, these models may be used to recover a single posterior

distribution $P(\text{state}|\text{spikes})$, which specifies the probability of a given state (stimulus, time, condition) conditioned on a pattern of spiking activity. This approach has been widely used to decode the spatial location of the animal (Zhang et al. 1998) as well as specific movements (Sanger 1996; Brown et al. 1998; Oram et al. 1999), visual stimuli (Ziaei et al. 2023), moments in time (Shahbaba et al. 2022), and many other task variables. Another approach is support vector machines (SVM), which is particularly common in fMRI analysis pipelines. This approach projects data into a high-dimensional space through feature augmentation in order to find a hyperplane that optimally separates the data, leading to decision boundaries in the original space that can be linear or non-linear (Cortes and Vapnik 1995; Smola and Schölkopf 2004; Chang and Lin 2011). A comparatively simpler approach is k -nearest neighbor (k-NN) algorithms, which leverage local similarity by classifying new data points based on a distance metric of their k nearest neighbors in the feature space. These algorithms exhibit competitive performance for diverse tasks, due to their non-parametric approach and flexibility.

3.2.7. Quantifying the Information Contained in Neural Activity (Information-Theoretic Methods)

Information theory provides a rich mathematical framework for quantifying how neural systems encode, transmit, and process signals. Despite the extensive use of information theory across neuroscience and cognitive science (Richmond and Optican 1990; Quian Quiroga and Panzeri 2009; Friston 2010; Timme and Lapish 2018; Varley 2025), ambiguity remains regarding the precise interpretation of some of its key concepts, particularly entropy (Varley 2025). Based on Shannon's foundational work in the mid-twentieth century (Shannon 1948), the Shannon information content quantifies the surprise associated with a single event, $I(A) = -\log p(A)$, so that unlikely events carry more information, while entropy, $H(X) = E[I(X)]$, captures the average surprise of a random variable. Shannon's work on information theory has been extended by applying it to physical and scientific systems, where entropy is interpreted as a measure of missing information (Brillouin 1956), and to statistical mechanics, where the maximum entropy principle is used to infer probability distributions under known constraints (Jaynes 1957). Contemporary work builds on these foundations, connecting information-theoretic principles with modern machine learning methods, including Bayesian modeling, Monte Carlo and variational techniques, clustering algorithms, and neural networks (MacKay 2003).

In machine learning, a maximum entropy model is the probability distribution that satisfies a given set of constraints while having the highest possible entropy. The resulting model is the least structured (least biased) distribution consistent with the information encoded in those constraints. This idea has a natural connection to statistical mechanics: physical systems in thermal equilibrium follow the Boltzmann distribution, which is the maximum entropy distribution under a fixed mean energy (Jaynes 1957). As a result, every maximum entropy distribution corresponds to

an associated energy function, which, in the case of a network of neurons, takes the form of an Ising model (Schneidman et al. 2003, 2006). Unlike traditional uses of the Ising model in machine learning, where it is motivated by assumptions about network dynamics (Hopfield 1982; Amit 1992), Schneidman et al. (2006) showed that this model arises directly as the least-structured distribution consistent with observed spike rates and pairwise correlations, representing an exact correspondence rather than an analogy. Building on this perspective, information theory and machine learning can be viewed as complementary paradigms for modeling complex systems, and the brain serves both as a complex system with a remarkable ability to compress and communicate information and as an inspiration for modern machine learning methods (MacKay 2003).

In the brain, individual neurons can be conceptualized as probabilistic communication channels, allowing their variability and potential capacity to be quantified. Mutual information is commonly used in neuroscience to measure the statistical dependence between spike trains and an external variable. If knowledge of a neuron's firing pattern reduces uncertainty about a given stimulus, mutual information quantifies the amount of information encoded in the neural activity. In other words, if the spike train reliably predicts the stimulus, it carries significant information about it. This approach has been successfully applied to infer an animal's position from hippocampal spiking activity (Skaggs et al. 1992; Barbieri et al. 2004), to quantify visual information in the lateral geniculate nucleus (Oram et al. 1999; Reinagel and Reid 2000), and to study coding properties of the retina (Laughlin et al. 1998). While these models provide valuable insights, they have limitations. Mutual information assumes knowledge of the true stimulus, which is often unknown in complex tasks, and its calculation can be intractable in high-dimensional settings.

Moreover, information theoretic methods are susceptible to high variance in the estimation of neural response curves in the case of small sample sizes, which limits their application (for a thorough review, see Panzeri et al. 2007). Although modern recording techniques enable simultaneous recording of many neurons, the number of subjects remain relatively small and the observed data tend to be temporally sparse. Importantly, recent theoretical work has proposed potential solutions to the temporal sparsity problem (Sheintuch et al. 2022) and multivariate extensions overcoming the computational costs of previous methods have been developed (Rosas et al. 2019; Newman et al. 2022; Varley et al. 2023). Given these advances, there has been renewed attention in leveraging these methods to test long-standing hypotheses about complex information dynamics in brain networks (Varley et al. 2023).

3.2.8. Population Inference vs. Individual Differences

Balancing population-level inference with individual-level understanding is a significant challenge in neuroscience, a challenge that has grown in complexity with the size and dimensionality of the datasets (Hampel et al. 2019; Kraus et al. 2023; Mattoni et al. 2025). In the extreme, one can view group statistics without individual insight as describing

an average brain that does not exist, whereas analyses focusing on individuals without group constraints risk overfitting and confounding idiosyncrasies for general principles. Perhaps we should view our ultimate goal as hierarchical understanding: identifying shared principles, processes, and strategies underlying specific functions, as well as individual (or subgroup) variations in the degree to which each process or strategy is engaged. In the end, the degree to which a given experimental or analytical design focuses on shared or individual features depends on the scientific question, and specifically on whether heterogeneity should be considered signal or noise in that context.

An emerging set of analytical tools may also be useful in this consideration. Given a finite amount of resources, scientists often have to choose between adding more subjects or collecting more data from a smaller number of subjects. In such cases, Bayesian utility theory can be used to quantify the expected information gain associated with each experimental design and to identify the design that maximizes this quantity (Lomeli et al. 2021). The optimal tradeoff depends on both technical and biological variability. While such tools were originally developed in the context of examining group effects, they can be extended to the investigation of individual differences or subgroups to provide a more rigorous framework to design this type of experiment and analyze the resulting data.

3.3. Opportunities to Leverage

3.3.1. Uncovering Emergent Properties of Neural Ensembles

Data from modern neuroscience techniques present a unique opportunity to leverage the larger yields of neural activity data toward uncovering emergent properties in neural systems. In fact, the traditional reductionist approach of studying components in isolation can be complemented by the integration of diverse datasets with advanced statistical and machine learning methods to examine their interconnectedness and interactions to promote understanding of the system as a whole. Such data-driven approaches complement classical experimental design by allowing researchers to identify complex, non-linear interactions within large-scale neural networks that give rise to emergent phenomena—properties that cannot be predicted from studying individual neurons alone. It is those emergent properties that have long been thought to underlie how the brain represents information and performs complex cognitive operations such as perception, memory, and decision-making.

3.3.2. Leveraging Recent Developments in Deep Learning

There has long been a positive feedback loop between neuroscience and artificial intelligence (AI; Cooper 2018; Hassabis et al. 2017; Reeke and Edelman 1988; Turing 1950; Zador et al. 2023). The CNN model described earlier is one specific architecture out of a much broader class of supervised deep learning algorithms that can be applied as a general tool for analysis of neural data. Deep learning approaches offer powerful statistical machine learning algorithms with a large number of adjustable

parameters which allow them to learn complex non-linear relationships in the data. They have been included across multiple steps of the data analysis pipeline, ranging from preprocessing steps to neural decoding, and their use in neuroscience is expected to continue to grow. Like other machine learning methods, deep learning approaches can be classified by their architectures (vanilla neural network, CNNs, RNNs, LSTMs, etc.) and by how they are trained (e.g., supervised, unsupervised, semi-supervised, or self-supervised). For example, the LFADS (Pandarinath et al. 2018) approach leverages sequential autoencoders, a non-linear dimensionality reduction technique sensitive to time, which is able to capture latent dynamics of neural spiking activity on timescales of a few milliseconds. Importantly, compared to more commonly used approaches in neuroscience such as PCA, LFADS allows for decoding trial-by-trial latent dynamics and incorporates non-linearities. Improvements have also been shown using deep learning architectures, such as CNNs, for spike estimation from calcium imaging data (Sebastian et al. 2021). How these models are trained has also been shown to be a key factor in their performance. For example, RADICaL (Zhu et al. 2022), which uses a similar architecture to LFADS but adds a noise emission model and a novel training strategy, has been shown to improve performance on calcium imaging data. Some approaches have also employed augmented training data—where the original training set is expanded or modified to improve model robustness—such as to enhance performance of long short-term memory (LSTM) architectures (Schlafly et al. 2025). Following a similar training philosophy, CASCADE uses a suite of supervised neural networks to estimate spikes from calcium imaging data (Rupprecht et al. 2021). Interestingly, no parameters need to be adjusted by the user during training, as CASCADE automatically resamples ground-truth data to match the sampling rate and noise level of the recordings.

Finally, deep learning provides another example in which analytical tool development can lead to new scientific questions and experiments. While deep learning architectures can simply be viewed as a tool to analyze data (e.g., similar to regression or PCA), they can also be viewed as a high-level computational model of a neural system, in which the nodes and edges of the network represent an abstract model of a biological neural system. As such, this type of model may be more likely to discover brain-like solutions to analytical problems. For example, an appropriately-tuned CNN model trained on the same images as rhesus macaques has been shown to exhibit similar layer-wise responses as those observed in the monkey's visual system (Yamins and DiCarlo 2016), with similar relationships also reported in the fusiform gyrus (Mohsenzadeh et al. 2020) and auditory cortex (Kell et al. 2018). Such integration between examining the representations and behavior of deep learning systems and of brain systems will facilitate the generation of future testable hypotheses that may have otherwise been difficult to formulate.

3.3.3. Opening the Black Box through Explainability Methods

Deep learning algorithms excel at identifying subtle patterns and relationships within vast datasets and thus can reveal

insights likely unattainable with traditional methods, which often struggle with the sheer volume and intricate nature of neural activity data. However, the complex non-linearity, hierarchical structure, and high-dimensional parameter space of deep learning approaches come at a high cost, namely that of explainability. Although solutions are being developed, this is an area of research with a considerable unmet need at this time. To mitigate this issue, several gradient-based and perturbation-based interpretability methods have been proposed (Simonyan et al. 2014; Zeiler and Fergus 2014; Zhou et al. 2015; Ribeiro et al. 2016; Selvaraju et al. 2017). These methods help interpret network features by calculating which features contribute most to the final prediction of the network, or which can detect discriminant patterns of activity within individual data samples. For more complex architectures, increasingly complex methods for explainability can be developed or adapted to address specific questions (Zhou et al. 2025). In addition, recent advances in physics-informed machine learning (PIML) integrate prior physical knowledge into data-driven models, reducing data requirements, improving generalizability, ensuring physical plausibility, and providing a more proactive mechanism for improving interpretability (Meng et al. 2025). Extending this direction, hybrid biologically informed machine learning (BIML) frameworks that couple mechanistic models with modern learning algorithms are rapidly gaining traction (Daryakenari et al. 2024).

Understanding how neural networks solve these complex problems is important for two main reasons. First, it is necessary for trusting the output, reproducing the findings, as well as improving task performance (Zeiler and Fergus 2013). Second, the modeling of large-scale neural networks holds immense promise for simulating brain activity and predicting behavior. As these tools continue to evolve, their integration with neuroscience research will offer a family of powerful new approaches to understand complex dynamic systems like the brain.

Nevertheless, important tradeoffs remain when selecting these analytical methods. Traditional statistical approaches are often designed to balance predictive power with interpretability, whereas many machine learning methods rely on highly complex models that achieve strong predictive performance at the cost of interpretability. Given the complexity of the problems involved, we expect both types of methods (and their integration) to be critical to the future of neuroscience. Much like we expect both top-down and bottom-up experimental approaches to remain critical.

3.3.4. Multimodal Data Integration Strategies

Neuroscience research will increasingly involve the analysis of large volumes of data from diverse sources, such as electrophysiological recordings, imaging data, molecular assays, and behavioral measurements. Effectively integrating these complex, multimodal datasets can provide a more comprehensive understanding of brain function, reveal novel neural mechanisms, and support more informed experimental design and therapeutic strategies (Bzdok and Yeo 2016). This integration can be performed by data fusion

frameworks, which are generally classified as early, intermediate, or late, depending on when fusion occurs. Early fusion merges raw modalities before feature extraction, thereby capturing rich cross-modal interactions but risking overfitting and poor performance with noisy or heterogeneous data. Late fusion combines predictions from separate models per modality; it is simple and robust but may miss complementary information across modalities. Intermediate fusion integrates low-dimensional latent representations from each modality to balance the extreme strategies above, but is dependent on the quality of the learned representations (Holbrook et al. 2017; Shahbaba et al. 2022; Arkhipov et al. 2025; Liang et al. 2025). To avoid making an *a priori* commitment to a particular fusion stage, recent methods such as cooperative learning (Ding et al. 2022) aim to automatically determine when fusion should occur. Note that data integration methods can also be leveraged to capture shared patterns across heterogeneous subjects (Sutter et al. 2024; Yuan et al. 2025).

