U.S. DEPARTMENT OF ENERGY | Office of Science

Bioimaging Capabilities to Enable Mapping of the Neural Connections in a Complex Brain

DOE Virtual Roundtable:

Summary Report

November 2 - 4, 2020

U.S. DEPARTMENT OF ENERGY | Office of Science

Bioimaging Capabilities to Enable Mapping of the Neural Connections in a Complex Brain

Summary of the Department of Energy (DOE) virtual roundtable on
Bioimaging Capabilities to Enable Mapping of the Neural Connections in a Complex Brain

Roundtable Co-Chairs

Rick Stevens	Associate Laboratory Director, Argonne National Laboratory
Galya Orr	Staff Scientist, Pacific Northwest National Laboratory

Department of Energy Contact

R. Adam Kinney	Office of the Principal Deputy Director, Office of Science

Speakers

Connectome Requirements	Ed Boyden, Bobby Kasthuri, Jeff Lichtman, Mark Schnitzer, Hongkui Zeng, Clay Reid
Bioimaging Technology	Stefan Vogt, Tianyi Mao, Na Ji, Wah Chiu, Xianghui Xiao, Carolyn Larabell, Mark Ellisman, Jim Werner, Lisa Miller, Tamas Varga, Scott Lea, Zhuhao Wu, Vincent De Andrade, Ken Hayworth, Scott Retterer
Computing and Data	Garrett Kenyon, Frank Alexander, Nicola Ferrier, Kris Bouchard, Bobbie-Jo Webb-Robertson, Katie Schuman, Tamie Borders

Administrative & Editorial Support Silvia Mulligan, Emily M. Dietrich

DOI: 10.2172/1821173

Contents

Executive Summary

Motivation

The Brain Research through Advancing Innovative Neurotechnologies® (BRAIN) Initiative has identified grand challenges that involve complex and multiscale lines of inquiry that will require new technological, scientific, and organizational inventions. These efforts will have the potential to impact public health at the scale of the Human Genome Project (HGP). Similar to the HGP, the Department of Energy (DOE) could play a pivotal role in helping to realize the potential of these efforts. One of the projects selected for pursuit under the BRAIN initiative, "Next-generation Technology for Microconnectivity Analysis," seeks to pilot technologies for developing comprehensive maps of all the neural connections in the brains of complex organisms such as the mouse, non-human primate, and, ultimately, human. This goal will require the co-development of novel high-throughput, highly parallel imaging technology with new data management and analysis tools coupled with high-performance computing optimized for artificial intelligence and machine learning.

The development and implementation of such large-scale experimental and data analysis infrastructure has been an area of expertise for DOE. DOE and its complex of national laboratories is uniquely experienced in scoping large-scale scientific capabilities that integrate multiple techniques from across the complex in order to address well-defined DOE mission needs and push scientific and technical capabilities beyond the current state of the art.

DOE's Office of Science (SC), through its BRAIN Initiative Working Group, convened a virtual roundtable on November 2 - 4, 2020 to develop a conceptual framework for the capabilities needed to enable an interdisciplinary community of researchers to fully map the neural connections initially in the brain of a mouse and ultimately in the brain of a human. Over 30 researchers and programmatic observers from diverse scientific backgrounds, including those working in both the basic and applied sciences, participated in the roundtable and conceptualized forward-leaning ideas. In developing the framework, participants were asked to address questions under the three themes: connectome requirements, bioimaging technology, and computing and data.

Connectome Requirements

Questions Posed: From a neuroscience perspective, what are the neuronal properties, as well as the imaging and computing requirements, to fully map the connections in the brains of a mouse, a human, and other relevant mammals? What will be the impact of scaling from one mouse brain to many mouse brains to one human brain to many human brains?

Findings: Connectivity maps at synaptic resolution (<10 nm) and long-range axonal projections (100 nm), registered with images of neural activity, will greatly enhance the potential for scientific discovery in connectomics. The development of a full connectome in a wild-type mouse would allow the community to establish methods before moving to the study of mice at different developmental stages or executing different tasks, and more complex brains like non-human primates or humans.

Challenges: No single imaging modality can currently provide all the information needed to realize complex brain connectomes at multiple scales that are registered with neural activity. Further, current technologies cannot achieve the throughput required to achieve whole brain connectome at synaptic resolution within a reasonable time frame.

Opportunities: DOE expertise in developing high throughput multiplexed technologies, required to support whole brain connectomes at multiple scales registered with activity, will revolutionize brain research and applications to preventing and treating neural diseases. The connectomes will be used to understand the rules of brain organization, such as cell type and synaptic diversity and network properties, and uncover structure-function relationships. Ultimately, the human connectome will be used to understand disease and other attributes such as learning and memory and the emergence of intelligence through evolution.

Bioimaging Technology Roadmap

Questions Posed: Focusing on imaging modalities, including optical, electron, and X-ray technologies, what are the technical R&D hurdles preventing the achievement of the goals defined under connectome requirements (i.e., resolution, sample preparation, and data acquisition in high-throughput and parallel configurations)?

Findings: Optical microscopy can image relatively thick samples (>100 µm) and large volumes (mm^3) in a reasonable time, can detect cell and molecular identities, and can image activity for functional characterization. Electron microscopy is the only modality currently capable of resolving synaptic nanostructure. X-ray methods are nondestructive and suited to imaging hydrated, cryopreserved samples. The high penetration power of X-rays enables imaging of large volumes (mm^3) at relatively high spatial resolution in 3D (<10-30 nm).

Challenges: All modalities are limited in some fashion, and no single modality can provide all the information needed. Optical and current X-ray methods cannot resolve synaptic nanostructures. EM methods are limited to thin sections that require massive stitching, registration, and segmentation, and provide limited information about molecular identity and neural activity. The approaches and protocols needed for X-ray imaging of brain tissue are insufficiently developed. All methods have sample preparation challenges (e.g., preservation, staining and sectioning) that must be overcome to both scale to more complex brains and to integrate multiple techniques.

