Data mining the functional architecture of the brain's circuitry

Adam S. Charles*

Abstract

The brain is a highly complex organ consisting of a myriad of subsystems that flexibly interact and adapt over time and context to enable perception, cognition, and behavior. Understanding the multi-scale nature of the brain, i.e., how circuit- and moleclular-level interactions build up the fundamental components of brain function, holds incredible potential for developing interventions for neurodegenerative and psychiatric diseases, as well as open new understanding into our very nature. Historically technological limitations have forced systems neuroscience to be *local* in *anatomy* (localized, small neural populations in single brain areas), in *behavior* (studying single tasks), in *time* (focusing on specific stages of learning or development), and in *modality* (focusing on imaging single biological quantities). New developments in neural recording technology and behavioral monitoring now provide the data needed to break free of *local* neuroscience to *global* neuroscience: i.e., understanding how the brain's many subsystem interact, adapt, and change across the multitude of behaviors animals and humans must perform to thrive. Specifically, while we have much knowledge of the anatomical architecture of the brain (i.e., the hardware), we finally are approaching the data needed to find the functional architecture and discover the fundamental properties of the *software* that runs on the hardware. We must take this opportunity to bridge between the vast amounts of data to discover this functional architecture which will face numerous challenges from low-level data alignment up to high level questions of interpretable mathematical models of behavior that can synthesize the myriad of datasets.

1 Introduction

With the constant advancement of new neural recording technologies [12, 8, 24], systems neuroscience has officially joined the era of big data [5, 3]. Simultaneous recordings of tens of neurons has given way to hundreds and thousands [30], with millions of neurons no longer a pipe dream. Moreover, behavioral methods have significantly improved in parallel [27, 20], offering new avenues to train and monitor more complex behaviors, including freely moving animals during neural imaging [29, 10], across organisms. With this explosion in data collection comes both opportunities to create a new data-driven view of neural function, but also challenges at every level from alignment to interpretability.

This opportunity will allow us to treat the brain as the interconnected system it is. For much of neuroscience history, studies at cellular resolution focused on local areas of the brain. Visual neuroscientists looked at the visual cortex, auditory processing was tested in auditory cortex, navigation in hippocampus, emotion and state in amygdala, etc. The brain, however, processes in parallel and distributed ways [23]. Inactivating LIP—an area implicated in decision making—does not necessarily stop an animal from being able to make a decision [13]. Studies in brain loss, and sensory loss redouble this observation, showing that the flexible brain substrate can move computations across the neural circuits to compensate for loss of tissue or to leverage unused resources [2]. Brain-wide recordings can now provide unbiased cellular-level scans that let us map out the *functional architecture*: where and how information spreads and transforms throughout the brain.

A functional architecture would provide a roadmap to the general principles underlying the flexibility, robustness, and efficiency of neural computation. It will give us baselines for core functions that are necessary in healthy brains, which in turn will improve understanding of how observed activity changes in, e.g., neurodegenerative and psychiatric disorders. The current state-of-the-art is to identify brain regions (anatomical areas) that have been linked to aspects of behavior and cognition. However, new global-brain observations see that "everything is everywhere" [6, 14] making it unclear if activity changes in a specific anatomical area must relate to a narrow set of functional deficits.

Quantitatively mapping the functional architecture requires combining a plethora of data taken across brain areas, tasks, and modalities capturing different biophysical signals with external observations, e.g., behavior. Bringing all these data to produce a holistic view of a single computational system that adapts and learns is the primary challenge. Here I discuss a number of these challenges, specifically focusing on challenges in mathe-

---

*Department of Biomedical Engineering, Center for Imaging Science (CIS), Mathematical Institute for Data Science (MINDS), Kavli Neuroscience Discovery Institute (NDI), Data Science & AI Institute, Johns Hopkins University, Baltimore, MD. Contact: adamsc@jhu.edu.