3.4. Lessons to Remember

Technical advances in recording techniques, particularly silicon probe and optical imaging technologies, have led to rapidly increasing yields of simultaneously recorded neurons. This transition to large-scale neural activity data has (appropriately) shifted the focus from the neuron-centric era of the past to a population-centric and distributed view of information processing across neural circuits. However, the remarkably complex nature of activity dynamics across neural populations, as well as the nonlinear relationships found in these data, have posed significant analytical challenges. As in the past, there remains a significant lag between the technical advances resulting in increasingly higher cell yields and the development of analytical tools to properly analyze the data produced. Moreover, as data collection methods continue to grow in scale, reliability, and ease of use, laboratories may now be collecting more data than they have time or in-house expertise to analyze. To handle these modern barriers in systems neuroscience, a marked transition from single investigators, to single labs, to multi-lab teams is becoming increasingly apparent and necessary.

4. The Future

“A clear grasp of simple and standardized statistical procedures will [...] go far to elucidate the principles of experimentation; but the procedures are themselves only the means to a more important end. Their part is to satisfy the requirements of sound and intelligible experimental design, and to supply the machinery for unambiguous interpretation. To attain a clear grasp of these requirements we need to study designs which have been widely successful in many fields, and to examine their structure in relation to the requirements of valid inference”

—Fisher, *The design of experiments* (1936)

How will the methods of today meet the needs of tomorrow? While established methods to record brain activity have yielded a wealth of new data and insights, our current

understanding of the brain can only be considered superficial. Further progress will depend on developments that can get us closer to capturing and modeling the brain’s full scale and complexity. Here, we outline how contemporary methods are being adapted and augmented to address the challenges and opportunities expected to define the future research landscape. We also emphasize throughout that methodological advancements are no substitute for good experimental design, practice, and scientific theory.

4.1. Recording Techniques: Future Developments

4.1.1. Exponential growth in channels, cell yields, and file sizes

At present, data collection for some neuroscience studies is approaching file sizes in the terabytes. Echoing Moore’s law (which describes the doubling of transistors on computer chips every 18 months), Stevenson and Kording showed that neural yields have doubled every ~ 7.4 years (Stevenson and Kording 2011). This rate was based on an analysis of over 50 neurophysiology studies as of 2011 and, when revisited almost a decade later in 2020, the predicted doubling appears to have been met if not surpassed (Urai et al. 2022; Figure 4). By extrapolating this relationship, one can surmise that recording from virtually *all* neurons of the mouse brain may be possible in the future, with estimates as early as a few decades from now (Urai et al. 2022). Should this pattern hold, in the case of the mouse brain with its ~ 70 million neurons (Herculano-Houzel et al. 2006), recording a 1-hour session at 30 kHz would reach ~ 15 petabytes. A primate brain recording with the same parameters could reach into the exabytes. Even recording from 1% of all neurons, which may be a more reasonable upper limit, would lead to staggering file sizes. While it is difficult to project future neuronal yields, one assumption is clear: the current growth rate

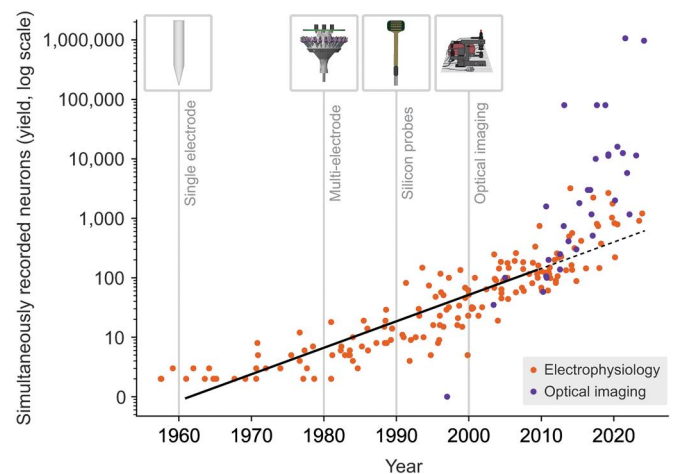


Figure 4. Exponential growth in neuronal yields. Each dot represents one publication, with electrophysiological and optical imaging studies indicated by red and blue dots, respectively. The solid line shows the original relationship (least-squares regression) described by Stevenson and Kording (2011) which indicates that neuronal yield has doubled every ~ 7.4 years up to that point. The dashed line shows the extrapolation of this relationship to the present time, which underestimated the recent drastic increases in yield obtained through optical imaging. Figure adapted from Stevenson and Kording (2011) and Urai et al. (2022).

is unlikely to slow and it will lead to significant challenges for data management and analysis.

4.2. Expected Analytical Challenges and Potential Opportunities

4.2.1. Issues with Data Storage, Maintenance, and Sharing

As the size of datasets grows in the future, the infrastructure to manage the data and support analysis will need to scale accordingly. This infrastructure includes the hardware and software solutions for storage and management of large datasets: whether in-house, supported by the home institution, or by a third-party company. Though data storage costs have tended to decrease over time, whether this trend will continue is unclear. The increased scale of datasets will also strain analysis pipelines since preprocessing, modeling, statistics, and hypothesis-testing steps will all increase in complexity. Thus, some current approaches may not continue to be practical analysis methods in the future. Similarly, when designing experiments, consideration of data sharing can no longer be an afterthought. At present, even smaller datasets are not always readily shared, and, if shared, they may come with minimal explanation, metadata, or documentation. Larger datasets will only increase the complexity and time required to share and interpret raw data, which is directly at odds with the expectation that data sharing will increase across neuroscience subdisciplines. To mitigate these barriers to reproducibility, the field is increasingly adopting standardized frameworks such as the Brain Imaging Data Structure (BIDS) for electrophysiology and optical imaging data (Gillon et al. 2025).

4.2.2. Unnecessary Duplication of Efforts

Neuroscience research requires a range of methodologies which span multiple levels of neural organization—from molecules, to neurons, to neural circuits, to brain regions (Figure 3a). This range, in turn, results in heterogeneity in experimental data types (e.g., continuous vs discrete), structure (e.g., independent units vs networks), and formats (e.g., file type). This variety significantly limits the ease to which data can be integrated, synthesized, and shared across labs. As an analogy, consider if every camera produced their own image format. All downstream processing of that image would require custom conversion scripts to be written across devices or services resulting in much duplicated work. The solution to this problem was to put into practice standards for file formats (e.g., .jpeg for images) which most computer users became comfortable and familiar with. In the same spirit, there has been a surge in interest to consolidate and standardize heterogeneous brain activity data formats, with a number of solutions proposed but with little convergence (Table 1). Perhaps a key factor driving this fragmented ecosystem is that migrating users to any new software platform requires a learning curve and may offer little, if any, immediately apparent benefit to the researcher. Further, the platforms themselves may suffer from restrictions such as

limited funding, lack of features, or falling short of meeting all of the users needs.

The central challenge of scientific software frameworks is simultaneously architecting them for flexibility (to handle the rapid development of new recording technologies and analysis pipelines), while maintaining ease of use and stability. Due to the time and engineering involved, this often results in data analysis routines largely based upon lab-specific code. Code is rarely re-used, and often re-written for each scientist or each project. This approach adds additional time to the research cycle, is susceptible to error, and impedes reproducibility and sharing with the broader community. Fortunately, a growing number of open source resources are now available to researchers (Table 2). While the neuroscience community has long had a culture of openly sharing code (usually upon request), the last decade has witnessed a marked increase in the amount of open source material available to researchers.

4.2.3. Properly Addressing the Curse of Multidimensionality

A major motivation for increasing the size of datasets in neuroscience is often rooted in statistics itself: increasing sample sizes, whether of neurons or subjects, leads to more adequate sampling and improved power. Increasing sample size is not a one-shot solution to testing hypotheses, however, particularly as larger datasets often bring increased heterogeneity, noise, and incidental endogeneity (Fan et al. 2014). Renewed and contextual consideration of statistical best practices may need to be emphasized in this big data era, as older methods risk becoming less applicable or are increasingly relegated to preprocessing steps of more complex analysis pipelines.

In an age where we will have more neurons to analyze than ever before, a corresponding increase in the number of false positives is unavoidable. This point was more thoroughly discussed recently (Harris 2020), in which ten common methods of controlling for nonsense correlations in neural data were examined. The author found that only two methods, pseudo session and session perturbation, reliably do not produce nonsense correlations between time series. The fact that these methods cannot be applied in all cases, however, highlights the importance of accounting for randomness early in the experimental design stage, especially for experiments examining changes in neural dynamics over time (Harris 2020).

A number of statistical considerations have been proposed to increase robustness against false positives as neural data scales. For example, when decoding data in a supervised setting, performance is typically quantified by the ratio of hits to misses of decoded trials (provided by a confusion matrix) or by cross-validation. Statistical significance can be determined by establishing a proper baseline through analytically defining the probability of a model achieving more hits than chance. For complex or unusual designs for which calculating such baselines may be difficult, resampling approaches such as permutation tests and bootstrapping offer broadly applicable solutions. Importantly, these resampling methods should preserve as many features and structural properties of the original data

Table 1. Examples of standardized data formats for brain activity data.

Format	Description	Citation	Modality
BRAINformat	The BRAINformat library aims to create a general framework for standardization of scientific data. It defines application-independent design concepts and modules, is open source, and utilizes the HDF5 data format.	Rübel et al. (2016)	Electrophysiology
KWIK	KWIK files were developed for the Klusta suite spike sorting software to contain the results of the spike sorting session. Based on HDF5, these files include information about the spike times and spike clusters, as well as all the metadata.	Kadir et al. (2014)	Electrophysiology
NEO	NEO succeeded the Neuroshare API, which provided a design specification for core neuroscience data objects. NEO similarly aims to decouple the data representation step in analysis pipelines to create a unified format facilitating downstream analysis in software pipelines, packages, and frameworks.	Garcia et al. (2014)	Electrophysiology
ONE	Developed to standardize data sharing among multiple laboratories involved in the International Brain Laboratory (IBL) consortium, ONE defines a set of conventions for naming and organizing data files. Because ONE utilizes a 'object.attribute.extension' naming scheme, it can be used across file formats.	International Brain Laboratory et al. (2023)	Electrophysiology, Optical imaging
NWB	NWB provides a common standard to share, archive, use, and build analysis tools for diverse neurophysiology data. Based on the HDF5 format, NWB was designed to support data from intracellular and extracellular electrophysiology, optical imaging, as well as behavioral data.	Teeters et al. (2015); Rubel et al. (2022)	Electrophysiology, Optical imaging
Orca	Orca (also called BORG) was developed by the Allen Institute for Brain Science and designed to store electrophysiology and optical imaging data. It is based on the HDF5 format.	Godfrey (2014); Friedsam (2016)	Electrophysiology, Optical imaging
NIX	The NIX file format was designed to store diverse neurophysiology data, including electrophysiology and optical imaging, as well as behavioral data. Based on HDF5, NIX aims to provide a flexible and robust standard for organizing and sharing complex neuroscience datasets in a hierarchical and extensible way.	Martone (2014)	Electrophysiology, Optical imaging
EDF, EDF+	An early standardized data format, EDF was used for the storage of sleep recordings and included EEG, respiration, body temperature, EMG, and other related data. With EDF+, the format was expanded to provide additional flexibility.	Kemp et al. (1992); Kemp and Olivan (2003)	EEG
MEF	MEF provides open source software to read, write, and process large volumes of neural data. It utilizes range encoding to provide a high degree of data compression, a 128-bit encryption system for securely storing patient information, and a redundancy check to verify the integrity of the compressed data.	Brinkmann et al. (2009)	EEG, Electrophysiology
BIDS	BIDS is a standard for organizing and describing neuroimaging and EEG data to promote reproducibility and interoperability. It defines a structured format to arrange files and metadata in order to facilitate the processing and analysis of datasets across tools and the sharing across labs.	Gorgolewski et al. (2016)	EEG, Neuroimaging
NIFTI	The NIFTI file format is designed to simplify the storage of neuroimaging data. It packages both image data and its associated metadata into a single file to facilitate analysis and sharing.	Cox et al. (2004)	Neuroimaging
DICOM	DICOM is an international standard for medical imaging. It defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use. This format is widely used in healthcare (X-ray, CT, MRI, ultrasound, etc.) as well as in neuroimaging.	NEMA (1993)	Neuroimaging, Other

An updated and expanded version of this list is maintained at www.neurodatascience.org.

as possible, except for the specific effect under investigation. For example, to test the influence of spike timing, jittering the precise timestamp of each spike by some random amount within a given interval allows to preserve all coarse temporal structure and remove all fine temporal structure (Oram et al. 1999; Harrison and Geman 2009; Platkiewicz et al. 2017; Grün et al. 2022).

4.3. Considerations and Recommendations for the Future of Neurodatascience

4.3.1. Lessons from the Past: Increasing Synergies between Analytical and Technical Development

The past decades of neuroscience research witnessed the field iterate through cycles which first promoted technological developments needed to acquire neural activity data, and later promoted the development of analytical tools needed to accurately and thoroughly process the resulting data. This cycle results in an inevitable lag that significantly

impedes scientific progress and does not maximize return on investment. One can easily understand why the collection of novel datasets would, at first thought, seem more valuable than further analysis of existing datasets to neuroscientists, funding agencies, and the general public. However, more balanced support of technological, scientific, and analytical developments would reduce this lag and be more efficient in the long term. In that regard, there are a few neuroscience funding mechanisms that focus on the development of analytical tools (e.g., under the BRAIN initiative or CRCNS) and the growth of such programs will be critical to address this growing need. While another approach to analytical development is folding it into experimental proposals, this reduces the effort on the project and, since reviewers rarely have expertise in both domains, funding such proposals can be more difficult.

It is important to realize that large datasets that are already collected but remain under-analyzed are costly—whether it be the opportunity cost of testing new hypotheses more quickly, further characterizing potential treatments, or

Table 2. Examples of open source analytical packages for brain activity data.