Opportunities: DOE's technology development expertise, coupled with co-location of modalities at individual laboratories, provides an opportunity to significantly advance high-throughput, massively parallel imaging with a concomitant development of the pipelines for image reconstruction. Co-location also offers unique opportunities to develop novel correlative imaging techniques combing EM, X-ray, and/or optical methods. DOE could leverage its unique position as steward of bright, coherent X-ray light sources to develop dedicated connectomics beamlines that will take advantage of the next generation of light sources. DOE expertise in the development of optics and detectors can be leveraged across all modalities. Finally, DOE's high bandwidth information backbone, ESnet, connects all DOE sites and presents a unique opportunity to explore a distributed "facility" for connectomics.

Computing and Data Roadmap

Questions Posed: What is the R&D required to address the computing and data environment needed to host and process connectome related datasets, as well as the algorithms and software stacks needed, to automate the handling and transformation of datasets generated?

Findings: At the resolution needed for resolving synaptic sub-structures, a mouse brain and human brain are approximately an exavoxel and zettavoxel, respectively. DOE's forthcoming exascale computers can enable significant progress on the mouse connectome, with future computers expected to be capable of full mouse connectome reconstruction if parallel progress is made in advancing imaging technology and analysis pipelines. To reconstruct a human connectome in a year, a million exaflop/s days would be required, which would require 10^4 exascale computers. Even with anticipated improvements in processing of 100x by 2025, an additional two orders of magnitude improvement in processing are needed.

Challenges: The current estimate of mouse image raw data is roughly the scale of the main filesystems on Exascale computers (10^{18} bytes). The human connectome is approximately 1000x larger. Storage capacity and storage throughput is growing more slowly than computing capacity improvements. The current workflows have major human in the loop components that would need higher levels of automation to be practical.

Opportunities: The co-location of imaging capabilities and advanced computers at multiple DOE laboratories positions DOE to make advances in (near) real time processing of connectomic data at the scale needed for the mouse and human brain. DOE's history of driving development of computing technology positions it to explore purpose designed software and hardware for connectomics that could yield the multiple orders of magnitude improvement in processing needed to reconstruct connectomes of complex brains in months or years, rather than decades. Advancing artificial intelligence (AI), machine learning, and visualization could enable human curators to develop better training data for AI algorithms designed to support connectomics.

Applications to DOE Missions

The discoveries and capabilities resulting from a large-scale effort to completely map the neural connections in the brain of a complex organism would benefit DOE's missions in science, energy, national security, and environmental stewardship. For example, the large-scale image-based 3D reconstruction of complex nanoscale structures required for the connectome could fill current gaps between molecular and cellular level studies and field studies at scales required for accurate, high-resolution modeling of the environment. These new capabilities could advance large scale data generation and analysis of omics data and assignments of functions to genomes to inform large scale systems biology models for engineering more resilient biological systems and the production of biofuel and bioproducts. Similar applications exist in imaging novel battery nanomaterial, large-scale analysis of semiconductors, and advanced manufacturing processes, including opportunities in designed self-assembly and complex polymers. Discoveries from building the mouse brain connectome may yield new ideas for AI models, new approaches to automated experimental control, and new insights into resilient, low-power neuromorphic computing that will benefit DOE's diverse mission space. Finally, the use cases of the connectome complement existing drivers for DOE in advancing computing and

the application of AI and data science, including the development of a supercomputing enabled imaging ecosystem that will accelerate scientific discovery across the physical sciences.

Introduction & Motivation

The National Institutes of Health (NIH) Brain Research through Advancing Innovative Neurotechnologies® (BRAIN) Initiative is pursuing a project entitled, "Next-generation Technologies for Microconnectivity Analysis," one of several transformative project concepts that was developed out of the findings of the Advisory Committee to the NIH Director BRAIN Initiative Working Group 2.0. This project seeks to pilot technologies for developing comprehensive maps of all the neural connections in the brains of complex organisms such as the mouse, non-human primate, and, ultimately, human. This goal will require the co-development of novel high-throughput, highly parallel imaging technologies with new data management and analysis tools coupled with high-performance computing optimized for artificial intelligence and machine learning. The development and implementation of such large-scale experimental and data analysis infrastructure has been an area of expertise for the Department of Energy (DOE). The discoveries and capabilities resulting from such a large-scale effort will benefit DOE's missions in science, energy, national security, and environmental stewardship.

As a first step, DOE's Office of Science (SC), through its BRAIN Initiative Working Group, convened a virtual roundtable on November 2 - 4, 2020 to develop a conceptual framework for the capabilities needed to fully map the neural connections initially in a mouse brain and ultimately in a human brain. Over 30 researchers from DOE national laboratories, academic institutions, and private research institutions, representing multiple scientific communities of interest, participated in the roundtable and were asked to address questions under the following three themes: I) Connectome Requirements, II) Bioimaging Technology Roadmap, and III) Computing and Data Roadmap. The findings derived from the Roundtable discussions and summarized in this report provide a concise summary of the required anatomical, molecular and physiological information for connectome reconstruction, the state-of-the-art imaging techniques for acquiring the large scale datasets, the computing methods and requirements for processing those datasets to reconstruct synapse resolving connectome models, and addressed the biological motivations for projects aiming to acquiring mouse, human, and other species connectomes to address foundational questions in neuroscience.

Below is a concise summary of the high-level findings and opportunities within each of the three thematic areas that were synthesized from the three-day meeting. In the following sections, a more detailed exploration of each theme is provided. The report concludes by offering a set of unique opportunities for DOE to contribute to a national-scale connectome effort, as well as a brief consideration of the ways in which the progress made in imaging and computational technologies through this effort could advance DOE's diverse mission space.

I. Connectome Requirements

Key neuronal properties required for reconstructing a whole brain connectome were described, including cell morphology and location**,** synapse anatomy, molecular identities, and functional

information that will be registered with microanatomical, morphological, and molecular information. All participating neuroscientists saw a great value in registering cortex-wide imaging of neural activity with the structural connectome map. Imaging neural activity is often done by voltage sensitive dyes or by calcium imaging using genetically encoded calcium indicators. Such approaches can be readily scaled to ~1 million cortical neurons.