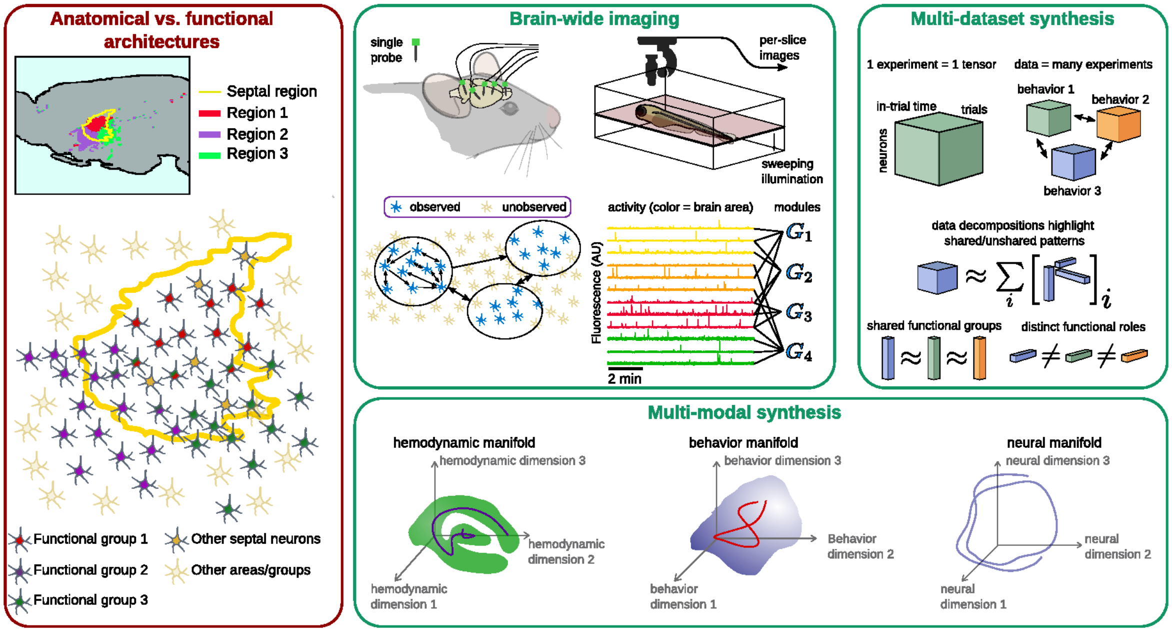

Figure 1: Challenges in discovering the brain's functional architecture. Left: many brain-wide recordings have found that functional relationships in the brain do not always follow the strict anatomical boundaries. Right: areas of advancing our ability to quantitatively find the functional architecture of the brain include synergizing across brain areas, across behavioral contexts, and across recording modalities.

matical modeling that will provide the language we need for a common framework.

2 Rethinking systems-level models

Defining the brain's functional architecture must account for the multiple, distributed systems that comprise brain-wide computations. Most fundamental models in neuroscience do not explicitly seek these subsystems and rather consider every recorded unit as an equal part of a shared state space. Mathematically, if the time-series of unit i is $x_i(t)$, then the vector $\boldsymbol{x}(t) = [x_1(t), \cdots, x_N(t)]^T$ evolves as $\boldsymbol{x}_t = g(\boldsymbol{x}_{t-1}; \theta)$, a Markovian model where the transition function $g(\cdot; \theta)$ are fit to data. Prominent examples of $g(\cdot; \theta)$ include linear dynamical systems (LDS) [7], switched LDS [16], and recurrent neural networks [28]. Non-Markovian models also exist, including generalized linear models (GLMs) [21] and RNNs with memory, e.g., Long-Short Term Memory (LSTM) and General Recurrent Units (GRUs). In all these models, the full system is treated as co-evolving at the same time-scale. Implicitly, the learned network connectivity can recommend disjoint systems, e.g., when the transition matrix is block-diagonal [31]. However, RNNs have a plethora of local minima, making it difficult to over-interpreting dynamics that are fit to finite, noisy data.

Explicitly models of modular dynamics is key to identifying the functional architecture's components. Recent efforts learn modular dynamics, for example the decomposed LDS (dLDS) [17]. dLDS treats dynamics as paths on a manifold $\boldsymbol{x}(t) \in \boldsymbol{M} \subset \mathbb{R}^N$, and finds a set of linear operators $\{\boldsymbol{G}_k\}_{k=1,\ldots,K}$ that, when applied to each point describes the trajectories along the manifold. Each operator describes a distinct set of interactions, and to ensure interpretability, dLDS further assumes that the use of the operators is *sparse*, i.e., the cardinal directions of the flow along the manifold aligns with learned dynamics. Similar manifold partitioning approach also take the approach [25] indicate that compositional descriptions of manifold trajectories appears to be a powerful tool to identify modular brain structure.

3 Challenges in multi-dataset analysis

Synthesizing data across the many recording conditions/context is another key challenge in finding the brain's functional architecture. Data-driven discovery is driven by correlations in the data, and if a task always recruits two systems simultaneously, no analysis can differentiate those systems.