Package	Description	Citation	Topic	Language
Cascade	Cascade translates calcium imaging $\Delta F/F$ traces into spiking probabilities or discrete spikes using deep learning networks.	Rupperecht et al. (2021)	Preprocessing (spike inference)	Python
BSD	A software package for inferring spike trains from fluorescence recordings. BSD features a fully unsupervised estimation of metaparameters and a blind sparse deconvolution algorithm.	Tubiana et al. (2020)	Preprocessing (spike inference)	MATLAB
LZeroSpikelnference	A collection of algorithms to determine the exact time at which a neuron spikes on the basis of its fluorescence trace using an $L0$ penalty.	Jewell and Witten (2017)	Preprocessing (spike inference)	R
Pynapple	A lightweight package designed to process a broad range of time-resolved data in systems neuroscience. It provides a set of data objects and methods aimed at simplifying the analysis of neurophysiological and behavioral data.	Viejo et al. (2023)	Neural signal processing	Python
GhostiPy	An open source software toolbox with implementations of various signal processing and spectral analysis tools. It prioritizes performance and efficiency through the use of parallelized algorithms.	Chu and Kemere (2021)	Neural signal processing	Python
Neurodsp	A software package specifically designed for analyzing neural time series data and their time-varying properties.	Cole et al. (2019)	Neural signal processing	Python
Chronux	An open source software package for the analysis of both point process and continuous data that includes preprocessing, exploratory analysis, and confirmatory analysis.	Mitra and Bokil (2008)	Neural signal processing	MATLAB
NetworkToolbox	A toolbox implementing network analysis and graph theory measures commonly used in neuroscience, cognitive science, and psychology.	Alexander (2018)	Graph theory and network analysis	R
BCTPY	A software toolbox implementing many graph theoretical measures to analyze brain activity. Based closely on the Brain Connectivity Toolbox for Matlab.	LaPlante (2013)	Graph theory and network analysis	Python
Brain Connectivity Toolbox	A popular toolbox used for complex brain network analysis that implements a number of graph theoretical measures.	Rubinov and Sporns (2010)	Graph theory and network analysis	MATLAB
Syntropy	An open source library for multivariate information theoretic analysis of discrete and continuous data.	Varley (2025)	Information theory	Python
Spyglass	An open source data analysis framework that facilitates the storage, visualization, analysis, and sharing of neuroscience data to promote more reproducible research.	Lee et al. (2024)	Analysis pipeline management	Python

An updated and expanded version of this list is maintained at www.neurodatascience.org.

maximizing the impact of federal funding. While most scientists are undoubtedly aware of this problem, this awareness has not generated sufficient urgency for concrete action. A change in culture at all levels will be critical to overcome these limitations and acknowledge that data analysis is a critical part of the modern scientific process.

4.3.2. Lessons from Other Big Data Fields

To overcome future challenges at the interface of neuroscience and data science, it may help to examine approaches taken by more mature fields. For example, astronomy and astrophysics have maintained collaborative international projects to map the sky dating to as early as 1887 (Turner 1912) and are currently undertaking petabyte-scale observing runs. Doing so has required a mature data infrastructure, pioneering approaches to open access of both publications and data, approaches to handle highly-collaborative and large authorship projects, and efforts to educate scientists on proper software development and data management practices.

As with the brain, the vastness of the cosmos is mirrored by the vastness of astrophysical data. It was recognized early that few, if any, astronomers would be able to locally store a

private copy of the data needed to conduct their analyses (Szalay and Gray 2001). For example, considered among the most important sky surveys in the field, the Sloan Digital Sky Survey (SDSS; Stoughton et al. 2002; Ahumada et al. 2020) has generated over 1.3 petabytes of data by the end of its most recent observing run. More contemporary survey projects will exceed this volume drastically. For example, the Rubin Observatory's Large Synoptic Survey Telescope (LSST) is expected to produce over 60 petabytes of data, with other proposed project estimates reaching into the exabytes (Bosch et al. 2019; Jones et al. 2020).

Management of this globally collected and vast scale of astronomical data has resulted in well established cyberinfrastructure. A key example is the Virtual Observatory (VO), which provides standards that describe all astronomical information resources worldwide, and enables universal access to these collections. For astronomers, VO acts as a gateway that seamlessly interrogates various data centers and offers powerful analysis and visualization tools within a unified system. This is achieved through standardized data, communication protocols, and a central registry displaying available services and functionalities. Crucially, VO established a dynamic and federated framework, rather than a singular software, which

has helped foster collaboration among data centers and developers (Brunner et al. 2002). This wealth of data is managed by a consortium of scientific and government organizations centered around astroinformatics and statistics (Zhang and Zhao 2015).

Analysis of large-scale astronomical data has resulted in many proper software and project management practices that have been shown to be successful in scientific software development. This is particularly true when domain scientists collaborate with software developers. For example, to address the need to coordinate multiple data collection organizations to detect supernovae, the Agile-Scrum project management process was shown to be effective for both scientists and developers (Baxter et al. 2022). For large-scale project management, the successful template of SDSS can be abstracted as follows: a large and invested research community, early data availability, open access, breadth of scope driven by specific aims, as well as a dedicated and integrated community of researchers, developers, and agencies. Lastly, long-term maintenance of project data is also critical. In a study of all articles published between 1997 and 2008 in the four main astronomy journals, the availability of linked material (URLs) markedly decayed over time; in fact, by 2011, 44% of links published a decade earlier were broken. Links to data hosted on astronomers' personal websites became unreachable faster than links to datasets on curated institutional sites (Pepe et al. 2014). This highlights the importance of shared repositories and of proper data sharing and archival approaches. Collectively, these approaches tested in other fields may provide a roadmap for future neuroscience research to follow.

4.3.3. Increased Importance of Team Science and Collaborative Platforms

The technical systems needed to observe billions of stars are more than a metaphor for the pursuit of observing billions of neurons in the brain. The informatics infrastructure developed for the Sloan Digital Sky Survey serves as the backbone for the Open Connectome Project (Vogelstein 2011), where a petabyte of semantically-labeled neuroscience electron microscopy data is freely available for analysis (Burns et al. 2013). Also mirroring the successful "amateur astronomy" model (Buxner et al. 2021), large-scale, public-facing, gamified neuroscience tools such as Eyewire have been developed (Kim et al. 2014). This crowdsourced approach has revealed novel circuits and cell types, and recently contributed to the neuronal wiring diagram of the adult *Drosophila melanogaster*, with 5×10^7 synapses and 139,255 neurons (Dorkenwald et al. 2024). The metaphor also extends to large-scale projects from the Allen Institute (Koch and Reid 2012). The institute has published comprehensive datasets on rodent neuroanatomy (Lein et al. 2007), connectivity (Oh et al. 2014), gene expression (Hawrylycz et al. 2012), and, more recently, neural activity data in awake mice (de Vries et al. 2023). In the latter project, the teams recorded over 40,000 neurons across 14 brain regions. Importantly, all of the data was shared publicly, with appropriate meta-data, and an SDK was even provided for selecting, downloading, and interacting with the vast dataset (de Vries et al. 2023).

Such "big science" projects (Börner et al. 2021), which have a long history in fields like physics (e.g., the Conseil Européen pour la Recherche Nucléaire; CERN) or genetics (e.g., the Human Genome Project), are becoming more common in neuroscience and have surpassed the proof-of-concept phase. A key example is the International Brain Laboratory (Abbott et al. 2017; International Brain Laboratory et al. 2025), a distributed network of neuroscientists studying decision-making. Bridging experimental and theoretical neuroscience laboratories, the collaboration standardized a steering-wheel task for head-fixed mice developed to investigate decision-making based on visual perception and on history of reward. Hundreds of mice have been trained successfully on the task with concurrent neural recordings, which allowed researchers to discover that motor and reward information was distributed across a network of hundreds of brain subregions (International Brain Laboratory et al. 2023). National-scale initiatives, such as the BRAIN initiative (Jorgenson et al. 2015) or the Human Brain Project can shape entire fields and provide the critical safety nets for high-risk/high-reward research. Lessons from their successes and shortcomings should continue to be distributed transparently and publicly to guide the planning of future endeavors (Frégnac 2023).

As scientific questions and experiments become increasingly complex, the era of individual labs having the expertise to perform all aspects of a project in-house is nearing its end. The next era of neuroscience will likely be defined not only by an ecosystem of tools, but by an ecosystem of labs each contributing to the broader mission of understanding the fundamental principles underlying brain functions and how they are disrupted in disease. Each lab will play its part in this ecosystem, bridging levels through the production of innovative methods, novel datasets and analyses, or theoretical models. Further, we would argue that the human capital will be as critical as those scientific products. Taking notes from highly collaborative fields such as physics to help solve the "academic credit assignment" problem will be important to maintain the teams and necessary expertise to support long-term projects. More integrated teams of experimentalists and data scientists may also help minimize the temporal gap between technical and analytical developments, as well as ensure method development is focused on solving the most important problems.

4.3.4. Unifying Data Infrastructure and Analysis Pipelines and Scaling with the Size of the Datasets

As described earlier, unifying data formats and analytical pipelines would help mitigate the duplication of efforts and inefficient workflows. Software ecosystems that become too fragmented limit reproducibility, benchmarking, and, consequently, analytical validity. For example, a re-analysis of fMRI studies using the three most widely-used analysis packages found differences in the results of key processing steps (Luppi et al. 2024). In addition, after asking 70 independent teams to analyze the same fMRI dataset, a study found striking heterogeneity in analytical approaches used and large variations in the results of hypothesis testing (Botvinik-Nezer et al. 2020). Similar variability would be expected in electrophysiological recordings, particularly given the high variability across spike

sorting techniques (Buccino et al. 2020). At the very least, this lack of integration adds considerable noise to the process of synthesizing the literature and building theoretical models. The wide range of methods available needs to be streamlined into a more coherent whole and the pros and cons of each software package considered as readily as the assumptions of statistical tests used.

Efforts to standardize and integrate the array of data formats across the variety of recording methodologies are ongoing, with a number of solutions already available (Table 1). One example receiving much attention at present is the Neurodata Without Borders format (NWB; Rübél et al. 2016; Teeters et al. 2015), which integrates many of the qualities and structures from prior data types. NWB aims to encompass a variety of data modalities, including intracellular and extracellular electrophysiological data, optical imaging data, behavioral task data, and the corresponding metadata. Based on HDF5, a common open format, each dataset and its accompanying metadata are included in a directory-like container structure. However, HDF5, and by extension NWB, are not without their shortcomings, including a complex specification, potential corruption risks, lack of data validation, large sizes, and performance issues (Greenfield et al. 2015). It is

unlikely that a single solution will ever be perfect or meet all needs, as the difficult problem of balancing speed, robustness, and reproducibility with broad applicability and ease of use will always be a moving target. Importantly, large technology companies have demonstrated the feasibility of solutions to rapidly and efficiently store, manage, and serve data to and from millions of users, and these technological advancements are expected to slowly migrate into academic science. It is critical that we remain dedicated to these efforts, and to maintaining robust and user-friendly conversion tools across top solutions, otherwise we will slowly revert to an inefficient and fragmented ecosystem of solutions.

Standardizing data formats also facilitates data sharing via open data repositories, with a growing number of freely available repositories specializing in holding neural data now accessible (Table 3). As these repositories have garnered more attention from users and funders, the ease of searching, retrieving, and analyzing data from many labs around the world has improved considerably. This growing amount of shareable data can help catalyze scientific progress by increasing statistical power through the pooling of participants and more effectively validating results across labs. It can also lower the barrier for involvement in neuroscience

Table 3. Examples of open repositories for neuroscience data.

Repository	Description	URL	Discipline
BossDB	BossDB is a volumetric database for 3D and 4D neuroscience data that focuses on electron microscopy and X-ray tomography.	bossdb.org	Neuroscience
Brain Image Library	The Brain Image Library is a public resource enabling researchers to deposit, analyze, mine, share, and interact with large brain image datasets. The library primarily focuses on microscopy data.	brainimagelibrary.org	Neuroscience
CRCNS	A result of a NSF-NIH sponsored collaborative research program, CRCNS is an archive that shares experimental neuroscience data. See Teeters et al. (2008) for a technical report.	crcns.org	Neuroscience
DABI	DABI is a data repository for human and animal intracranial recordings and allows users to search, visualize, and analyze multimodal data. Data providers maintain full control of data sharing privileges and can manage their data with a user-friendly interface. See Duncan et al. (2023) for a technical report.	dabi.loni.usc.edu	Neuroscience
DANDI	The DANDI archive is a repository for sharing and publishing electrophysiology data. It aims to standardize data sharing and analysis, enabling researchers to easily access, contribute, and reuse complex datasets.	dandiarchive.org	Neuroscience
EBRAINS	EBRAINS provides access to a free and open database of neuroscience data, computational models, and software.	ebrains.eu	Neuroscience
NEMAR	NEMAR is a data, tools, and compute resource for EEG, MEG, and iEEG data. Data are formatted according to the BIDS standards and shared publicly by authors through openneuro.org .	nemar.org	Neuroscience
NeMO	The NeMO repository is the access point for neuroscience-related genomic data. It provides a web portal for searchable access to raw data, and a range of data visualization and analysis tools.	nemoarchive.org	Neuroscience
OpenNeuro	OpenNeuro is a platform for sharing MRI, MEG, EEG, intracranial EEG, and other neuroimaging datasets. It is organized according to the BIDS standard. See Markiewicz et al. (2021) for a technical report.	openneuro.org	Neuroscience
Dryad	Dryad is a data publishing platform committed to the open availability and routine reuse of research data. Dryad curates all data and metadata and publishes exclusively under a Creative Commons Public Domain License.	datadryad.org	Multiple
Figshare	Figshare is a freely available data publishing platform where researchers can share and receive credit for multiple types of scholarly output (e.g., datasets, figures, posters, code, papers, preprints, etc.). Support for larger datasets is also being incorporated.	figshare.com	Multiple
OSF	OSF is a free and open source project management tool that supports researchers throughout their entire project lifecycle in open science best practices. In addition to manuscript hosting, OSF provides tools to host code and data.	osf.io	Multiple
Vivli	Vivli has developed a global data sharing and analytics platform. With a focus on sharing participant-level data from completed clinical trials, the nonprofit aims to broadly serve the international research community.	vivli.org	Multiple
Zenodo	Designed by researchers from the Conseil Européen pour la Recherche Nucléaire (CERN) data center, Zenodo's original aim was for the long-term preservation of digital objects for high-energy physics. It has since expanded into data storage for a number of disciplines.	zenodo.org	Multiple

An updated and expanded version of this list is maintained at www.neurodatascience.org.

research by providing access to data to labs without the resources to undertake costly data collection (Insel et al. 2004; Kennedy 2006; Liu and Ascoli 2007). Such cloud-based solutions may be an opportunity to solve the problem of accessing very large data sets that are difficult to download, and to improve efficiency by bringing the analysis closer to the data (Bell et al. 2007). Therefore, investments in such distributed data archival systems are prudent and offer a high return on investment. However, their management should also be carefully planned: they should be controlled by scientists, tolerant of lapses in funding, and robust over multiple decades. Questions of data ownership, from the level of the team performing the work to that of the country funding it, will need to be addressed. Developing guidelines and incentives for openly sharing and re-analyzing federally-funded data will likely help mitigate these concerns and aid this transition (Sim et al. 2020).