Both microscale (~10 nm resolution) and mesoscale (~100 nm resolution) connectomes (also referred to as a projectome) are needed. A microscale, or synaptic, connectome requires a ~10x10x10 nm voxel resolution over a 1 cm^3 volume of mouse brain. There are 75-100 million neurons and 100 billion synapses in the mouse brain. For a mouse brain, the data size is estimated to be ~1 exabyte (depending on voxel size). However, the human brain is 1000 times larger. The initial goal could be a complete microscale connectome of a wild-type mouse brain, during which approaches would be established. This effort could be followed by mapping the connectomes of mice executing specific tasks, or mice at different developmental stages. These connectomes will be used to understand the rules of brain organization, such as cell type and synaptic diversity and network properties, and uncover structure-function relationships. Ultimately, the human connectome will be used to understand disease and other attributes such as learning and memory and the emergence of intelligence through evolution.

In parallel to a microscale connectome, it was suggested that a complete mesoscale connectome for the mouse brain (petascale, achievable in 1 year) and the human brain (exascale, achievable in 5-10 years) be pursued. A complete human brain projection map could have immediate applications in clinical neurology. Accurate maps of long-distance axons could increase and change the understanding of existing treatments and inform the design of new treatments. It was estimated that with ~100 fluorescence microscopes (e.g., light sheet microscopes), a mesoscale connectome of the human brain at 100 nm resolution could be achieved in 1 year. An array of such microscopes, together with significant computation at the edge for quality control, preprocessing, and data compression, along with the resources needed to deploy machine learning at the exascale, is well within the province of DOE to enable.

II. Bioimaging Technologies

A combination of imaging modalities will be required to achieve whole brain connectomes. Electron microscopy (EM) is currently the only technology that can accurately resolve a synapse, including nanoscale structures such as synaptic vesicles and membranes (<10 nm). However, current EM technologies can only image thin volumes and requires massive registration, stitching, and segmentation to reconstruct the connectome within larger volumes. Current EM techniques are therefore limited in throughput to visualizing millimeter-scale dimensions in a reasonable timeframe. In addition, EM currently has a relatively limited ability to distinguish between multiple distinct molecular entities, although advances in this direction have been made. In contrast, fluorescence microscopy, combined with expansion microscopy, can only reach 20-40 nm resolution at best, but can clearly distinguish between multiple molecular entities, as well as image neural activity for functional characterization. Optical microscopy approaches can image relatively thick samples (>100 µm) and large volumes (mm^3)

in a reasonable time. Correlated light and electron microscopy (without harming ultrastructure) could provide ultrastructural signposts for cell and synapse types. X-ray techniques, with their nondestructive, high penetration power can image, in principle, large volumes (mm^3) with <10-20 nm resolution in hydrated cryo-preserved tissue with minimal sample preparation. However, X-ray technologies have yet to demonstrate the ability to resolve synapses and require the development of approaches and protocols for connectome imaging. Similar to EM, X-ray imaging techniques are currently limited in the ability to distinguish molecular identities, although some genetically encoded probes (i.e., lanthanide-binding tags) have been developed. Combining X-ray and electron or light microscopy modalities has been proposed for hierarchical, multiscale imaging. All imaging modalities, with no exceptions, will require massive parallelization and automation to reach the goal of mapping one mouse brain, let alone one human brain, in a year.

Key technologies under each of the imaging modalities have been discussed. Under EM technologies, a hybrid thick sectioning + broad ion milling approach called GCIB-SEM (Gas Cluster Ion Beam SEM) + MultiSEM has been proposed to have what it takes to scale all the way up to mouse and human brains. By collecting thick sections (100 nm to 10 microns) on sturdy silicon wafers, GCIB-SEM should be able to achieve the 100% reliability required even when wide blocks and frequent knife changes are considered. GCIB-SEM can be implemented in a massively parallel fashion by passing wafers among potentially thousands of MultiSEM and broad ion milling machines in a manner similar to a semiconductor manufacturing plant. An estimated 20 MultiSEMs would be required for reconstructing one mouse brain, and 3000 for reconstructing one human brain, in 5 years. Developments in the MultiSEM optical column and redesign for mass-production are likely to increase throughput and decrease cost. Approaches to optimize sample preparation and increase throughput have been discussed.

Under optical microscopy technologies, a combined expansion microscopy + lattice light sheet microscopy + Brainbow (or other multiplex neuronal labeling) approach has been proposed for connectomics. With current technology, using 8x expansion (effective resolution of 12.5 nm x 12.5 nm x 25 nm) and 5 fluorescent colors (3 for Brainbow and 2 for pre- and post-synaptic markers), a whole mouse brain connectome (500 mm^3, image size ~1280 petabyte) was estimated to be acquired in ~14 years, and a whole human brain (1,200,000 mm^3) in ~35,000 years. Increasing throughput by increasing the field of view to a larger format or arrays of spatial light modulators and cameras could decrease significantly the estimated timeframes. It was estimated that with such improvements, one human brain connectome could be achieved in less than 5 years. Operation of multiple imaging systems in parallel and automation of sample preparation, as well as investments in data storage and analysis (i.e., registration, stitching, segmentation), will be required.

Functional characterization could significantly increase the value and impact of the connectome. Several fluorescence microscopy approaches for imaging neural activity were discussed, including adaptive optics, Bessel focus in a two-photon fluorescence mesoscope, 16-beam two-photon microscope, and 'Crystal Skull' imaging by epifluorescence. Large-scale

parallelization of such systems could provide DOE with an opportunity to significantly impact the field.

Multiple approaches to using X-ray imaging technologies for brain connectivity mapping were discussed. The application of full-field X-ray holographic nano-tomography (XNH) to image millimeter-scale volumes with sub-100-nm resolution (down to ~30 nm) has been demonstrated, enabling reconstruction of dense wiring in mouse nervous tissue. Correlative XNH and EM was further applied to reconstruct hundreds of cortical pyramidal cells. To accelerate neuronal reconstructions, a convolutional neural network was trained to automatically segment neurons from XNH volumes. Thus, XNH could bridge a key gap between light and electron microscopy, providing a new avenue for neural circuit discovery. Scanning X-ray microscopy techniques such as ptychography and X-ray fluorescence (XRF) imaging have demonstrated local 3D imaging with high detection sensitivities, complementing the resolution gap between the optical and electron imaging techniques at the 10~100 nm range. Ptychography can potentially reach a resolution below 10 nm, while XRF microscopy can provide nanoscale resolution images with elemental sensitivity. In general, the application of X-ray imaging techniques to brain connectivity is in early stages and will require the establishment of methods and protocols.

