*Scalable End-to-end Neuroscience Workflows using Advanced CI*

line 1: 1st Given Name Surname   
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID  
  
line 1: 2nd Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID  
  
line 1: 3rd Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID  
  
line 1: 4th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 5th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 6th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—component, formatting, style, styling, insert (key words)

# Introduction

Neuroscience research is increasingly becoming data-intensive from genomic to behavior levels, and there is an increasing need to integrate and analyze voluminous data being generated at all these levels for advanced explorations into the functioning of brains. Importantly, such research also requires multi-disciplinary expertise, e.g., in computer science, engineering, biology, psychology, to effectively reverse engineering the functioning of brains [1]. In parallel, interest among computer science and engineering researchers is growing in *computational* and *hardware* aspects of neural engineering, and among educators who are creating neuroscience majors at universities across the nation [2]. The rapidly growing data of multiple types, and spread across wide temporal (milli-seconds to days) and spatial (nanometers to meters) creates unique challenges related to integration, management, and analysis, for cutting-edge neuroscience research.

Challenges in advanced computational neuroscience research. Computation in neuroscience involves software for modeling of neurons, synapses and networks at biophysical first-principles level to cognitive modeling and to statistical/machine learning model levels [3]. Equally or perhaps more important are issues related to *neuroinformatics* that deals with integration of diverse types of neural data, such as neuroimaging, electrophysiology, and genetic information [4], and a variety of techniques to analyze such data, including for other applications such as *neuroimaging*, and *brain-machine-interfaces* [5]. The widespread interest in neuroscience among students at all levels and the public in general has led to *reverse-engineering the brain* being recognized as a grand challenge by the National Academy of Engineering [6]. However, neuroscience researchers including graduate students, postdoctoral fellows, and faculty—the *targeted communities in this project*—face challenges in accessing relevant resources, and in interacting with collaborators in related disciplines to advance research in neuroscience. Our proposal will focus on developing advanced CI with applicability to computational neuroscience at the molecular, cellular and systems/network levels, with the tools also have applicability to several of the other areas above.

Advanced research in computational neuroscience critically relies on access to distributed resources, including multiple software packages (reviewed in [7]), high-performance computing (HPC)/cluster computing with large numbers of cores, fast storage, hardware accelerators such as graphics processing units (GPUs), data sharing/collaboration capabilities, big data management (e.g., using Apache Hadoop [8], Apache Spark [9]) and data archival (e.g., NeuronDB [10]). Further, the availability of advanced cyberinfrastructure (CI) technologies including distributed/heterogeneous resources (e.g., FABRIC [11], CloudLab [12]) provide new opportunities to conduct large-scale neuroscience computations. In addition, the targeted communities lack training in effectively using these CI resources to improve their scientific productivity and pursue large-scale data-enabled investigations. Such technologies are crucial for science teams [13] as they collaborate to reverse engineer the human brain, with its 100 billion neurons and 150 trillion synapses. To address these challenges, our proposed project aims to design, develop, and evaluate *end-to-end scalable neuroscience workflows* on FABRIC, which is a programmable infrastructure with extensible networking elements and large amounts of compute and storage capabilities scattered throughout the network, to potentially transform neuroscience research and training.

Challenges in developing advanced CI for neuroscience research.Neuroscience continues to be challenged by *the data deluge* due to rapid advances in data acquisition, processing, management, and archival technologies. For instance, even for modeling single neurons at a single level of scale (cellular, Figure 2) the compute time for biologically realistic models is inordinate, taking hours and weeks [REF]. This, together with the fact that the human brain has ~100 billion neurons, precludes the generation of realistic models even for a region of the brain, let alone the brain itself, even for rodents [REF]. Note that other levels of neuroscience such as genomic and molecular on the lower side of the ‘cellular’ level, and behavioral and clinical on the higher side, are presently not conceivable. So, critical multi-level investigation of brain circuits, a goal of advanced research and of federal agencies such as NSF and NIH are presently not possible.