4.3.5. The Open-Source Movement and Importance of Best Practices for Interpretable and Shareable Code

Analyses are increasingly being implemented in free and open-source programming languages such as *Python* or *R*, utilizing open-source analysis and visualization packages (e.g., *numpy*, *pandas*, *sci-kit*, *matplotlib*, *pytorch*, *ggplot*) and development environments (e.g., *jupyter*, *VS Code*, *RStudio*, *git*, etc), which facilitate code sharing and development. Nevertheless, sharing code is often not enough. The code shared must be useful, well-documented, and robust (Riquelme and Gjorgjieva 2021). Fortunately, the prevalence of open-source tools has resulted in a growing number of neuroscience-specific data analysis frameworks and toolboxes (Table 2). These tools include packages ranging from single-use functions to polished GUI-based analysis platforms that are being utilized by a growing number of labs.

Importantly, this open-source ethos is spreading to other aspects of neuroscience and data science as well. Hardware designs and specifications are increasingly becoming freely available, ranging from circuit designs for microcontrollers (Akam et al. 2022), custom lab equipment and 3D printed tools (Tek et al. 2008), camera systems (Saxena et al. 2018), to recording devices such as silicon probes (Yang et al. 2020). Efforts to support this ethos, such as Open Ephys (Siegle et al. 2017) or Open Behavior, can accelerate the rate at which individual labs design and implement complex studies. They also provide extra information on experimental approaches and are not locked behind proprietary doors, which can help data scientists without extensive neuroscience knowledge better understand how the data were collected.

The rapid integration of Large Language Models (LLMs) is increasingly altering the landscape of computational research by offering a powerful tool to accelerate the development of programming workflows, data preprocessing, and exploratory analysis. While these tools significantly lower technical barriers by automating boilerplate code generation and offering rapid debugging support, their usage bears important considerations. For instance, the stochastic nature of generative AI can produce “hallucinations”, syntactically correct but methodologically flawed code, that

may compromise research integrity if left unchecked. Consequently, LLM-assisted programming may ultimately demand a shift in the scientist’s role from a sole author to a critical auditor. Despite the convenience of these tools, a foundational understanding of the underlying algorithms and adherence to best practices remain necessary to ensure reproducibility.

4.3.6. Preserving the Need for Scientific Theory and Proper Experimental Design

In regard to theory, neuroscience is infamous as a field that is “data rich, but theory poor,” (Voytek 2022). This is well exemplified by the cycle described earlier, in which the development of techniques to collect more data is prioritized over the development of analyses and theories. Despite this, foundational findings and theories have emerged: one classic example is the discovery and subsequent experimental studies of “place cells” in the 1970s and 1980s, which predated the development of theoretical models to test and the experimental approaches to test them (e.g., projection-specific or cell-type-specific manipulations). While such a strong bottom-up focus is expected early in a field’s history, as neuroscience has matured it would now be counterproductive. In other words, at this time, not involving theory earlier in the scientific process will only delay scientific progress.

How can we facilitate the integration of scientific theory into increasingly more complex neuroscience experiments? One goal should be to lower the barrier for engaging theorists. In addition to improving access to the data (as discussed above), we should target hurdles limiting the integration of experimental data, such as a lack of tools to aggregate and compare metadata across experiments. To achieve this, collaborations between theorists and experimentalists will become increasingly important, as theorists may lack the necessary domain knowledge to properly interpret such data. In addition to theorists, consulting statisticians and data scientists early in the experimental design process can help mitigate future problems that may only be revealed when formalizing data analysis plans, such as underpowered study designs or lack of adequate controls. As noted early on by Fisher (1938), “To consult the statistician after an experiment is finished is often merely to ask [them] to conduct a post mortem examination. [They] can perhaps say what the experiment died of.”

What have been the strengths and weaknesses of the field of neuroscience in incorporating proper experimental design practices? In terms of strengths, we should note that neuroscience has consistently valued, and invested in, the technical development needed to improve the quality and size of the dataset. Neuroscientists (and other life scientists) have also made concerted efforts to make data and methods more reproducible and easier to access (Schottdorf et al. 2024). In terms of weaknesses, neuroscience has faced a range of well-documented statistical concerns in recent years that can compromise the reliability of published findings. While these concerns extend to most empirical sciences, neuroscience is thought to be particularly vulnerable because of the unusually complex, high-dimensional, noisy, and hierarchical nature of its data.

For clarity, here we divide these concerns into three main categories. The first category is related to misinterpretations or misuses of p -values and of null hypothesis significance testing in general. Examples include p -hacking, optional stopping, flexible data inclusion, and dichotomization of results (e.g., $p = 0.049$ as significant, $p = 0.051$ as non-significant), as well as common misinterpretations of what a p -value represents (Gelman and Stern 2006; Simmons et al. 2011; Head et al. 2015; Calin-Jageman and Cumming 2019). The second category is related to the common use of small sample sizes in the field, which can be limited for technical or logistical reasons. Specifically, many neuroscience studies were shown to be too small to detect realistic effect sizes, which can increase the rate of false negatives, exaggerate effect sizes, and decrease the probability that significant findings reflect real effects (Ioannidis 2005; Button et al. 2013). Much attention has been called to these two categories of concerns in recent years, which should improve with the resulting increased emphasis on quantitative training for neuroscientists. The last category includes problems specific to the scale and statistical structure of neural data. For instance, the large number of units involved (e.g., thousands of voxels, hundreds of neurons) is a challenge for balancing the risk of false positives and false negatives (Maris and Oostenveld 2007; Eklund et al. 2016). In addition, voxels, neurons, and trials are often treated as independent observations despite known nested and correlated structures (Aarts et al. 2014; Yu et al. 2022). Whether neuroscientists are aware of these issues or not, most labs do not have easy access to advanced statistical methods, which can result in oversimplified or incorrect analyses. Therefore, improved training of neuroscientists may not suffice to address this category of concerns—closer collaborations between experimentalists and data scientists will be needed as well.

Finally, we should strive to view the investigative team as the whole field, instead of any specific lab. Some questions may be more important than others in a given field and the order in which they are addressed should mirror that. Some questions may be answered with existing datasets, whereas others may require new data, and the decision process should include a consideration of the associated costs and benefits. While not everyone will agree on the details, a consensus can still be reached on a long-term experimental plan that, importantly, incorporates scientific theory as well as statistics and experimental design from its inception. This would require a considerable change in culture for neuroscientists accustomed to the existing model, in which each lab has their own independent research program that directly competes with other labs investigating the same topic. As such, the importance (and benefits) of scientific theory and sound experimental design would need to be consistently communicated to trainees and senior investigators.

4.3.7. Increased Need for Interdisciplinary Training Programs and Initiatives

A new generation of scientists at the intersection of neuroscience and data science (i.e., neurodatascience) will be needed to tackle the projected increase in scale and complexity of neuroscience data in the future. The key challenge

is that the curricula for neuroscience and data science students tend to be vastly different such that they face asymmetrical deficits in core competencies, the fundamental skills required to execute scientific work, and core concepts, the foundational knowledge that defines a field. For instance, a 2017 survey of leaders in computational neuroscience identified that students in the life sciences often lack quantitative training, while those from quantitative backgrounds frequently lack biological foundations and experience with biological data (Goldman and Fee 2017). The researchers found that these disparities are exacerbated by a scarcity of training resources and cultural barriers regarding the adoption of computational techniques.

Despite their importance, quantitative core competencies are frequently absent from modern neuroscience training (Ramirez 2020). A primary obstacle to implementation is the breadth of requisite skills, which can include calculus, linear algebra, differential equations, statistics, probability, and proficiency in high-level programming languages (Python, R, MATLAB, etc.). For students and researchers already managing heavy workloads, acquiring such a comprehensive range of data science proficiencies independently can prove prohibitive. Conversely, students entering from quantitative backgrounds often lack the foundational neuroscience concepts necessary to contextualize their technical skills. Unlike deterministic computational systems, biological environments are characterized by stochasticity, high variability, and complex hierarchical organization. Without adequate grounding in these principles, students risk applying models that are mathematically rigorous yet physiologically implausible, or misinterpreting data due to a lack of context regarding experimental limitations and assumptions. Data science students and researchers are also burdened with heavy workloads and acquiring such domain knowledge independently can prove equally prohibitive.

The long-term solution to this problem is to develop interdisciplinary training programs at the undergraduate and graduate level, where neuroscience trainees can receive more advanced quantitative training and data science trainees can receive more neuroscience training. Such programs can build on recent efforts integrating quantitative training into neuroscience curricula (Grisham et al. 2016), integrating neuroscience core concepts into non-domain scientists curricula (Chen et al. 2023) as well as encouraging broader coursework (Grisham et al. 2016; Cooper et al. 2025). Other training initiatives, such as workshops or short courses, will also play a key role. For example, Neuromatch Academy successfully delivered a fully online, 3-week Computational Neuroscience Summer School to 1,757 students, supported by 191 teaching assistants. Crucially, the majority of their materials remain freely available, including slides, lectures, and code (van Viegen et al. 2021). Ultimately, education extends beyond formal instruction; accessible code libraries, comprehensive documentation, and tutorials play a significant role in lowering barriers for trainees. Moving forward, the field must prioritize the dissemination of best practices for interdisciplinary training to facilitate their broader adoption and assess their effectiveness.

Acknowledgments

We dedicate this work to our teams and colleagues who share our desire to leave the comfort of one's field to learn about another. KWC would like to thank Dr. James McGaugh for insightful discussions on past practices in neuroscience. We maintain a repository of updated tables and related materials at www.neurodatascience.org.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was partially supported by awards from National Institutes of Health (NIH) (R01-MH115697, R01-DC017687), and National Science Foundation (NSF) (CAREER IOS-1150292, DGE-1839285, BCS-1439267, and NCS-FR-2319618).

ORCID

Keiland W. Cooper  <http://orcid.org/0000-0002-0358-9645>
 Babak Shahbaba  <http://orcid.org/0000-0002-8102-1609>
 Norbert J. Fortin  <http://orcid.org/0000-0002-6793-6984>