III. Computation and Data

The exascale computers that DOE is deploying in the 2021/2022 timeframe have the compute capability and the storage resources to enable progress on the mouse connectome in any of the major imaging modalities. The current estimate is that the DOE machines planned for 2025 would be capable of storing and processing the data for a full mouse connectome at the microscale (~10 nm) in approximately one year of computing time. The current estimate of mouse image raw data is roughly the scale of the main filesystems on exascale computers (10^{18} bytes), though it is believed that methods may be improved to reduce this by a factor of 10. The human connectome is approximately 1000x larger. Storage capacity and storage throughput is growing more slowly than computing capacity improvements, and current workflows have major “human in the loop” components that would need higher levels of automation to be practical.

Based on current approaches to connectome data processing and reconstruction, it would take on the order of a million exaflop/s days to process the imaging data from a human brain connectome. Doing that in one year (assuming massive progress in imaging that would pace the computing to produce a human connectome in one year) would require 10^4 exascale computing systems. It is expected that by 2030, it will be feasible to build (classical) computing systems that are roughly 100x faster than exascale systems being deployed in the 2022 timeframe (i.e., 100 exaflop/s). That leaves an additional factor of 100x improvement needed to accomplish the computing in one year. There are several paths to accomplishing this factor of 100x. Given the possibilities outlined in this report, it seems likely, with sufficient investment and focus, that a computer system could be developed that would be optimized for human brain connectome computing tasks that would achieve the needed 10,000x over exascale systems being designed today. How much beyond that is not clear. To satisfy the computing requirements for 10 or 100

human connectomes per year would entail sustaining the rate of development of capabilities for another one or two decades beyond the timeframe it becomes feasible for the first human connectome.

In addition to imaging and reconstruction tasks, a major potential use of large-scale computing resources will be analysis and simulation of neuronal circuits produced from connectome efforts. Brain-scale neural network computation is becoming possible on today's largest machines and in the cloud. Recent natural language models based on artificial neural networks, such as GPT-3[1], contain 175 billion parameters (compared to mouse brain of 100 billion synapses). These deep neural network models are considerably simpler and more regular than biological brains, however they are providing use cases for "neural" style computation at scale that is a starting point. Other groups have developed more biologically-oriented neural network computational models and the scale of these simulations is also increasing. Two general scenarios are emerging. The first is the use of DOE computing platforms to model, using computational neuroscience approaches, biological brains. These simulations provide insight into neuronal processes, circuit dynamics, functional dynamics, etc. The second approach is to develop computer hardware based on the ideas from brains. These neuromorphic systems replace conventional CPU designs with simplified models of neurons as their computational base. The most sophisticated of these systems have neuronal hardware that can be parametrized to approximate behavior of biological neurons. Connectomics-based reconstructions of neurons and circuits may also inform future computer hardware concepts, particularly if the underlying principles behind biological neural circuits can be understood. Simulating circuits, integrating function information, and ultimately whole brain level simulations will be needed to gain insights from brain scale datasets.

The following sections summarize the main themes covered during the roundtable. The information is taken from the participants' responses to charge questions posed prior to the event, their presentations during the roundtable, and the various discussion sessions.

I. Connectome Requirements

Key neuronal properties required for reconstructing a whole brain connectome

- *Cell morphology and location* of all neurons, and ideally other cells such as glia, within a given volume, with enough resolution to unambiguously resolve nanoscale compartments and fine structures such as axons, dendrites, synapses, philopodia and dendritic spines.
- *Synapse anatomy* (strength) including features such as the number of vesicles, the size of axonal boutons or presynaptic active zones and whether a single bouton synapses with multiple targets.
- *Molecular identity* to identify cell types, and ideally synapse types, with enough spatial resolution to distinguish where molecules are (e.g., within a dendrite vs. within a synapse or spine). Identity can be also inferred from the synapse morphology and anatomy (i.e., many inhibitory synapses look different from excitatory synapses).
- *Function* determined by live cell imaging of calcium dynamics, and ideally activity-linked voltage fluctuations, of at least a subset of the mapped neurons and circuits, and co-registration of the functional information with the above microanatomical, morphological, and molecular information.

Functional information will greatly increase connectomics value and impact

All participating neuroscientists saw a great value in having functional information included. It was pointed out that scientific value of connectomic data will greatly *increase* if we can register it with cortex-wide imaging of neural activity as these recordings are vital for modeling mammalian brain functions. Such functional characterization will enable testing of connectivity-based models and theories of neural computation. Imaging neural activity approaches are described in detail under II.a. Optical Microscopy and are often done by voltage sensitive dyes or by calcium imaging using genetically encoded calcium indicators (GECIs)[2]. Existing optical modalities for imaging and optogenetic manipulation of brain activity can be readily scaled to ~1 million cortical neurons. Such large-scale functional characterization has been demonstrated, showing that approximately 1 million neurons can be optically accessible in awake mice[3]. Also mentioned was that functional imaging will provide only partial information. Ideally, with future capability, every neuron could be functionally imaged throughout the normal behaving life of the mouse until the day of sacrifice.

Both Microscale (~10 nm resolution) and Mesoscale (~100 nm resolution) connectomes are needed

For reconstructing a microscale (synaptic resolution) connectome, a ~10x10x10 nm voxel resolution over a 1 cm^3 volume of mouse brain is required. Increasing the resolution beyond that would likely increase the efficacy of segmentation algorithms but the data size would increase as well. There are 75-100 million neurons and 100 billion synapses in the mouse brain. Data size (depending on voxel size) will probably be no more than 1 million terabytes (1 exabyte) for a mouse brain. The human brain is 1000 times larger. With current technology, petascale connectomes of cortical columns (1 mm^3) have been generated in one year. It was

estimated that an exascale connectome of mouse brain (1 cm^3) could be generated over 5-10 years, while a zettascale connectome of human brain (10 cm^3) could be generated over 10-20 years.