The diversity of software packages and data standards used by the neuroscience community have further exacerbated the situation. Figure 3 shows that as network size increases, only simplified models are being developed presently - column 2 has simple models as the first four columns and multi-compartmental realistic one as the fifth, and the realistic model is absent in column 5 due to the inordinate time taken and the lack of adequate CI support. So, advanced CI holds promise for tackling this challenge. Furthermore, end-to-end, and scalable workflows that seamlessly integrate neuroscience applications with CI technologies is necessary to make this accessible to the neuroscientists to enable them to advance cutting-edge neuroscience research.

*HPC vs Distributed Computing for Neuroscience*.

Traditional HPC resources are widely used for computational neuroscience workloads. The Neuroscience Gateway (NSG) [14] is an exemplar for engaging computational neuroscientists to leverage HPC resources (using XSEDE/ACCESS [15]) with a low barrier to entry. In fact, NSG continues to remain the leading consumer of ACCESS resources – consuming more than 40% more than the next user. However, the technology landscape is constantly evolving creating new opportunities for innovation and training. The growing popularity of cloud computing for HPC applications [16-18] demands new research workflows that integrate neuroscience with advanced CI beyond traditional HPC environments. While past neuroscience development efforts have targeted the use of HPC through middleware services/workflows [19], the barrier to entry was high. Hence, the CI adoption was poor. The heterogeneity of hardware, high-speed networking, and the ability to customize one’s computing environment in advanced CI resources demands rethinking how end-to-end neuroscience workflows can be seamlessly integrated with these resources.

Scaling neural network simulations has been an ongoing challenge in the neuroscience community. Migliore et al. [20] investigated the scaling of NEURON network simulations using Message Passing Interface (MPI) and multiprocessor systems such as a Beowulf cluster and a supercomputer. They observed that spike communication overhead was less of a bottleneck compared to cache memory effects. As the number of processors was increased, the problem size became small enough that a supercomputer achieved better speedup. Later Hines et al. [21] showed how models implemented in NEURON can be parallelized using MPI for executing on local resources (e.g., a single computer, a cluster of computers connected by Ethernet, a multiprocessor computer) as well as supercomputers.

Recent efforts have explored the use of GPUs for accelerating neuronal biophysical simulations (e.g., Arbor [22], CoreNeuron [23]). They focused on simulating large scale neuronal networks with multiple compartments. More recently, NeuroGPU [24] accelerated multi-compartment neuron simulations using GPUs and was 200 times faster than NEURON simulation on a single core. It used CUDA to optimize memory management on the GPUs. NeuroGPU was also faster than NEURON parallelized using MPI. Another recent effort used Docker containers to accelerate NEURON models using MPI in a commodity cluster [25]. There in growing interest in scaling neuroscience computations beyond traditional HPC environments, including using distributed computing.

Created in 2019, FABRIC is a nationwide NSF research infrastructure and testbed for advancing computer science research and science applications. It is a programmable infrastructure with extensible networking elements and large amounts of compute and storage capabilities scattered throughout the network. High speed optical links interconnect 30 geographically distributed sites on FABRIC via different Layer 2 and Layer networking services [26]. FABRIC has Internet2 connectivity to public clouds such as Amazon Web Services (AWS) and Microsoft Azure. A FABRIC node, which contains a rack of compute, storage, and networking devices, is equipped with cutting-edge processors, large amounts of RAM, non-volatile memory express (NVMe) drives, GPUs, and 100/200 Gbps SmartNICs. Experimenters on FABRIC can choose virtual machine (VM) or container configurations for provisioning resources. Indeed, FABRIC is an advanced CI resource that can have a transformative impact on next-generation neuroscience research and education. To the best of our knowledge, *none have explored this opportunity for neuroscience*.