References

- Aarts E et al. 2014. A solution to dependency: using multilevel analysis to accommodate nested data. *Nat Neurosci*. 17(4):491–496. <https://doi.org/10.1038/nn.3648>
- Abbott LF et al. 2017. An international laboratory for systems and computational neuroscience. *Neuron*. 96(6):1213–1218. <https://doi.org/10.1016/j.neuron.2017.12.013>
- Abeles M, Goldstein MH. 1977. Multispike train analysis. *Proc IEEE*. 65(5):762–773. <https://doi.org/10.1109/PROC.1977.10559>
- Aellen FM et al. 2021. Convolutional neural networks for decoding electroencephalography responses and visualizing trial by trial changes in discriminant features. *J Neurosci Methods*. 364:109367. <https://doi.org/10.1016/j.jneumeth.2021.109367>
- Aertsen AM et al. 1989. Dynamics of neuronal firing correlation: modulation of 'effective connectivity'. *J Neurophysiol*. 61(5):900–917. <https://doi.org/10.1152/jn.1989.61.5.900>
- Aharoni D et al. 2019. All the light that we can see: a new era in miniaturized microscopy. *Nat Methods*. 16(1):11–13. <https://doi.org/10.1038/s41592-018-0266-x>
- Ahn J-R, Lee H-W, Lee I. 2019. Rhythmic pruning of perceptual noise for object representation in the hippocampus and perirhinal cortex in rats. *Cell Rep*. 26(9):2362–2376.e4. <https://doi.org/10.1016/j.celrep.2019.02.010>
- Ahrens MB et al. 2012. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*. 485(7399):471–477. <https://doi.org/10.1038/nature11057>
- Ahumada R et al. 2020. The 16th data release of the sloan digital sky surveys: first release from the APOGEE-2 southern survey and full release of eBOSS spectra. *ApJS*. 249(1):3. <https://doi.org/10.3847/1538-4365/ab929e>
- Akam T et al. 2022. Open-source, Python-based, hardware and software for controlling behavioural neuroscience experiments. *Elife*. 11: e67846. <https://doi.org/10.7554/eLife.67846>
- Al-Awami AK et al. 2014. NeuroLines: a subway map metaphor for visualizing nanoscale neuronal connectivity. *IEEE Trans Vis Comput Graph*. 20(12):2369–2378. <https://doi.org/10.1109/TVCG.2014.2346312>
- Allen EA, Erhardt EB, Calhoun VD. 2012. Data visualization in the neurosciences: overcoming the curse of dimensionality. *Neuron*. 74(4):603–608. <https://doi.org/10.1016/j.neuron.2012.05.001>
- Amit DJ. 1992. Modeling brain function. [accessed 2025 Dec 10]. <https://ui.adsabs.harvard.edu/abs/1992mbf.book>
- Arkhipov A et al. 2025. Integrating multimodal data to understand cortical circuit architecture and function. *Nat Neurosci*. 28(4):717–730. <https://doi.org/10.1038/s41593-025-01904-7>
- Barabási DL et al. 2023. Neuroscience needs network science. *J Neurosci*. 43(34):5989–5995. <https://doi.org/10.1523/JNEUROSCI.1014-23.2023>
- Barbieri R et al. 2004. Dynamic analyses of information encoding in neural ensembles. *Neural Comput*. 16(2):277–307. <https://doi.org/10.1162/089976604322742038>
- Bates AS et al. 2020. The natverse, a versatile toolbox for combining and analysing neuroanatomical data. *Elife*. 9:e53350. <https://doi.org/10.7554/eLife.53350>
- Baxter AL et al. 2022. Collaborative experience between scientific software projects using Agile Scrum development. *Softw Pract Exp*. 52(10):2077–2096. <https://doi.org/10.1002/spe.3120>
- Bell G, Gray J, Szalay A. 2007. Petascale computational systems. *arXiv*. <https://doi.org/10.48550/arXiv.cs/0701165>
- Berger H. 1929. Über das Elektroencephalogramm des Menschen. *Archiv Psychiatr*. 87(1):527–570. <https://doi.org/10.1007/BF01797193>
- Bigdely-Shamlo N, Mullen T, Kothe C, Su K-M, Robbins KA. 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front Neuroinform*. 9:16. <https://doi.org/10.3389/fninf.2015.00016>
- Borisjuk GN et al. 1985. A new statistical method for identifying interconnections between neuronal network elements. *Biol Cybern*. 52(5):301–306. <https://doi.org/10.1007/BF00355752>
- Börner K, Silva FN, Milojević S. 2021. Visualizing big science projects. *Nat Rev Phys*. 3(11):753–761. <https://doi.org/10.1038/s42254-021-00374-7>
- Bosch J et al. 2019. An overview of the LSST image processing pipelines. <https://doi.org/10.48550/arXiv.1812.03248>
- Botvinik-Nezer R et al. 2020. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*. 582(7810):84–88. <https://doi.org/10.1038/s41586-020-2314-9>
- Brinkmann BH, Bower MR, Stengel KA, Worrell GA, Stead M. 2009. Large-scale electrophysiology: acquisition, compression, encryption, and storage of Big Data. *J Neurosci Methods*. 180(1):185–192. <https://doi.org/10.1016/j.jneumeth.2009.03.022>
- Brillinger DR. 1988. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cybern*. 59(3):189–200. <https://doi.org/10.1007/BF00318010>
- Brillouin L. 1956. Science and information theory. Dover Publications. <https://doi.org/10.1063/1.3059856>
- Brown EN et al. 1998. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci*. 18(18):7411–7425. <https://doi.org/10.1523/JNEUROSCI.18-18-07411.1998>
- Brown EN, Kass RE, Mitra PP. 2004. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat Neurosci*. 7(5):456–461. <https://doi.org/10.1038/nn1228>
- Brunner RJ et al. 2002. Massive datasets in astronomy. In: *Handbook of massive data sets*. Kluwer Academic Publishers. p 931–979. https://doi.org/10.1007/978-1-4615-0005-6_27
- Bruzzzone M et al. 2021. Whole brain functional recordings at cellular resolution in zebrafish larvae with 3D scanning multiphoton microscopy. *Sci Rep*. 11(1):11048. <https://doi.org/10.1038/s41598-021-90335-y>
- Buccino AP et al. 2020. SpikeInterface, a unified framework for spike sorting. *Elife*. 9:e61834. <https://doi.org/10.7554/eLife.61834>
- Burns R et al. 2013. The open connectome project data cluster: scalable analysis and vision for high-throughput neuroscience. In: *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*. Association for Computing Machinery. p 10–11. <https://doi.org/10.1145/2484838.2484870>
- Button KS et al. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 14(5):365–376. <https://doi.org/10.1038/nrn3475>
- Buxner SR, Fitzgerald MT, Freed, RM. 2021. Chapter 8. Amateur astronomy: engaging the public in astronomy through exploration, outreach, and research. In: Kaminski AP, editors. *Space science and public engagement*. Elsevier. p 143–168. <https://doi.org/10.1016/B978-0-12-817390-9.00002-6>

- Buzsáki G, Anastassiou CA, Koch C. 2012. The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. *Nat Rev Neurosci*. 13(6):407–420. <https://doi.org/10.1038/nrn3241>
- Buzsáki G, Draguhn A, Buzsáki G. 2004. Neuronal oscillations in cortical networks. *Science*. 304(5679):1926–1929. <https://doi.org/10.1126/science.1099745>
- Bzdok D, Yeo BTT. 2016. The future of data analysis in the neurosciences. *arXiv*. <https://doi.org/10.48550/arXiv.1608.03465>
- Calin-Jageman RJ, Cumming G. 2019. Estimation for better inference in neuroscience. *eNeuro*. 6(4):ENEURO.0205-19.2019. <https://doi.org/10.1523/ENEURO.0205-19.2019>
- Campagnola L et al. 2015. VisPy: harnessing the GPU for fast, high-level visualization. In: *Python in Science Conference*, Austin, Texas. p 91–96. <https://doi.org/10.25080/Majora-7b98e3ed-00e>
- Chang C-C, Lin C-J. 2011. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2(3):1–27. <https://doi.org/10.1145/1961189.1961199>
- Chapin JK, Nicolelis MAL. 1999. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *J Neurosci Methods*. 94(1):121–140. [https://doi.org/10.1016/S0165-0270\(99\)00130-2](https://doi.org/10.1016/S0165-0270(99)00130-2)
- Chaudhuri R et al. 2019. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat Neurosci*. 22(9):1512–1520. <https://doi.org/10.1038/s41593-019-0460-x>
- Chen A et al. 2023. Community-derived core concepts for neuroscience higher education. *CBE Life Sci Educ*. 22(2):ar18. <https://doi.org/10.1187/cbe.22-02-0018>
- Chen S et al. 2019. The multivariate Hawkes process in high dimensions: beyond mutual excitation. *arXiv*. <https://doi.org/10.48550/arXiv.1707.04928>
- Chen T-W et al. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*. 499(7458):295–300. <https://doi.org/10.1038/nature12354>
- Christensen A, Massara GP. 2018. NetworkToolbox: methods and measures for brain, cognitive, and psychometric network analysis. <https://journal.r-project.org/articles/RJ-2018-065/>
- Chu JP, Kemere CT. 2021. GhostPy: an efficient signal processing and spectral analysis toolbox for large data. *eNeuro*. 8(6):ENEURO.0202-21.2021. <https://doi.org/10.1523/ENEURO.0202-21.2021>
- Chung S, Abbott LF. 2021. Neural population geometry: an approach for understanding biological and artificial neural networks. *Curr Opin Neurobiol*. 70:137–144. <https://doi.org/10.1016/j.conb.2021.10.010>
- Churchland MM et al. 2012. Neural population dynamics during reaching. *Nature*. 487(7405):51–56. <https://doi.org/10.1038/nature11129>
- Claudi F et al. 2021. Visualizing anatomically registered data with brainrender. *Elife*. 10:e65751. <https://doi.org/10.7554/eLife.65751>
- Cole S, Donoghue T, Gao R, Voytek B. 2019. NeuroDSP: a package for neural digital signal processing. *JOSS*. 4(36):1272. <https://doi.org/10.21105/joss.01272>
- Collinger JL et al. 2013. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet*. 381(9866):557–564. [https://doi.org/10.1016/S0140-6736\(12\)61816-9](https://doi.org/10.1016/S0140-6736(12)61816-9)
- Combrisson E et al. 2019. Visbrain: a multi-purpose GPU-accelerated open-source suite for multimodal brain data visualization. *Front Neuroinform*. 13:14. <https://doi.org/10.3389/fninf.2019.00014>
- Cooley JW, Tukey JW. 1965. An algorithm for the machine calculation of complex Fourier series. *Math Comp*. 19(90):297–301. <https://doi.org/10.1090/S0025-5718-1965-0178586-1>
- Cooper KW. 2018. Can the machine understand: an evidence based approach to the Chinese room. *IJUR*. 4(1):82–85. <https://doi.org/10.14434/ijur.v4i1.24547>
- Cooper KW et al. 2025. Challenges that novices face in applying core concepts to neuroscience contexts. *Front Educ*. 10:1488892. <https://doi.org/10.3389/educ.2025.1488892>
- Cortes C, Vapnik V. 1995. Support-vector networks. *Mach Learn*. 20(3):273–297. <https://doi.org/10.1007/BF00994018>
- Couey JJ et al. 2013. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci*. 16(3):318–324. <https://doi.org/10.1038/nn.3310>
- Cox RW et al. 2004. A (sort of) new image data format standard: niFTI-1. Presented at the 10th Annual Meeting of the Organization for Human Brain Mapping, Budapest, Hungary.
- Creutzfeldt O, Jung R. 1961. Neuronal discharge in the cat's motor cortex during sleep and arousal. In: *Ciba Foundation Symposium—The Nature of Sleep*. John Wiley & Sons, Ltd. p 131–170. <https://doi.org/10.1002/9780470719220.ch7>
- Csicsvari J et al. 2003. Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J Neurophysiol*. 90(2):1314–1323. <https://doi.org/10.1152/jn.00116.2003>
- Cunningham JP, Yu BM. 2014. Dimensionality reduction for large-scale neural recordings. *Nat Neurosci*. 17(11):1500–1509. <https://doi.org/10.1038/nn.3776>
- Dabagia M, Kording KP, Dyer EL. 2023. Comparing high-dimensional neural recordings by aligning their low-dimensional latent representations. *Nat. Biomed. Eng* 7:337–343. <https://doi.org/10.1038/s41551-022-00962-7>
- Daryakenari NA et al. 2024. AI-Aristotle: a physics-informed framework for systems biology gray-box identification. *PLOS Comput Biol*. 20(3):e1011916. <https://doi.org/10.1371/journal.pcbi.1011916>
- Davidson TJ, Kloosterman F, Wilson MA. 2009. Hippocampal replay of extended experience. *Neuron*. 63(4):497–507. <https://doi.org/10.1016/j.neuron.2009.07.027>
- Diggelmann R et al. 2018. Automatic spike sorting for high-density microelectrode arrays. *J Neurophysiol*. 120(6):3155–3171. <https://doi.org/10.1152/jn.00803.2017>
- Ding DY et al. 2022. Cooperative learning for multiview analysis. *Proc Natl Acad Sci U S A*. 119(38):e2202113119. <https://doi.org/10.1073/pnas.2202113119>
- Dorkenwald S et al. 2024. Neuronal wiring diagram of an adult brain. *Nature*. 634(8032):124–138. <https://doi.org/10.1038/s41586-024-07558-y>
- Dumermuth G, Flüher H. 1967. Some modern aspects in numerical spectrum analysis of multichannel electroencephalographic data. *Med Biol Eng*. 5(4):319–331. <https://doi.org/10.1007/BF02479095>
- Duncan D et al. 2023. Data archive for the BRAIN initiative (DABI). *Sci Data*. 10(1):83. <https://doi.org/10.1038/s41597-023-01972-z>
- Eklund A, Nichols TE, Knutsson H. 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci U S A*. 113(28):7900–7905. <https://doi.org/10.1073/pnas.1602413113>
- El-Yaagoubi AB et al. 2025. Methods for brain connectivity analysis with applications to rat local field potential recordings. *Entropy*. 27(4):328. <https://doi.org/10.3390/e27040328>
- Everts MH et al. 2009. Depth-dependent halos: illustrative rendering of dense line data. *IEEE Trans Vis Comput Graph*. 15(6):1299–1306. <https://doi.org/10.1109/TVCG.2009.138>
- Fan J, Han F, Liu H. 2014. Challenges of big data analysis. *Natl Sci Rev*. 1(2):293–314. <https://doi.org/10.1093/nsr/nwt032>
- Ferezou I, Bolea S, Petersen CCH. 2006. Visualizing the cortical representation of whisker touch: voltage-sensitive dye imaging in freely moving mice. *Neuron*. 50(4):617–629. <https://doi.org/10.1016/j.neuron.2006.03.043>
- Ferguson JE, Boldt C, Redish AD. 2009. Creating low-impedance tetrodes by electroplating with additives. *Sens Actuators A Phys*. 156(2):388–393. <https://doi.org/10.1016/j.sna.2009.10.001>
- Frégac Y. 2023. Flagship afterthoughts: could the human brain project (HBP) have done better? *eNeuro*. 10(11):ENEURO.0428-23.2023. <https://doi.org/10.1523/ENEURO.0428-23.2023>
- Friston K. 2010. The free-energy principle: a unified brain theory? *Nat Rev Neurosci*. 11(2):127–138. <https://doi.org/10.1038/nrn2787>
- Fuchs R, Hauser H. 2009. Visualization of multi-variate scientific data. *Comput Graphics Forum*. 28(6):1670–1690. <https://doi.org/10.1111/j.1467-8659.2009.01429.x>
- Fuhs MC, Touretzky DS. 2006. A spin glass model of path integration in rat medial entorhinal cortex. *J Neurosci*. 26(16):4266–4276. <https://doi.org/10.1523/JNEUROSCI.4353-05.2006>
- Fukushima K. 1980. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*. 36(4):193–202. <https://doi.org/10.1007/BF00344251>