The initial goal could be a complete microscale connectome of a wild-type mouse brain, during which approaches will be established. This effort could be followed by mapping the connectomes of mice executing specific tasks, or of mice at different developmental stages, among other goals. A suggestion was to focus on mapping the connectome of a nonhuman primate brain before moving to map the connectome of a human brain.

In parallel with the pursuit of a microscale connectome, a complete mesoscale projectome at ~100 nm resolution for the mouse brain (petascale achievable in 1 year) and the human brain (exascale achievable in 5-10 years) could also be pursued. A complete human brain projection map will have immediate applications in clinical neurology. Deep brain stimulation influences brain function predominantly by stimulating axons, even in gray matter. Accurate maps of long-distance axons could increase and change the understanding of existing treatments and inform the design of new ones. Similarly, traumatic brain injury typically damages long-distance connections in the brain. Thus, clinical understanding of brain injuries requires a better understanding of the anatomy of these pathways.

Methods for mesoscale connectomes have been established (i.e., modest expansion with immunostaining). Imaging platforms (i.e., light sheet microscopy) are inexpensive (~$200K). It was estimated that with 100 such microscopes, ~5 petavoxel will be generated per day, achieving a mesoscale connectome of the human brain at 100 nm resolution in 1 year. It has been pointed out that an array of such microscopes, together with significant computation at the edge for quality control, preprocessing, and data compression, along with the resources needed to deploy machine learning at the exascale, could be well within the province of DOE.

How the connectomes will be used

The connectomes will be used to understand the rules of brain organization, including cell type diversity and network properties; uncover structure-function relationship, including connectomes of functionally interrogated brains; and bridge between molecularly defined cell types and connection properties. The connectomes could also be used for neuronal classification using Patch-seq, a multi-modal technique developed to comprehensively profile single neurons that can bridge transcriptomic and morpho-electrical properties.[4] Ultimately, the human connectome will be used to understand disease, learning, and other attributes such as the emergence of intelligence through evolution.

A critical challenge in comparing mammalian brain connectomes is that there are only a few identified individual neurons that can be compared. Therefore, comparing the connectivity of neuronal types is possible, but is also a challenging task. In mammals, the connectome is mainly the product of experience (activity), hence function could be inferred from structure (i.e., neurons that fire together wire together). It has been pointed out that constructing the entire human connectome, however, may always be problematic due to a couple of reasons: 1) death

defined by brain death (i.e., a flat electroencephalogram (EEG)), which means brain tissue is irreparably damaged; and 2) a technique for staining the entire human brain is unknown.

No single imaging modality is sufficient; a combination of imaging modalities will be required

Electron microscopy is currently the only approach that can accurately resolve a synapse, including nanoscale structures such as synaptic vesicles and membranes (<10 nm). However, the Achilles heel of electron microscopy is molecular identities. While advances have been made, electron microscopy currently has a relatively limited ability to distinguish between multiple distinct molecular constituents and the throughput to visualize millimeter-scale dimensions in a reasonable timeframe. In contrast, the current resolution of optical microscopy combined with expansion microscopy can only reach ~40 nm, but the method can clearly distinguish between multiple molecular entities, as well as image neural activity for functional characterization. It was suggested that correlated light and electron microscopy (without harming ultrastructure) could provide ultrastructural signposts for cell and synapse types. The potential of correlated cryogenic imaging techniques to eliminate distortion induced by fixation and embedding has been mentioned[5,6].

X-Ray techniques can image, in principle, mm thick samples with <10-20 nm resolution in cryo-preserved samples. The first application of X-ray holographic nano-tomography to reconstructing dense brain tissue (mm thick at 30-100 nm resolution) was published in 2020[7]. This work showed that X-ray imaging approaches, while promising, will have to be further developed for detecting synapses with the resolution and clarity achieved by electron microscopy. Similar to electron microscopy, X-ray imaging techniques are currently limited in the ability to distinguish molecular identities, although some genetically encoded probes (i.e., lanthanide-binding tags) have been developed. Combining X-ray and electron or light microscopy modalities has been proposed for hierarchical, multiscale imaging. It has been suggested that X-ray imaging could be used for assaying the quality of staining for EM in large brain volumes, essential to know before undergoing a long (i.e., years) imaging project. All imaging modalities, with no exceptions, will require massive parallelization and automation to reach the goal of mapping one mouse brain, let alone one human brain, in a year.

II. Bioimaging Technologies

II.a. Optical Microscopy

Main Opportunity. Can image relatively thick samples (>100 µm) and large volumes (mm^3) in a reasonable time. Can detect cell and molecular identities. Can image activity for functional characterization. Optical systems are relatively inexpensive.

Main Challenge. Resolution is currently limited to 20-40 nm, which is insufficient for imaging nano-scale structures such as membranes or synaptic vesicles.

Approaching EM resolution in thicker tissue samples with molecular selectivity

Optical imaging techniques and fluorescence tagging for detecting cell and molecular identities, as well as functional characterization, have been developed specifically for connectomics. The major approaches are described below.

Expansion microscopy introduces a polymer gel into fixed cells and tissues, and chemically induces swelling of the polymer by almost two orders of magnitude. This approach enables scalable super-resolution microscopy using diffraction-limited confocal microscopy, which is faster and can image thicker samples (>100 µm). The effective resolving power of expansion microscopy is 20-40 nm, where the axial and lateral effective resolutions are improved by the same factor. Although not sufficient for imaging membranes or synaptic vesicles, this resolution is sufficient for detecting pre- and post-synaptic contacts defined by distinct fluorescent colors[8]. Using iterative expansion microscopy to expand biological specimens by ~20x (~4.5 × 4.5), ~25 nm resolution imaging of cells and tissues has been demonstrated by conventional fluorescence microscopes. This approach has been used to visualize synaptic proteins, as well as the detailed architecture of dendritic spines, in mouse brain circuitry[9]. Expansion microscopy has been also combined with long-read untargeted and targeted *in situ* RNA sequencing, yielding nanoscale-resolution maps of RNAs for thousands of genes throughout dendrites and spines in neurons of the mouse hippocampus[10].