Scalable, end-to-end neuroscience workflows involve both neural model simulation and machine learning along with interactions with biologists/experimenters/animals. Using FABRIC for these workflows poses interesting challenges that are discussed next. (a) FABRIC’s programmability allows experimenters to set up a heterogenous cluster with CPUs, GPUs, and/or SmartNICs, spanning a single site or across multiple sites. End-to-end neuroscience workflows, which include neural model simulation and machine learning, will pose new challenges in load balancing and resource utilization in such a heterogenous cluster. These challenges must be addressed to achieve good scalability and performance.

(b) Neural model simulations can access public databases during execution. These databases should be accessed in near real-time for the simulations to execute fast. As these databases are large, they must be partitioned and distributed carefully across the cluster to enable low-latency access. One possibility is to employ distributed in-memory database technologies with the ability to slide and dice data efficiently. (c) Model-in-loop experimentation allows collaboration across institutions and research labs. For example, a rodent can be controlled at Rutgers University using a neural simulation running at University of Missouri. To enable such experimentation on FABRIC, low-latency, high-speed data transfer between geographically separated FABRIC sites is necessary. To achieve this, appropriate network paths must be dynamically selected on FABRIC to achieve real-time coordination.

In summary, the features of advanced CI resources (i.e., FABRIC) need to be carefully tailored to the diverse neuroscience research cases. Yet, the workflow of a typical neuroscience user should be minimally disrupted to encourage adoption.

# METHOD

## Case Studies

* Running on one FABRIC node vs two nodes
* Data management for large-scale models – need to access data from the neurobiologist (and databases) and interface with them in real-time…using machine learning example
* Real-time coordination of model-in-loop experiments

## XXXXXXXX

Identify applicable funding agency here. If none, delete this text box.

……………..

# RESULTS

**Scalable and efficient single- and multi-cell neural model simulation**

A diagram of a network model

Description automatically generated

TO DO:

* What is the limit for one GPU?

Case 1

* One VM with 1 GPU vs 2 or more GPUs --- novelty?

Case 2

* One VM with one GPU vs two VMs with 1 GPU each - run on 1 node

TO CHECK: Does the present MPI code consider two nodes?

**Scalable and efficient data management for large-scale neural model simulation**

***A diagram of a cloud

Description automatically generated***

**Clinical level** – Use Case: **Linking psychosocial functions with intrinsic functional connectivity of brain activity**. Collaborator - **Dr. Seung Lark Lim**from the Kansas City campus.

**Efficient real-time multi-site protocols for model-in-loop experimentation**

**A diagram of a cloud computing system

Description automatically generated**

# DISCUSSION

The proposed integrated neuroscience research workflows on NSF-funded testbeds will be broadly available to researchers and students internationally for fostering innovations in neuroscience and CI.

The training materials will be incorporated into existing courses at the four state-wide campuses of the University of Missouri System, and into summer training workshops targeted separately for CI Users (e.g., scientific domain researchers in neuroscience, engineering, psychology) and for CI Contributors (e.g., CI researchers/software engineering).

##### Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[1] J. Ngai, “BRAIN 2.0: Transforming neuroscience,” *Cell,* vol. 185, no. 1, pp. 4-8, 2022/01/06/, 2022.

[2] B. Latimer, D. Bergin, V. Guntu, D. J. Schulz, and S. S. Nair, “Integrating Model-Based Approaches Into a Neuroscience Curriculum—An Interdisciplinary Neuroscience Course in Engineering,” *IEEE Transactions on Education,* vol. 62, no. 1, pp. 48-56, 2019.

[3] P. Dayan, and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, Cambridge, MA: MIT Press, 2005.

[4] T. M. Morse, “Neuroinformatics: from bioinformatics to databasing the brain,” *Bioinform Biol Insights,* vol. 2, pp. 253-64, May 14, 2008.

[5] !!! INVALID CITATION !!! [5-8].

[6] NAE, *Grand Challenges for Engineering, National Academy of Engineering engineeringchallenges.org/*, 2008.

[7] !!! INVALID CITATION !!! [10].

[8] T. White, *Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale, 4th Edition*: O'Reilly Media, 2015.

[9] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, and e. al., “Apache spark: a unified engine for big data processing,” *Communications of the ACM*, no. 11, pp. 56-65, 2016.