- Gallego JA et al. 2018. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat Commun.* 9(1):4233. <https://doi.org/10.1038/s41467-018-06560-z>
- Ganguli S et al. 2008. One-dimensional dynamics of attention and decision making in LIP. *Neuron.* 58(1):15–25. <https://doi.org/10.1016/j.neuron.2008.01.038>
- Ganguli S, Sompolinsky H. 2012. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu Rev Neurosci.* 35(1):485–508. <https://doi.org/10.1146/annurev-neuro-062111-150410>
- Garcia S et al. 2014. Neo: an object model for handling electrophysiology data in multiple formats. *Front Neuroinform.* 8:10. <https://doi.org/10.3389/fninf.2014.00010>
- Gardner RJ et al. 2022. Toroidal topology of population activity in grid cells. *Nature.* 602(7895):123–128. <https://doi.org/10.1038/s41586-021-04268-7>
- Gelman A, Stern H. 2006. The difference between ‘significant’ and ‘not significant’ is not itself statistically significant. *Am Stat.* 60(4):328–331. <https://doi.org/10.1198/000313006X152649>
- Gerstein GL, Clark WA. 1964. Simultaneous studies of firing patterns in several neurons. *Science.* 143(3612):1325–1327. <https://doi.org/10.1126/science.143.3612.1325>
- Gerstein GL, Perkel DH. 1969. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science.* 164(3881):828–830. <https://doi.org/10.1126/science.164.3881.828>
- Gerstein GL, Kiang NY-S. 1960. An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophys J.* 1(1):15–28. [https://doi.org/10.1016/S0006-3495\(60\)86872-5](https://doi.org/10.1016/S0006-3495(60)86872-5)
- Gillon CJ et al. 2025. Open data in neurophysiology: advancements, solutions & challenges. *eNeuro.* 12(11):ENEURO.0486-24.2025. <https://doi.org/10.1523/ENEURO.0486-24.2025>
- Glaser E. 1971. Separation of neuronal activity by waveform analysis. *Adv Biomed Eng.* 1:77–136.
- Glaser JI et al. 2020. Machine learning for neural decoding. *eNeuro* 31 7 (4) ENEURO.0506-19.2020; <https://doi.org/10.1523/ENEURO.0506-19.2020>
- Glover GH. 2011. Overview of functional magnetic resonance imaging. *Neurosurg Clin N Am.* 22(2):133–139, vii. <https://doi.org/10.1016/j.nec.2010.11.001>
- Goldman MS, Fee MS. 2017. Computational training for the next generation of neuroscientists. *Curr Opin Neurobiol.* 46:25–30. <https://doi.org/10.1016/j.conb.2017.06.007>
- Gorgolewski KJ et al. 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data.* 3(1):160044. <https://doi.org/10.1038/sdata.2016.44>
- Grass AM, Gibbs FA. 1938. A Fourier transform of the electroencephalogram. *J Neurophysiol.* 1(6):521–526. <https://doi.org/10.1152/jn.1938.1.6.521>
- Griffanti L et al. 2017. Hand classification of fMRI ICA noise components. *Neuroimage.* 154:188–205. <https://doi.org/10.1016/j.neuroimage.2016.12.036>
- Greenfield P, Droettboom M, Bray E. 2015. ASDF: a new data format for astronomy. *Astron Comput.* 12:240–251. <https://doi.org/10.1016/j.ascom.2015.06.004>
- Griffith J, Horn G. 1963. Functional coupling between cells in the visual cortex of the unrestrained cat. *Nature.* 199(4896):876–895. <https://doi.org/10.1038/199876a0>
- Griffith JS, Horn G, Sassoon GT. 1965. A device for transferring neurophysiological pulse data automatically from magnetic to computer input tape. *Med Electron Biol Engng.* 3(3):239–248. <https://doi.org/10.1007/BF02474305>
- Grisham W et al. 2016. Proposed training to meet challenges of large-scale data in neuroscience. *Front Neuroinform.* 10:28. <https://doi.org/10.3389/fninf.2016.00028>
- Groppe DM et al. 2017. iELVis: an open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. *J Neurosci Methods.* 281:40–48. <https://doi.org/10.1016/j.jneumeth.2017.01.022>
- Grün S et al. 2022. Emerging principles of spacetime in brains: meeting report on spatial neurodynamics. *Neuron.* 110(12):1894–1898. <https://doi.org/10.1016/j.neuron.2022.05.018>
- Guardamagna M et al. 2022. The Hybrid Drive: a chronic implant device combining tetrode arrays with silicon probes for layer-resolved ensemble electrophysiology in freely moving mice. *J Neural Eng.* 19(3):036030. <https://doi.org/10.1088/1741-2552/ac6771>
- Hafting T et al. 2005. Microstructure of a spatial map in the entorhinal cortex. *Microstruct Spatial Map Entorhinal Cortex.* 436(7052):801–806. <https://doi.org/10.1038/nature03721>
- Hampel H et al. 2019. The Alzheimer precision medicine initiative. *J Alzheimers Dis.* 68(1):1–24. <https://doi.org/10.3233/JAD-181121>
- Han J et al. 2024. Impact of impedance levels on recording quality in flexible neural probes. *Sensors.* 24(7):2300. <https://doi.org/10.3390/s24072300>
- Harris KD. 2020. Nonsense correlations in neuroscience. *bioRxiv.* 20201129.402719. <https://doi.org/10.1101/2020.11.29.402719>
- Harrison MT, Geman S. 2009. A rate and history-preserving resampling algorithm for neural spike trains. *Neural Comput.* 21(5):1244–1258. <https://doi.org/10.1162/neco.2008.03-08-730>
- Hassabis D et al. 2017. Neuroscience-inspired artificial intelligence. *Neuron.* 95(2):245–258. <https://doi.org/10.1016/j.neuron.2017.06.011>
- Hawkes AG. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika.* 58(1):83–90. <https://doi.org/10.1093/biomet/58.1.83>
- Hawrylycz MJ et al. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature.* 489(7416):391–399. <https://doi.org/10.1038/nature11405>
- Head ML et al. 2015. The extent and consequences of p-hacking in science. *PLoS Biol.* 13(3):e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- Heeger DJ, Ress D. 2002. What does fMRI tell us about neuronal activity? *Nat Rev Neurosci.* 3(2):142–151. <https://doi.org/10.1038/nrn730>
- Helmchen F et al. 2001. A miniature head-mounted two-photon microscope: high-resolution brain imaging in freely moving animals. *Neuron.* 31(6):903–912. [https://doi.org/10.1016/S0896-6273\(01\)00421-4](https://doi.org/10.1016/S0896-6273(01)00421-4)
- Henze DA et al. 2000. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol.* 84(1):390–400. <https://doi.org/10.1152/jn.2000.84.1.390>
- Herculano-Houzel S, Mota B, Lent R. 2006. Cellular scaling rules for rodent brains. *Proc Natl Acad Sci U S A.* 103(32):12138–12143. <https://doi.org/10.1073/pnas.0604911103>
- Hilgen G et al. 2017. Unsupervised spike sorting for large-scale, high-density multielectrode arrays. *Cell Rep.* 18(10):2521–2532. <https://doi.org/10.1016/j.celrep.2017.02.038>
- Holbrook A et al. 2017. A Bayesian supervised dual-dimensionality reduction model for simultaneous decoding of LFP and spike train signals. *Stat.* 6(1):53–67. <https://doi.org/10.1002/sta4.137>
- Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A.* 79(8):2554–2558. <https://doi.org/10.1073/pnas.79.8.2554>
- Hubel DH. 1959. Single unit activity in striate cortex of unrestrained cats. *J Physiol.* 147(2)(2):226–238. <https://doi.org/10.1113/jphysiol.1959.sp006238>
- Insel TR et al. 2004. Limits to growth: why neuroscience needs large-scale science. *Nat Neurosci.* 7(5):426–427. <https://doi.org/10.1038/nn0504-426>
- International Brain Laboratory et al. 2023. A brain-wide map of neural activity during complex behaviour. Preprint Neuroscience. <https://doi.org/10.1101/2023.07.04.547681>
- International Brain Laboratory et al. 2025. 20 lessons in team science: learning from the experience of the International Brain Laboratory. *PsyArXiv.* https://doi.org/10.31234/osf.io/rfzch_v1
- Ioannidis JPA. 2005. Why most published research findings are false. *PLoS Med.* 2(8):e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Jewell S, Witten D. 2017. Exact spike train inference via L0 optimization. <https://doi.org/10.48550/arXiv.1703.08644>
- Jaynes ET. 1957. Information theory and statistical mechanics. *Phys Rev.* 106(4):620–630. <https://doi.org/10.1103/PhysRev.106.620>
- Jones L et al. 2020. Evaluating the rubin observatory legacy survey of space and time (LSST) observing strategy—with an eye on small body science. AAS Division of Planetary Science meeting #52, id. 110.02. *Bulletin of the American Astronomical Society, Vol. 52, No. 6.*

- Jorgenson LA et al. 2015. The BRAIN initiative: developing technology to catalyse neuroscience discovery. *Phil Trans R Soc B*. 370(1668): 20140164. <https://doi.org/10.1098/rstb.2014.0164>
- Juczewski K et al. 2020. Stress and behavioral correlates in the head-fixed method: stress measurements, habituation dynamics, locomotion, and motor-skill learning in mice. *Sci Rep*. 10(1):12245. <https://doi.org/10.1038/s41598-020-69132-6>
- Kadir SN, Goodman DFM, Harris KD. 2014. Klustakwik. <http://klustakwik.github.io/klustakwik/>
- Kass RE et al. 2018. Computational neuroscience: mathematical and statistical perspectives. *Annu Rev Stat Appl*. 5(1):183–214. <https://doi.org/10.1146/annurev-statistics-041715-033733>
- Kass RE, Eden U, Brown EN. 2014. Analysis of neural data. Springer Science + Business Media. <https://doi.org/10.1007/978-1-4614-9602-1>
- Kass RE, Kelly RC, Loh W-L. 2011. Assessment of synchrony in multiple neural spike trains using loglinear point process models. *Ann Appl Stat*. 5(2B):1262–1292. <https://doi.org/10.1214/10-AOAS429>
- Kell AJE et al. 2018. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*. 98(3):630–644.e16. <https://doi.org/10.1016/j.neuron.2018.03.044>
- Kelly RC et al. 2007. Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *J Neurosci*. 27(2): 261–264. <https://doi.org/10.1523/JNEUROSCI.4906-06.2007>
- Godfrey K. 2014. Orca file format for e-phys and o-phys neurodata. In: Neurodata Without Borders Hackathon. http://crcns.org/files/data/nwb/h1/NWBh1_09_Keith_Godfrey.pdf
- Kemp B, Olivan J. 2003. European data format “plus” (EDF+), an EDF alike standard format for the exchange of physiological data. *Clin Neurophysiol*. 114(9):1755–1761. [https://doi.org/10.1016/s1388-2457\(03\)00123-8](https://doi.org/10.1016/s1388-2457(03)00123-8)
- Kemp B, Värri A, Rosa AC, Nielsen KD, Gade J. 1992. A simple format for exchange of digitized polygraphic recordings. *Electroencephalogr Clin Neurophysiol*. 82(5):391–393. [https://doi.org/10.1016/0013-4694\(92\)90009-7](https://doi.org/10.1016/0013-4694(92)90009-7)
- Kennedy D. 2006. Where’s the beef? Missing data in the information age. *Neuroinformatics*. 4(4):271–273. <https://doi.org/10.1385/NI:4:4:271>
- Kim JS et al. 2014. Space–time wiring specificity supports direction selectivity in the retina. *Nature*. 509(7500):331–336. <https://doi.org/10.1038/nature13240>
- Kim TH et al. 2016. Long-term optical access to an estimated one million neurons in the live mouse cortex. *Cell Rep*. 17(12):3385–3394. <https://doi.org/10.1016/j.celrep.2016.12.004>
- Kobak D et al. 2016. Demixed principal component analysis of neural population data. *Elife*. 5:e10989. <https://doi.org/10.7554/eLife.10989>
- Koch C, Reid RC. 2012. Observatories of the mind. *Nature*. 483(7390): 397–398. <https://doi.org/10.1038/483397a>
- Kraus B et al. 2023. Insights from personalized models of brain and behavior for identifying biomarkers in psychiatry. *Neurosci Biobehav Rev*. 152:105259. <https://doi.org/10.1016/j.neubiorev.2023.105259>
- Krizhevsky A. 2014. One weird trick for parallelizing convolutional neural networks. *arXiv*. <https://doi.org/10.48550/arXiv.1404.5997>
- Krüger, J. (1983) Simultaneous individual recordings from many cerebral neurons: techniques and results. In: Reviews of physiology, biochemistry and pharmacology. Vol. 98. Springer Berlin Heidelberg. p 177–233. <https://doi.org/10.1007/BFb0033868>
- Lambert RC et al. 2018. Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *J Neurosci Methods*. 297:9–21. <https://doi.org/10.1016/j.jneumeth.2017.12.026>
- LaPlante R. 2013. bctpy: brain connectivity toolbox for Python. <https://pypi.org/project/bctpy/>
- Laughlin SB, de Ruyter van Steveninck RR, Anderson JC. 1998. The metabolic cost of neural information. *Nat Neurosci*. 1(1):36–41. <https://doi.org/10.1038/236>
- Lecun Y et al. 1998. Gradient-based learning applied to document recognition. *Proc IEEE*. 86(11):2278–2324. <https://doi.org/10.1109/5.726791>
- Lee AK, Brecht M. 2018. Elucidating neuronal mechanisms using intracellular recordings during behavior. *Trends Neurosci*. 41(6):385–403. <https://doi.org/10.1016/j.tins.2018.03.014>
- Lee DD, Seung HS. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*. 401(6755):788–791. <https://doi.org/10.1038/44565>
- Lee KH et al. 2025. Spyglass: a framework for reproducible and shareable neuroscience research. *Elife*. 14:RP108089. <https://doi.org/10.7554/eLife.108089.1>
- Lein ES et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*. 445(7124):168–176. <https://doi.org/10.1038/nature05453>
- Lewicki MS. 1998. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network*. 9(4):R53–R78. https://doi.org/10.1088/0954-898X_9_4_001
- Li Y, Dalphin N, Hyland BI. 2013. Association with reward negatively modulates short latency phasic conditioned responses of dorsal raphe nucleus neurons in freely moving rats. *J Neurosci*. 33(11): 5065–5078. <https://doi.org/10.1523/JNEUROSCI.5679-12.2013>
- Liang Z, Qu A, Shahbaba B. 2025. Meta fusion: a unified framework for multimodality fusion with mutual learning. *arXiv*. <https://doi.org/10.48550/arXiv.2507.20089>
- Liu Y, Ascoli GA. 2007. Value added by data sharing: long-term potentiation of neuroscience research. *Neuroinform*. 5(3):143–145. <https://doi.org/10.1007/s12021-007-0009-0>
- Lomeli LM et al. 2021. Optimal experimental design for mathematical models of haematopoiesis. *J R Soc Interface*. 18(174):20200729. <https://doi.org/10.1098/rsif.2020.0729>
- Luo L, Callaway EM, Svoboda K. 2018. Genetic dissection of neural circuits: a decade of progress. *Neuron*. 98(2):256–281. <https://doi.org/10.1016/j.neuron.2018.03.040>
- Luppi AI et al. 2024. Systematic evaluation of fMRI data-processing pipelines for consistent functional connectomics. *Nat Commun*. 15(1):4745. <https://doi.org/10.1038/s41467-024-48781-5>
- MacDonald CJ et al. 2011. Hippocampal ‘time cells’ bridge the gap in memory for discontinuous events. *Neuron*. 71(4):737–749. <https://doi.org/10.1016/j.neuron.2011.07.012>
- Machens CK, Romo R, Brody CD. 2010. Functional, but not anatomical, separation of ‘what’ and ‘when’ in prefrontal cortex. *J Neurosci*. 30(1):350–360. <https://doi.org/10.1523/JNEUROSCI.3276-09.2010>
- MacKay DJC. 2003. Information theory, inference, and learning algorithms. Cambridge University Press. <https://www.inference.org.uk/itprnn/book.pdf>
- Mackevicius EL et al. 2019. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. *Elife*. 8:e38471. <https://doi.org/10.7554/eLife.38471>
- Magland J et al. 2020. SpikeForest, reproducible web-facing ground-truth validation of automated neural spike sorters. *Elife*. 9:e5167. <https://doi.org/10.7554/eLife.55167>
- Mante V et al. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. 503(7474):78–84. <https://doi.org/10.1038/nature12742>
- Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*. 164(1):177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Markiewicz CJ et al. 2021. The OpenNeuro resource for sharing of neuroscience data. *Elife*. 10:e71774. <https://doi.org/10.7554/eLife.71774>
- Martone M et al. 2020. NIX – Neuroscience information exchange format. *F1000Research*. 9:358. <https://doi.org/10.7490/f1000research.1117858.1>
- Mattoni M et al. 2025. Group-to-individual generalizability and individual-level inferences in cognitive neuroscience. *Neurosci Biobehav Rev*. 169:106024. <https://doi.org/10.1016/j.neubiorev.2025.106024>
- McNaughton BL et al. 2006. Path integration and the neural basis of the ‘cognitive map’. *Nature Rev Neuro*. 7(8):663–678. <https://doi.org/10.1038/nrn1932>
- McNaughton BL, Barnes CA, O’Keefe J. 1983. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp Brain Res*. 52(1):41–49. <https://doi.org/10.1007/BF00237147>
- Meng C et al. 2025. When physics meets machine learning: a survey of physics-informed machine learning. *Mach Learn Comput Sci Eng*. 1(1):20. <https://doi.org/10.1007/s44379-025-00016-0>