Lattice light sheet microscopy provides non-invasive, fast live cell imaging of large volumes. With a resolution of ~200 nm x 200 nm x 320 nm, image acquisition can be done at 300 frames per second. This technique, which uses ultrathin light sheets from two-dimensional optical lattices, enables imaging three-dimensional (3D) dynamics with resolutions spanning four orders of magnitude in space and time[11]. Combined lattice light sheet microscopy and expansion microscopy has been demonstrated to enable fast image acquisition (1200x faster than EM) of Drosophila brain tissue (360 µm x 660 µm x 90 µm volume) at 56 x 56 x 92 nm resolution with 4x expansion. Using this combined approach, the detection of synaptic proteins at dendritic spines, myelination along axons, and presynaptic densities at dopaminergic neurons have been demonstrated in every fly brain region[12].

Brainbow has been developed to address the following need—detailed analysis of neuronal network architecture requires labelling uniquely many individual cells within a population. Brainbow enables the visualization of synaptic circuits by genetically labelling neurons with

multiple, distinct colors[13,14,15]. Brainbow, in combination with expansion microscopy, has been used to image mouse brain circuitry[16]. Brainbow has been further combined with multi-round stripping and immunostaining to simultaneously interrogate morphology, molecular markers, and connectivity in the same brain section using expansion microscopy[17].

Combined techniques for optical microscopy-based connectomics

The combined *Expansion microscopy + Lattice light sheet microscopy + Brainbow* (or other multiplex neuronal labeling) approach has been proposed for connectomics. With current technology, using 8x expansion (effective resolution of 12.5 nm x 12.5 nm x 25 nm) and 5 fluorescent colors (3 for Brainbow and 2 for pre- and post-synaptic markers), estimates are that a whole mouse brain connectome (500 mm^3, image size ~1280 petabyte) could be acquired in ~14 years, and a whole human brain (1,200,000 mm^3) in ~35,000 years. Increasing throughput by increasing the field of view of lattice light sheet microscopy to a larger-format or arrays of spatial light modulators (SLM or phase plates) and cameras could decrease significantly the estimated timeframes. With 10 x 10 arrays, 100x increase in throughput will be achieved (at 100x data rate—not supportable by standard laboratory computer hardware) with estimates to support seven whole mouse brain connectomes per year. With another 100x increase in throughput, it is estimated that one human brain connectome could be achieved in less than five years. The development of brighter probes and faster cameras could also significantly increase throughput. Operation of multiple imaging systems in parallel and automation of sample preparation, as well as investments in data storage and analysis (i.e., registration, stitching, segmentation), will be required.

Approaches for imaging neural activity: functional characterization

The techniques described below use relatively inexpensive optical imaging systems that mostly have been developed in individual labs. Large-scale parallelization of such systems could provide DOE with an opportunity to significantly impact the field with minimal investment. As mentioned earlier, the community agreed that cortex-wide recordings of neural activity could significantly increase the value and impact of the connectomes.

Adaptive optics has been applied to accurately measure bouton activity deep in the cortex, determining the response properties of ~28,000 thalamic boutons and ~4,000 cortical neurons in layers 1-5 of the primary visual area (V1) of an awake mouse.[18]

Bessel focus module incorporated into a two-photon fluorescence mesoscope, the imaging of neurons and neural circuits over large volumes (2.7 mm^3) at high speed and subcellular resolution, has been demonstrated. Using this approach, rapid mesoscale volumetric imaging of neural activity with synaptic resolution has been achieved. The technology has been applied to calcium imaging of entire dendritic spans of neurons as well as neural ensembles within multiple cortical regions over two hemispheres of the awake mouse brain[19]. Using this approach, it is estimated that ~100,000 neurons within 1 mm^3 cortical column could be imaged at 1 Hz, generating 100 MB per volume. With 1 volume per second, 100 MB per second could be generated. Over 100,000 sec (27.8 hours) recording of the same neurons will generate 10

TB. Characterizing the activity of all neocortical neurons (~100 mm^3 volume) could then generate 1 petabyte.

Simultaneous recordings from thousands of cortical neurons with shared sensory inputs have been achieved using a *16-beam, two-photon microscope*, developed in the lab to monitor activity across the mouse primary visual cortex. Quantifying the information conveyed by large neural ensembles is critical for understanding, for example, the role of correlated noise in coding fidelity[20].

By creating a curved glass replacement to the dorsal cranium (a *'Crystal Skull'*), long-term optical access to 30-40 neocortical brain areas, or an estimated 800,000–1,100,000 individual neurons across the dorsal surface of neocortex, has been achieved in behaving mice[3]. Using this approach, cellular- and sub-cellular-level resolution of neural morphology across the cortex, and large-scale imaging of neural Ca^{2+} dynamics across the cortex, have been achieved in active mice. This approach has been applied using epifluorescence, as well as two-photon microscopy, revealing dendritic morphologies throughout the neocortex. It allowed time-lapse imaging of individual cells, and yielded estimates of >1 million accessible neurons per mouse by serial tiling.

II.b. Electron Microscopy & Tomography

Main Opportunity. The only technology that currently can resolve the synapse nanostructure such as synaptic vesicles and membranes.

Main Challenge. Currently limited to thin sections and requires massive registration, stitching, and segmentation.

Electron microscopy and tomography approaches for connectomics

Neural tissue is generally prepared using conventional methods of aldehyde fixation, heavy metal infiltration, and either serial transmission EM (TEM) or scanning EM (SEM) with serial blockface, or slice-and-view Focused Ion Beam (FIB). Array tomography was further developed as a correlative light and electron microscopy technique that produces large volume sets of 3D data using immuno-fluorescence labeling, followed by backscatter electron detection (BSD)-SEM imaging of the same section arrays. This approach adds contextual information to the physiological dataset from fluorescence microscopy, highlighting function in addition to structure. Using these techniques, the entire connectome of various model organisms has been determined[21]. All these techniques are limited to ~10 nm resolution, and the likelihood of deviation from the native structure caused by sample preparations remains a concern.