The existence of distinct systems is likely driven by the need to flexibly perform multiple tasks. Recent studies on machine learning systems trained to perform multiple tasks have found that in fact multi-task RNNs

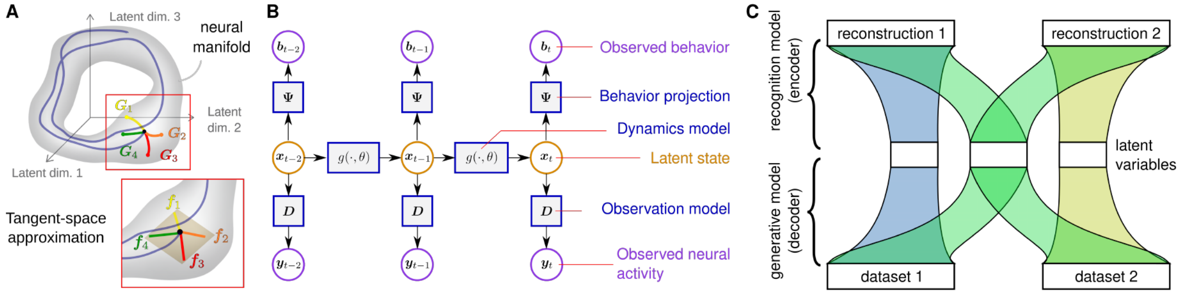

Figure 2: Example neural modeling approaches. A: defining the local geometry of the neural data can provide insight into the key features of the dynamics that define the flows across the manifold. B: dynamical latent states are a key model for combining brain and behavior into a single meaningful system-level model. C: relating disparate datasets requires finding shared and private sources of variance in an interpretable framework.

develop internal modules to support the flexible processing [9].

In neuroscience, finding such modules thus relies on the synthesis of datasets taken across experimental sessions, individuals, and even labs. Neuron-to-neuron alignment is impossible in this case, requiring systems-level alignment instead. Recent efforts have leveraging graph-based multi-matrix (or multi-tensor) decompositions where shared factors can be constrained to represent similar functions, linking between datasets [18, 19]. This link can be, e.g., in neurons if overlapping population are observed across contexts/tasks. Alternatively, the link can be in the trial structure for similar behaviors. While promising, this approach requires some shared trial or neural structure, which will likely need to be modified in each analysis.

4 Challenges in interpretability

Thus far I have avoided mention of the modern artificial neural networks (ANNs) that now form the centerpieces of large-scale data mining in other applications. This is due to the black box nature of ANNs that prevent the types of interpretability necessary for scientific discovery. Scientists need to find relationships between variables that extrapolate our understanding to broader relationships. Thus models such as manifolds, causal analysis, and simple linear systems tend to still be used extensively despite the increased ease of training and deploying ANNs. Specifically, ANNs sacrifice interpretability for expressivity, and so they can *interpolate* on the training data, they cannot *extrapolate* beyond.

This is not to say that no ANN tools can aid neuroscientists. For example, in variational autoencoders, sparsity and independence in the latent layers can promote interpretable disentangling of the data representation [11]. Moreover, emerging approaches in explainable AI [26, 4], can identify key features in the dataset that drive the ANNs. These latter approaches do not provide the extrapolation necessary for scientific discovery, rather they are self-contained explanations of the original data.

Dynamical systems have similar trade-offs in expressivity and interpretability. Large RNNs with LSTM and GRU nodes can predict future data with high accuracy, however the learned RNN parameters are not unique, limiting the insights they can provide about the system interactions. Moreover, nonlinear systems are now being modeled in latent spaces, i.e., low-dimensional representations of the neural data $\boldsymbol{x}(t) = \boldsymbol{\Phi}(\boldsymbol{z}(t))$. When $\boldsymbol{\Phi}(\cdot)$ is nonlinear, also including nonlinear dynamics results in an unidentifiability up to any arbitrary invertible function. Thus, theoretical advances are also needed to understand when such combinations of extremely expressive models creates statistically unidentifiable models.

5 Challenges in multi-modal data

Finally, the brain has more than pyramidal neurons. There is a rich biophysical infrastructure of neurons, astrocytes, vasculature, etc. that modulate neural function beyond direct connections [22] and are therefore a part of the brain's functional architecture. These signals can be measured, e.g., using fMRI, functional ultrasound, optical imaging of neurotransmitter indicators, etc. Moreover, behavioral monitoring provides captures the environment and eventual motor output of the brain.

Sythesizing across data modalities require a single model that describes multiple modalities. Such data fusion is often seeks a latent state $\boldsymbol{z}$ where each mode $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ can be "read out" from the latent state, e.g., $\boldsymbol{x}_1 = f_1(\boldsymbol{z})$ and $\boldsymbol{x}_2 = f_2(\boldsymbol{x}_2)$ [1]. While data fusion can identify one such joint representation, the relationship of multiple datasets is often more nuanced with some information in one mode that is not in the other. A more complete approach separates the latent space into shared ($\boldsymbol{z}_s$) and private ($\boldsymbol{z}_1$ and $\boldsymbol{z}_2$) information, i.e., $\boldsymbol{z} \rightarrow \{\boldsymbol{z}_s, \boldsymbol{z}_1, \boldsymbol{z}_2\}$

ANN architectures, e.g., cross-encoders with private paths, can learn private latents. However, their high expressiveness causes leakage of private information into the shared variables and vice versa. This leakage, while not necessarily damaging in engineering applications, can cause erroneous scientific conclusions about shared brain function. More recent "butterfly" architectures pair multiple cross-encoders with adversarial predictions to minimize such leakage [15].