- Mitra P, Bokil H. 2007. Observed brain dynamics. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195178081.001.0001>
- Millet D. 2002. The origins of EEG. In: ISHN 2002 Annual Meeting. [accessed 2025 Mar18]. <https://www.semel.ucla.edu/ishn/ab24-2002.htm>.
- Mohsenzadeh Y et al. 2020. Emergence of visual center-periphery spatial organization in deep convolutional neural networks. *Sci Rep*. 10(1):4638. <https://doi.org/10.1038/s41598-020-61409-0>
- Moore GP, Perkel DH, Segundo JP. 1966. Statistical analysis and functional interpretation of neuronal spike data. *Annu Rev Physiol*. 28(1):493–522. <https://doi.org/10.1146/annurev.ph.28.030166.002425>
- Mountcastle VB, Davies PW, Berman AL. 1957. Response properties of neurons of cat's somatic sensory cortex to peripheral stimuli. *J Neurophysiol*. 20(4):374–407. <https://doi.org/10.1152/jn.1957.20.4.374>
- Newman EL et al. 2022. Revealing the dynamics of neural information processing with multivariate information decomposition. *Entropy*. 24(7):930. <https://doi.org/10.3390/e24070930>
- Nguyen JP et al. 2016. Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A*. 113(8):E1074–E1081. <https://doi.org/10.1073/pnas.1507110112>
- Oh SW et al. 2014. A mesoscale connectome of the mouse brain. *Nature*. 508(7495):207–214. <https://doi.org/10.1038/nature13186>
- Ohki K et al. 2005. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*. 433(7026):597–603. <https://doi.org/10.1038/nature03274>
- Okatan M, Wilson MA, Brown EN. 2005. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Comput*. 17(9):1927–1961. <https://doi.org/10.1162/0899766054322973>
- O'Keefe J, Recce ML. 1993. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*. 3(3):317–330. <https://doi.org/10.1002/hipo.450030307>
- Oram MW et al. 1999. Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J Neurophysiol*. 81(6):3021–3033. <https://doi.org/10.1152/jn.1999.81.6.3021>
- Pang R, Lansdell BJ, Fairhall AL. 2016. Dimensionality reduction in neuroscience. *Curr Biol*. 26(14):R656–R660. <https://doi.org/10.1016/j.cub.2016.05.029>
- Pachitariu M et al. 2016. Fast and accurate spike sorting of high-channel count probes with KiloSort. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. [accessed 2024 Feb 13]. <https://proceedings.neurips.cc/paper/2016/hash/1145a30ff80745b56fb0cecf65305017-Abstract.html>
- Pachitariu M et al. 2017. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*. 061507. <https://doi.org/10.1101/061507>
- Pachitariu M, Stringer C, Harris KD. 2018. Robustness of spike deconvolution for neuronal calcium imaging. *J Neurosci*. 38(37):7976–7985. <https://doi.org/10.1523/JNEUROSCI.3339-17.2018>
- Pandarinarath C et al. 2018. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat Methods*. 15(10):805–815. <https://doi.org/10.1038/s41592-018-0109-9>
- Panzeri S et al. 2007. Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol*. 98(3):1064–1072. <https://doi.org/10.1152/jn.00559.2007>
- Pastor L et al. 2021. A unified framework for neuroscience morphological data visualization. *Appl Sci*. 11(10):4652. <https://doi.org/10.3390/app11104652>
- Pepe A et al. 2014. How do astronomers share data? reliability and persistence of datasets linked in AAS Publications and a qualitative study of data practices among US astronomers. *PLoS One*. 9(8):e104798. <https://doi.org/10.1371/journal.pone.0104798>
- Pellegrino A, Stein H, Cayco-Gajic NA. 2024. Dimensionality reduction beyond neural subspaces with slice tensor component analysis. *Nat Neurosci*. 27(6):1199–1210. <https://doi.org/10.1038/s41593-024-01626-2>
- Perkel DH, Gerstein GL, Moore GP. 1967. Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys J*. 7(4):419–440. [https://doi.org/10.1016/S0006-3495\(67\)86597-4](https://doi.org/10.1016/S0006-3495(67)86597-4)
- Pfau D, Pnevmatikakis EA, Paninski L. 2013. Robust learning of low-dimensional dynamics from large neural ensembles. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2013/hash/47a658229eb2368a99f1d032c8848542-Abstract.html
- Pillow JW et al. 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*. 454(7207):995–999. <https://doi.org/10.1038/nature07140>
- Pinsker H, Eberly L. 1982. Spike train analysis in neuroethological studies. *J Electrophysiol Tech*. 8:39–85.
- Platkiewicz J, Stark E, Amarasingham A. 2017. Spike-centered jitter can mistake temporal structure. *Neural Comput*. 29(3):783–803. https://doi.org/10.1162/NECO_a_00927
- Poldrack RA, Mumford JA, Nichols TE. 2011. Handbook of functional MRI data analysis. Cambridge University Press. <https://doi.org/10.1017/CBO9780511895029>
- Quiñan Quiroga R, Panzeri S. 2009. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci*. 10(3):173–185. <https://doi.org/10.1038/nrn2578>
- Reinagel P, Reid RC. 2000. Temporal coding of visual information in the thalamus. *J Neurosci*. 20(14):5392–5400. <https://doi.org/10.1523/JNEUROSCI.20-14-05392.2000>
- Ramezan R et al. 2022. A multivariate point process model for neural spike trains. In: 2022 Conference on Cognitive Computational Neuroscience. 2022 Conference on Cognitive Computational Neuroscience. Cognitive Computational Neuroscience. <https://doi.org/10.32470/CCN.2022.1153-0>
- Ramirez JJ. 2020. Undergraduate neuroscience education: meeting the challenges of the 21st century. *Neurosci Lett*. 739:135418. <https://doi.org/10.1016/j.neulet.2020.135418>
- Redish AD, Elga, Adam N, Touretzky DS. 1996. A coupled attractor model of the rodent head direction system. *Network Comput Neural Syst*. 7(4):671–685. https://doi.org/10.1088/0954-898X_7_4_004
- Reeke GN, Edelman GM. 1988. Real brains and artificial intelligence. *Daedalus*. 117(1):143–173.
- Richmond BJ, Optican LM. 1990. Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. II. Information transmission. *J Neurophysiol*. 64(2):370–380. <https://doi.org/10.1152/jn.1990.64.2.370>
- Ribeiro MT, Singh S, Guestrin C. 2016. ‘Why should I trust you?’ Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ‘16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA. ACM. p 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Riquelme JL, Gjorgjieva J. 2021. Towards readable code in neuroscience. *Nat Rev Neurosci*. 22(5):257–258. <https://doi.org/10.1038/s41583-021-00450-y>
- Rosas FE et al. 2019. Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Phys Rev E*. 100(3-1):032305. <https://doi.org/10.1103/PhysRevE.100.032305>
- Rossant C et al. 2016. Spike sorting for large, dense electrode arrays. *Nat Neurosci*. 19(4):634–641. <https://doi.org/10.1038/nn.4268>
- Rübel O. et al. (2016). Methods for specifying scientific data standards and modeling relationships with applications to neuroscience. *Front Neuroinform*. 10:48. <https://doi.org/10.3389/fninf.2016.00048>
- Rübel O et al. 2022. The neurodata without borders ecosystem for neurophysiological data science. *Elife*. 11:e78362. <https://doi.org/10.7554/eLife.78362>
- Rubinov M, Sporns O. 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*. 52(3):1059–1069. <https://doi.org/10.1016/j.neuroimage.2009.10.003>
- Rupprecht P et al. 2021. A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. *Nat Neurosci*. 24(9):1324–1337. <https://doi.org/10.1038/s41593-021-00895-5>
- Rybakken E, Baas N, Dunn B. 2019. Decoding of neural data using cohomological feature extraction. *Neural Comput*. 31(1):68–93. https://doi.org/10.1162/neco_a_01150
- Sadtler PT et al. 2014. Neural constraints on learning. *Nature*. 512(7515):423–426. <https://doi.org/10.1038/nature13665>