A key challenge in EM is the tradeoff between section thickness versus z-resolution, which undermines reliability. Several EM-connectomics techniques involve sectioning brain tissue into ultrathin (30-50 nm) sections and acquiring one tiled image of each section with either a multibeam scanning electron microscope (MultiSEM)[22] or TEM[23]. The cutting and collection of such ultrathin sections becomes notoriously unreliable as section thickness drops below 50 nm

(becoming especially difficult for wide tissue blocks as would be required for whole brains). The z-resolution is still barely sufficient to trace neural circuitry and increases the chances of accumulating unacceptable numbers of tracing errors. While acceptable for small-scale connectomics, such a tradeoff will be unacceptable for whole mouse and human brains. Due to the tiny diameters of spine necks and long-range axons, as well as the overall density of neuropil, isotropic image acquisition with approximately 10 x 10 x 10 nm voxels is preferred for reliable automated tracing.

The blockface ion milling approach FIB-SEM[24] is able to achieve 10 nm isotropic resolution and has been used to produce synapse-level connectome of the fly central brain[21]. However, this approach is currently incompatible with the fast, parallel imaging needed for whole brains. FIB milling is too slow, and its mill area is too small for mapping larger brains. A *hybrid thick sectioning + broad ion milling approach called GCIB-SEM* (Gas Cluster Ion Beam SEM)[25] has been developed specifically for whole brain connectome mapping. GCIB-SEM allows for reliable sectioning because it can work with sections from 100 nm thick to 10 microns thick. Thick sections are collected onto sturdy silicon wafers and are put through sequential rounds of broad ion milling and MultiSEM surface imaging resulting in a 10 x 10 x 10 nm voxel volume dataset of each thick section. Volume datasets of sequential thick sections are then stitched together into a single dataset covering the original tissue volume and suitable for automated connectome tracing.

A *GCIB-SEM's broad ion milling + MultiSEM* approach has been proposed to have what it takes to scale all the way up to mouse and human brains. By collecting thick sections (100 nm to 10 microns) on sturdy silicon wafers, GCIB-SEM should be able to achieve the 100% reliability required even when wide blocks and frequent knife changes are considered. GCIB-SEM can be implemented in a massively parallel fashion by passing wafers among potentially thousands of MultiSEM and broad ion milling machines in a manner similar to a semiconductor manufacturing plant. All machines are interchangeable, so machine failure is not an issue. Voxel resolution and signal-to-noise can be optimized for reliable tracing without impacting reliability of cutting. An estimated 20 MultiSEMs will be required for reconstructing one mouse brain, and 3000 for reconstructing one human brain, in 5 years. Developments in the MultiSEM optical column and redesign for mass-production are likely to increase throughput and decrease cost.

Cryo-electron tomography (cryoET) can resolve subcellular structures *in situ* without chemical fixation, embedding in plastic, or staining[26]. Correlated cryo-fluorescence light microscopy (cryoFLM) and cryoET of frozen, hydrated cells has been used to localize targeted molecules with fluorescence tags in the 3D context of cells and tissues at a higher resolution than any other imaging technique[27]. A current limitation is the penetration power of 300 keV electron microscopes, which restricts specimen thickness to ~500 nm, much thinner than typical eukaryotic cells. A focused ion beam scanning electron microscopy equipped with a cryo-specimen stage (cryoFIB-SEM) can be used both to produce consecutive SEM images after each step of milling (Mill and View), and to process the specimen further to produce thin lamellae (~100-300 nm) for subsequent transfer to a TEM for cryoET[28]. The Mill and View images can have sufficient resolution to visualize subcellular features to better than 10 nm, while the

cryoET of selected regions can resolve to sub-nanometer resolution with sub-volume classification and averaging.

A *correlated workflow,* including consecutive steps of cryoFLM, cryoFIB-SEM, Mill and View and cryoET, has been proposed to operate on the same frozen, hydrated specimen[6]. However, this approach has not been applied in its entirety to any organism and thus it is difficult to predict its applicability to the connectomics or make realistic estimates of the needed resources. This approach could potentially provide an unprecedented resolvability of membrane bound organelles, cytoskeleton filaments, protruding membrane proteins, and large macromolecular assemblies in the neurons. The entire cell, from cell body to the tip of the neurites in a frozen, hydrated state could potentially be imaged and analyzed at nanometer resolution without chemical fixation, staining or drying. With no demonstrated application, it would be difficult to determine how information acquired by this correlated approach will complement similar information from the current practice with plastic embedded, chemically fixed neuron samples done with either SEM or TEM at room temperature.

Two *3D EM methods,* "multi-tilt" TEM tomography[29],[30] and serial block SEM (using an ultramicrotome), have been compared. Comparing a TEM image of an ultrathin section (~60 nm ultramicrotome section) with block-face backscatter-electron SEM image showed a better image by TEM, although it also demonstrated the need to mitigate section distortion and compression in the TEM image for reliable serial section reconstructions. To improve the accuracy and efficiency of 3D EM, such as the diamond knife-based serial block-face SEM, X-ray microscopy (XRM) has been developed to provide nondestructive imaging of the samples to pinpoint specific regions of interest. Given the excellent contrast and resolution that can be achieved with XRM when specimens are stained with protocols compatible with EM, it has been suggested that XRM can dramatically enhance the value of EM data, either by revealing how the EM data fits into a larger context and/or by improving the ability to perform correlated light microscopy and EM imaging[31].

The application of energy filtering automated most-probable loss (MPL) tomography to intermediate voltage electron microscopy (IVEM) has been described. For thick, selectively stained biological specimens, this method has been shown to produce a dramatic increase in resolution of the projections and the computed volumes compared with standard unfiltered TEM. This improvement is particularly evident at the large tilt angles required to improve tomographic resolution in the z-direction. The method dramatically improves resolution compared with unfiltered TEM and expands the utility of IVEM to specimens up to 3 µm thick[32].

Tags for imaging molecular identities by EM are described here. As mentioned earlier, unambiguous identification of specific molecules in electron micrographs presents a challenge. To address this challenge, genetic tags that generate electron-dense contrasts on a specific protein or subcellular compartment have been developed (i.e., Mini-Single Oxygen Generator (SOG), Split Mini-SOG, Split Horseradish Peroxidase (HRP), APEX2)[33]. Another approach for simultaneous visualization of multiple molecular species by EM uses sequential, localized deposition of different lanthanides by photosensitizers, small molecule probes, or peroxidases.