6 Conclusion

I aim here to lay out a key opportunity in mining the depths of now-available neural data: discovering the brain's functional architecture. In this endeavor, the field will have to so solve at a minimum the mentioned challenges, specifically 1) the synthesis of data collected across brain areas, behaviors, and modalities, 2) the synthesis of brain and behavior data, and 3) the development of interpretable AI that goes beyond the explainable AI currently used in engineering applications.

In the emerging solutions, one emerging theme is the importance of data geometry, specifically going beyond topology and into how the curvature and tangent spaces relate to dynamics. Another theme is the importance of statistically independent representations, which is related to the sparsity that is enjoying a rebound in use from its ability to induce interpretability into regression-type problems. These advances and more will hopefully soon provide new insights into brain function.

References

[1] Andrew et al. Deep canonical correlation analysis. In *ICML*, 2013.

[2] Bedny et al. Language processing in the occipital cortex of congenitally blind adults. *PNAS*, 108(11), 2011.

[3] Benisty et al. Review of data processing of functional optical microscopy for neuroscience. *Neurophotonics*, 9(4), 2022.

[4] Bharti et al. Sufficient and necessary explanations (and what lies in between). *arXiv*, 2024.

[5] Charles et al. Toward community-driven big open brain science: open big data and tools for structure, function, and genetics. *Ann. rev. of neuro.*, 43, 2020.

[6] Chen et al. Brain-wide neural activity underlying memory-guided movement. *Cell*, 187(3), 2024.

[7] Churchland et al. Neural population dynamics during reaching. *Nature*, 487(7405), 2012.

[8] Demas et al. High-speed, cortex-wide volumetric recording of neuroactivity at cellular resolution using light beads microscopy. *bioRxiv*, 2021.

[9] Driscoll et al. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nat. Neuro.*, 27(7), 2024.

[10] El Hady et al. Chronic brain functional ultrasound imaging in freely moving rodents performing cognitive tasks. *J. Neuro. Methods*, 403, 2024.

[11] Geadah et al. Sparse-coding variational autoencoders. *Neur. Comp.*, 2024.

[12] Jun et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679), 2017.

[13] Katz et al. Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature*, 535(7611), 2016.

[14] Koay et al. Sequential and efficient neural-population coding of complex task information. *bioRxiv*, 2021.

[15] Koukuntla et al. Unsupervised discovery of the shared and private geometry in multi-view data. *arXiv*, 2024.

[16] Linderman et al. Bayesian learning and inference in recurrent switching linear dynamical systems. In *AISTATS*, 2017.

[17] Mudrik et al. Decomposed linear dynamical systems (dlds) for learning the latent components of neural dynamics. *JMLR*, 25(59), 2024.

[18] Mudrik et al. Sibblings: Similarity-driven building-block inference using graphs across states. *ICML*, 2024.

[19] Mudrik et al. Creimbo: Cross ensemble interactions in multi-view brain observations. *ICLR*, 2025.

[20] Nath et al. Using deeplabcut for 3d markerless pose estimation across species and behaviors. *Nature Protocols*, 14(7), 2019.

[21] Pillow et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207), 2008.

[22] Randi et al. Neural signal propagation atlas of caenorhabditis elegans. *Nature*, 623(7986), 2023.

[23] Rumelhart et al. A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(45-76), 1986.

[24] Song et al. Volumetric two-photon imaging of neurons using stereoscopy (vTwINS). *Nat. Meth.*, 14(4), 2017.

[25] Tafazoli et al. Building compositional tasks with shared neural subspaces. *bioRxiv*, 2024.

[26] Teneggi et al. Fast hierarchical games for image explanations. *IEEE TPAMI*, 45(4), 2022.

[27] Wu et al. Deep graph pose: a semi-supervised deep graphical model for improved animal pose tracking. *NeurIPS*, 33, 2020.

[28] Yu et al. Extracting dynamical structure embedded in neural activity. *NeurIPS*, 18, 2005.

[29] Zong et al. Large-scale two-photon calcium imaging in freely moving mice. *Cell*, 185(7), 2022.

[30] Stevenson & Kording. How advances in neural recording affect data analysis. *Nat. Neuro.*, 14(2), 2011.

[31] Perich & Rajan. Rethinking brain-wide interactions through multi-region 'network of networks' models. *Cur. op. in neurobio.*, 65, 2020.