- Sanger TD. 1996. Probability density estimation for the interpretation of neural population codes. *J Neurophysiol.* 76(4):2790–2793. <https://doi.org/10.1152/jn.1996.76.4.2790>
- Sarveniazi A. 2014. An actual survey of dimensionality reduction. *AJCM.* 04(02):55–72. <https://doi.org/10.4236/ajcm.2014.42006>
- Skaggs WE, McNaughton BL, Gothard KM. 1992. An information-theoretic approach to deciphering the hippocampal code. *Adv Neural Inf Process Syst.* 5:1030–1037.
- Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. 2019. High-dimensional geometry of population responses in visual cortex. *Nature.* 571(7765):361–365. <https://doi.org/10.1038/s41586-019-1346-5>
- Saxena R, Barde W, Deshmukh SS. 2018. Inexpensive, scalable camera system for tracking rats in large spaces. *J Neurophysiol.* 120(5):2383–2395. <https://doi.org/10.1152/jn.00215.2018>
- Schlaflly ED et al. 2025. A data augmentation procedure to improve detection of spike ripples in brain voltage recordings. *Neurosci Res.* 215:15–26. <https://doi.org/10.1016/j.neures.2024.07.005>
- Schneidman E et al. 2003. Network information and connected correlations. *Phys Rev Lett.* 91(23):238701. <https://doi.org/10.1103/PhysRevLett.91.238701>
- Schneidman E et al. 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature.* 440(7087):1007–1012. <https://doi.org/10.1038/nature04701>
- Schottdorf M, Yu G, Walker EY. 2024. Data science and its future in large neuroscience collaborations. *Neuron.* 112(18):3007–3012. <https://doi.org/10.1016/j.neuron.2024.08.017>
- Sebastian J et al. 2021. Signal-to-signal neural networks for improved spike estimation from calcium imaging data. *PLoS Comput Biol.* 17(3):e1007921. <https://doi.org/10.1371/journal.pcbi.1007921>
- Sejnowski TJ, Churchland PS, Movshon JA. 2014. Putting big data to good use in neuroscience. *Nat Neurosci.* 17(11):1440–1441. <https://doi.org/10.1038/nn.3839>
- Selvaraju RR et al. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- Shahbaba B et al. 2022. Hippocampal ensembles represent sequential relationships among an extended sequence of nonspatial events. *Nat Commun.* 13(1):787. <https://doi.org/10.1038/s41467-022-28057-6>
- Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J.* 27(3):379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sheintuch L, Rubin A, Ziv Y. 2022. Bias-free estimation of information content in temporally sparse neuronal activity. *PLoS Comput Biol.* 18(2):e1009832. <https://doi.org/10.1371/journal.pcbi.1009832>
- Shneiderman B. 2003. The eyes have it: a task by data type taxonomy for information visualizations. In: Bederson BB, Shneiderman BEN, editors. *The craft of information visualization.* Morgan Kaufmann (Interactive Technologies). p 364–371. <https://doi.org/10.1016/B978-155860915-0/50046-9>
- Siegle JH et al. 2017. Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. *J Neural Eng.* 14(4):045003. <https://doi.org/10.1088/1741-2552/aa5eea>
- Sim I et al. 2020. Time for NIH to lead on data sharing. *Science.* 367(6484):1308–1309. <https://doi.org/10.1126/science.aba4456>
- Simmons JP, Nelson LD, Simonsohn U. 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci.* 22(11):1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simonyan K, Vedaldi A, Zisserman A. 2014. Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv.* <https://doi.org/10.48550/arXiv.1312.6034>
- Skaggs WE et al. 1995. A model of the neural basis of the rat's sense of direction. *Adv Neural Inf Process Syst.* 7:173–180.
- Skaggs WE et al. 1996. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus.* 6(2):149–172. [https://doi.org/10.1002/\(sici\)1098-1063\(1996\)6:2<149::aid-hipo6>3.0.co;2-k](https://doi.org/10.1002/(sici)1098-1063(1996)6:2<149::aid-hipo6>3.0.co;2-k)
- Smetters D, Majewska A, Yuste R. 1999. Detecting action potentials in neuronal populations with calcium imaging. *Methods (San Diego, Calif.).* 18(2):215–221. <https://doi.org/10.1006/meth.1999.0774>
- Smola AJ, Schölkopf B. 2004. A tutorial on support vector regression. *Stat Comput.* 14(3):199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Sporns O. 2011. *Networks of the brain.* MIT Press.
- Sporns O, Betzel RF. 2016. Modular brain networks. *Annu Rev Psychol.* 67(1):613–640. <https://doi.org/10.1146/annurev-psych-122414-033634>
- Stevenson IH, Kording KP. 2011. How advances in neural recording affect data analysis. *Nat Neurosci.* 14(2):139–142. <https://doi.org/10.1038/nn.2731>
- Stosiek C et al. 2003. In vivo two-photon calcium imaging of neuronal networks. *Proc Natl Acad Sci U S A.* 100(12):7319–7324. <https://doi.org/10.1073/pnas.1232232100>
- Stoughton C et al. 2002. Sloan digital sky survey: early data release. *Astron J.* 123(1):485–548. <https://doi.org/10.1086/324741>
- Stringer C, Pachitariu M. 2024. Analysis methods for large-scale neuronal recordings. *Science.* 386(6722):eadp7429. <https://doi.org/10.1126/science.adp7429>
- Sussillo D. 2014. Neural circuits as computational dynamical systems. *Curr Opin Neurobiol.* 25:156–163. <https://doi.org/10.1016/j.conb.2014.01.008>
- Sutter TM et al. 2024. Unity by diversity: improved representation learning for multimodal vaes. *Adv Neural Inf Process Syst.* 37:74262–74297. <https://doi.org/10.52202/079017-2363>
- Szalay A, Gray J. 2001. The world-wide telescope. *Science.* 293(5537):2037–2040. <https://doi.org/10.1126/science.293.5537.2037>
- Tang X, Li L. 2023. Multivariate temporal point process regression. *J Am Stat Assoc.* 118(542):830–845. <https://doi.org/10.1080/01621459.2021.1955690>
- Taube JS. 2007. The head direction signal: origins and sensory-motor integration. *Annu Rev Neurosci.* 30(1):181–207. <https://doi.org/10.1146/annurev.neuro.29.051605.112854>
- Teeters JL, Harris KD, Millman KJ, Olshausen BA, Sommer FT. 2008. Data sharing for computational neuroscience. *Neuroinformatics.* 6(1):47–55. <https://doi.org/10.1007/s12021-008-9009-y>
- Teeters JL et al. 2015. Neurodata without borders: creating a common data format for neurophysiology. *Neuron.* 88(4):629–634. <https://doi.org/10.1016/j.neuron.2015.10.025>
- Tek P et al. 2008. Rapid prototyping for neuroscience and neural engineering. *J Neurosci Methods.* 172(2):263–269. <https://doi.org/10.1016/j.jneumeth.2008.03.011>
- Trautmann EM et al. 2019. Accurate estimation of neural population dynamics without spike sorting. *Neuron.* 103(2):292–308.e4. <https://doi.org/10.1016/j.neuron.2019.05.003>
- Timme NM, Lapish C. 2018. A tutorial for information theory in neuroscience. *eNeuro.* 5(3):ENEURO.0052-18.2018. <https://doi.org/10.1523/ENEURO.0052-18.2018>
- Truccolo W et al. 2005. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol.* 93(2):1074–1089. <https://doi.org/10.1152/jn.00697.2004>
- Tubiana J et al. 2020. Blind deconvolution for spike inference from fluorescence recordings. *J Neurosci Methods.* 342:108763. <https://doi.org/10.1016/j.jneumeth.2020.108763>
- Tudor M, Tudor L, Tudor KI. 2005. [Hans Berger (1873-1941)–the history of electroencephalography]. *Acta Medica Croatica Casopis Hrvatske Akademije Medicinskih Znanosti.* 59(4):307–313.
- Turing AM. 1950. Computing machinery and intelligence. *Mind.* 59:433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Turner HH. 1912. *The great star map: being a brief general account of the international project known as the astrographic chart.* London: John Murray.
- Urai AE et al. 2022. Large-scale neural recordings call for new insights to link brain and behavior. *Nat Neurosci.* 25(1):11–19. <https://doi.org/10.1038/s41593-021-00980-9>
- Varley TF et al. 2023. Information-processing dynamics in neural networks of macaque cerebral cortex reflect cognitive state and

- behavior. *Proc Natl Acad Sci U S A*. 120(2):e2207677120. <https://doi.org/10.1073/pnas.2207677120>
- Varley TF. 2025. Information theory for complex systems scientists: what, why, and how. *Phys. Rep.* 1148:1–55. <https://doi.org/10.1016/j.physrep.2025.09.007>
- Verkhatsky A, Parpura V. 2014. History of electrophysiology and the patch clamp. In: Martina M, Taverna S, editors. *Patch-clamp methods and protocols, Methods in Molecular Biology*. Springer. p 1–19. https://doi.org/10.1007/978-1-4939-1096-0_1
- van Viegen T et al. 2021. Neuromatch academy: teaching computational neuroscience with global accessibility. *Trends Cogn Sci*. 25(7): 535–538. <https://doi.org/10.1016/j.tics.2021.03.018>
- Viejo G et al. 2023. Pynapple, a toolbox for data analysis in neuroscience. *Elife*. 12:RP85786. <https://doi.org/10.7554/eLife.85786>
- Vogelstein JT. 2011. Q&A: what is the open connectome project? *Neural Syst Circuits*. 1(1):16. <https://doi.org/10.1186/2042-1001-1-16>
- Voytek B. 2022. The data science future of neuroscience theory. *Nat Methods*. 19(11):1349–1350. <https://doi.org/10.1038/s41592-022-01630-z>
- de Vries SE, Siegle JH, Koch C. 2023. Sharing neurophysiology data from the Allen Brain Observatory. *Elife*. 12:e85550. <https://doi.org/10.7554/eLife.85550>
- Vu M-AT et al. 2018. A shared vision for machine learning in neuroscience. *J Neurosci*. 38(7):1601–1607. <https://doi.org/10.1523/JNEUROSCI.0508-17.2018>
- Vyas S et al. 2020. Computation through neural population dynamics. *Annu Rev Neurosci*. 43(1):249–275. <https://doi.org/10.1146/annurev-neuro-092619-094115>
- Warbrick T. 2022. Simultaneous EEG-fMRI: what have we learned and what does the future hold? *Sensors (Basel)*. 22(6):2262. <https://doi.org/10.3390/s22062262>
- Weng G et al. 2024. Time-varying generalized linear models: characterizing and decoding neuronal dynamics in higher visual areas. *Front Comput Neurosci*. 18:1273053. <https://doi.org/10.3389/fncom.2024.1273053>
- Wilson MA, McNaughton BL. 1993. Dynamics of the hippocampal ensemble code for space. *Science*. 261(5124):1055–1058. <https://doi.org/10.1126/science.8351520>
- Wise KD, Angell JB, Starr A. 1970. An integrated-circuit approach to extracellular microelectrodes. *IEEE Trans Biomed Eng*. 17(3):238–247. <https://doi.org/10.1109/TBME.1970.4502738>
- Wood F et al. 2004. On the variability of manual spike sorting. *IEEE Trans Biomed Eng*. 51(6):912–918. <https://doi.org/10.1109/TBME.2004.826677>
- Xu Z et al. 2019. A comparison of neural decoding methods and population coding across thalamo-cortical head direction cells. *Front Neural Circuits*. 13:75. <https://doi.org/10.3389/fncir.2019.00075>
- Yamins DLK, DiCarlo JJ. 2016. Using goal-driven deep learning models to understand sensory cortex. *Nat Neurosci*. 19(3):356–365. <https://doi.org/10.1038/nn.4244>
- Yang L et al. 2020. Open source silicon microprobes for high throughput neural recording. *J Neural Eng*. 17(1):016036. <https://doi.org/10.1088/1741-2552/ab581a>
- Yang W, Yuste R. 2017. In vivo imaging of neural activity. *Nat Methods*. 14(4):349–359. <https://doi.org/10.1038/nmeth.4230>
- Yger P et al. 2018. A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *Elife*. 7:e34518. <https://doi.org/10.7554/eLife.34518>
- Yu BM et al. 2009. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J Neurophysiol*. 102(1):614–635. <https://doi.org/10.1152/jn.90941.2008>
- Yu Z et al. 2022. Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. *Neuron*. 110(1):21–35. <https://doi.org/10.1016/j.neuron.2021.10.030>
- Yuan Y et al. 2025. Optimal transport based cross-domain integration for heterogeneous data. *J Am Stat Assoc*. 120(551):1449–1462. <https://doi.org/10.1080/01621459.2025.2540653>
- Zador A et al. 2023. Catalyzing next-generation Artificial Intelligence through NeuroAI. *Nat Commun*. 14(1):1597. <https://doi.org/10.1038/s41467-023-37180-x>
- Zeiler MD, Fergus R. 2013. Visualizing and understanding convolutional networks. *arXiv*. <https://doi.org/10.48550/arXiv.1311.2901>
- Zeiler MD, Fergus R. 2014. Visualizing and understanding convolutional networks. In: Fleet D et al., editors. *Computer Vision – ECCV 2014*. Springer International Publishing. p. 818–833. https://doi.org/10.1007/978-3-319-10590-1_53
- Zhang J et al. 2022. A CNN-transformer hybrid approach for decoding visual neural activity into text. *Comput Methods Programs Biomed*. 214:106586. <https://doi.org/10.1016/j.cmpb.2021.106586>
- Zhang K. 1996. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*. 16(6):2112–2126. <https://doi.org/10.1523/JNEUROSCI.16-06-02112.1996>
- Zhang K et al. 1998. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J Neurophysiol*. 79(2):1017–1044. <https://doi.org/10.1093/cercor/6.3.406>
- Zhang Y, Zhao Y. 2015. Astronomy in the big data era. *CODATA*. 14: 11. <https://doi.org/10.5334/dsj-2015-011>
- Zhang T, Azghadi MR, Lammie C, Amirsoleimani A, Genov R. 2023. Spike sorting algorithms and their efficient hardware implementation: a comprehensive survey. *J Neural Eng*. 20(2):021001. <https://doi.org/10.1088/1741-2552/acc7cc>
- Zhou B et al. 2015. Learning deep features for discriminative localization. *arXiv*. <https://doi.org/10.48550/arXiv.1512.04150>
- Zhou W et al. 2025. A model-agnostic graph neural network for integrating local and global information. *Journal of the American Statistical Association*. 120(550), 1225–1238. <https://doi.org/10.1080/01621459.2024.2404668>
- Zhu F et al. 2022. A deep learning framework for inference of single-trial neural population dynamics from calcium imaging with sub-frame temporal resolution. *Nat Neurosci*. 25(12):1724–1734. <https://doi.org/10.1038/s41593-022-01189-0>
- Ziaei N et al. 2023. Bayesian time-series classifier for decoding simple visual stimuli from intracranial neural activity. In: Liu, F., Zhang, Y., Kuai, H., Stephen, E.P., Wang, H. (eds.), *Brain Informatics. BI 2023. Lecture Notes in Computer Science*, vol 13974. Cham: Springer. https://doi.org/10.1007/978-3-031-43075-6_20
- Zong W et al. 2022. Large-scale two-photon calcium imaging in freely moving mice. *Cell*. 185(7):1240–1256.e30. <https://doi.org/10.1016/j.cell.2022.02.017>