[10] R. A. McDougal, T. M. Morse, T. Carnevale, L. Marenco, R. Wang, M. Migliore, P. L. Miller, G. M. Shepherd, and M. L. Hines, “Twenty years of ModelDB and beyond: building essential modeling tools for the future of neuroscience,” *Journal of Computational Neuroscience,* vol. 42, no. 1, pp. 1-10, February 01, 2017.

[11] I. Baldin, A. Nikolich, J. Griffioen, I. I. S. Monga, K. C. Wang, T. Lehman, and P. Ruth, “FABRIC: A National-Scale Programmable Experimental Network Infrastructure,” *IEEE Internet Computing,* vol. 23, no. 6, pp. 38-47, 2019.

[12] D. Duplyakin, R. Ricci, A. Maricq, G. Wong, J. Duerig, E. Eide, L. Stoller, M. Hibler, D. Johnson, K. Webb, A. Akella, K. Wang, G. Ricart, L. Landweber, C. Elliott, M. Zink, E. Cecchet, S. Kar, and P. Mishra, “The Design and Operation of CloudLab,” in 2019 USENIX Annual Technical Conference (USENIX ATC 19), 2019, pp. 1--14.

[13] M. Parashar, A. Friedlander, E. Gianchandani, and M. Martonosi, “Transforming Science through Cyberinfrastructure,” *Commun. ACM,* vol. 65, no. 8, pp. 30–32, jul, 2022.

[14] S. Sivagnanam, K. Yoshimoto, N. T. Carnevale, and A. Majumdar, “The Neuroscience Gateway: Enabling Large Scale Modeling and Data Processing in Neuroscience,” in Proceedings of the Practice and Experience on Advanced Research Computing, Pittsburgh, PA, USA, 2018.

[15] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr, “XSEDE: Accelerating Scientific Discovery,” *Computing in Science & Engineering,* vol. 16, no. 5, pp. 62-74, 2014.

[16] "High Performance Computing - AWS," /aws.amazon.com/hpc/, 2022.

[17] "High Performance Computing Solutions," /cloud.google.com/solutions/hpc, 2022.

[18] "Azure high-performance computing," /azure.microsoft.com/en-us/solutions/high-performance-computing, 2022.

[19] P. Calyam, and S. S. Nair, “Science Gateway Development to aid Cyber and Software Automation for Neuroscience Researchers and Educators, Austin TX (Sept 25-27, 2018),” *13th Gateway Computing Environments Conference (Gateways)*, 2018.

[20] M. Migliore, C. Cannia, W. W. Lytton, H. Markram, and M. L. Hines, “Parallel network simulations with NEURON,” *Journal of Computational Neuroscience,* vol. 21, no. 2, pp. 119-129, 2006/10/01, 2006.

[21] M. L. Hines, and N. T. Carnevale, “Translating network models to parallel hardware in NEURON,” *J Neurosci Methods,* vol. 169, no. 2, pp. 425-55, Apr 30, 2008.

[22] N. A. Akar, B. Cumming, V. Karakasis, A. Küsters, W. Klijn, A. Peyser, and S. Yates, "Arbor - A Morphologically-Detailed Neural Network Simulation Library for Contemporary High-Performance Computing Architectures." pp. 274–282.

[23] P. Kumbhar, M. Hines, J. Fouriaux, A. Ovcharenko, J. King, F. Delalondre, and F. Schürmann, “CoreNEURON : An Optimized Compute Engine for the NEURON Simulator,” *Front Neuroinform,* vol. 13, pp. 63, 2019.

[24] !!! INVALID CITATION !!! [27].

[25] R. S. Iyengar, and M. Raghavan, "MPI Parallelization of NEUROiD Models Using Docker Swarm." pp. 655-660.

[26] P. Ruth, I. Baldin, K. Thareja, T. Lehman, X. Yang, and E. Kissel, "FABRIC Network Service Model." pp. 1-6.

[27] !!! INVALID CITATION !!! [51, 73, 74].