A detailed view of biological structures is then created by overlaying conventional electron micrographs with pseudocolor lanthanide elemental maps, derived from distinctive electron energy-loss spectra (EELS) of each lanthanide deposit via in-column energy-filtered transmission electron microscope (EFTEM). This results in multicolor EM images analogous to multicolor fluorescence but with the benefit of the full spatial resolution of EM. The power of the multicolor EM approach has been demonstrated by visualizing hippocampal astrocytes to show that processes from two astrocytes can share a single synapse[34].

Sample preparation

Staining. Current state of the art in sample preparation includes Osmium for staining blocks that are ~1 mm^3. Active research into staining substantially larger volumes is being pursued and several promising approaches are being developed for both 10 mm^3 blocks via immersion and whole brains via perfusion (i.e., no cracks and extracellular space retention). The need for heavy metal staining of an entire human brain is perhaps *the biggest technical hurdle* to obtaining a human connectome. One promising route to overcome this hurdle would be to vibratome the fixed brain into 1 mm slabs prior to staining and plastic embedding. Doing so would mean that existing EM staining protocols[35] could be applied. However, such near lossless vibratomy has never been demonstrated but is theoretically possible if the brain is sufficiently fixed and custom diamond knives are developed.

Sectioning. Diamond knives that can cut larger sections would need to be developed. One idea has been to develop custom diamond knives made of thin CVD diamond wafers whose edges are sharpened by broad ion milling. Optimal imaging volume is thought to be voxels of 500 nm^3 (4 x 4 x 30 nm), providing traceable processes of almost all objects (>99%). Sections thicker than 30 nm become progressively harder to disambiguate the finest processes especially when they run parallel to the plane of imaging. Such 4 x 4 nm lateral resolution represents Nyquist sampling for an 8 nm osmium stained membrane.

Approaches to increase throughput

Parallelization is key. It has been proposed that parallelization could potentially be achieved by dividing the tissue into manageable smaller volumes. Throughput is then proportional to the number of devices running simultaneously. If one microscope can acquire 40 TB of image data per day, then 25 microscopes can generate 1 PB per day and thus 1 EB (one mouse brain) in 2 years 9 months. To obtain a mouse connectome in a year would require generating ~2 PB per day, which would require ~40 high speed imaging devices running continuously. To obtain a human brain connectome in the same time scale, this would be multiplied by ~1000.

There is a need for developing advanced optics and large-scale and sensitive imaging detectors to increase throughput, as well multispectral detection approaches. There is also a need for developing new labeling approaches for detecting and mapping molecular specificities by EM. The development of correlated light and electron microscopy in one system is key, including for cryo-based approaches.

There is a need for automation of sample preparation, sample loading and instrument operation, and techniques for sample freezing and trimming or dividing to suitable sizes. There is a need to develop artificial intelligence (AI) to guide data collection and perform unsupervised AI annotation, improve image reconstruction to optimize map accuracy, and develop approaches for integrating data acquired by different systems (i.e., cryo-light microscopy, mill-and-view FIB-SEM and cryoET). Finally, annotated data for public sharing would be essential.

II.c. X-Ray Microscopy & Tomography

Main Opportunity. Nondestructive, high penetration power for imaging large volumes (mm^3). High spatial resolution in 3D (<10-30 nm). Can image hydrated, cryopreserved tissue with minimal sample preparation.

Main Challenge. Has yet to demonstrate the ability to resolve synapses. Approaches and protocols for imaging brain tissue are yet to be developed.

X-ray microscopy and tomography approaches for connectomics

The advantage of *X-ray imaging modalities* is their nondestructive, high penetration power combined with high spatial resolution in 3D, as well as minimal sample processing due to the cryogenic environment. Depending on the approach, the resolution can reach ~10 nm. About 20 nm resolution can be reached with full-field X-ray imaging via transmission X-ray microscope (TXM) or projection microscope in phase contrast holography mode, which is thought to be advantageous for thicker samples and connectomics. About 10 nm resolution can be reached using scanning X-ray imaging (via ptychography). The detection limit is set by the X-ray cross section for low Z elements in a brain and radiation damage caused by the X-rays. Considering throughput and dose-dependent sample degradation, an optimal resolution range for high-speed X-ray imaging is in the range of 50-100 nm through the full-field mode.

The application of *full-field X-ray holographic nano-tomography* (XNH) to image millimeter-scale volumes with sub-100-nm resolution (down to ~30 nm) has been demonstrated[7], enabling reconstruction of dense wiring in mouse nervous tissue. Correlative XNH and EM was further applied to reconstruct hundreds of cortical pyramidal cells. To accelerate neuronal reconstructions, a convolutional neural network was trained to automatically segment neurons from XNH volumes. Thus, XNH can bridge a key gap between light and electron microscopy, providing a new avenue for neural circuit discovery.

Since full-field X-ray imaging is nondestructive, the same sample can be further processed and imaged by other modalities, including high resolution X-ray imaging. Even with the high-throughput of full-field imaging, the total scan time of a mouse brain at the highest possible resolution is still prohibitively long. Given its nondestructive detection nature, *full-field X-ray imaging* can provide a hierarchical multiresolution (from micrometer to tens of nanometer) imaging complementary to other approaches that can provide the ultimate resolution for micro-connectome mapping of a whole mouse or human brain, as well as providing 3D

datasets, required for verifying co-registration of multimodal imaging methods. X-ray tomography approaches are excellent for assaying the quality of staining in large brain volumes, essential to know before undergoing a long (i.e., years) imaging project. In principle, bright X-ray sources can also be used to generate high-resolution images without physical sectioning. However, the state of the art cannot yet detect synapses.

Full-field transmission *soft x-ray tomography* was discussed in detail. It provides 3D views of whole, hydrated cells in their native state under cryo-preserved conditions with natural contrast and no need for staining. The maximum penetration depth is currently ~20 µm and the isotropic resolution is 35-50 nm. Optics and algorithms remain limiting factors but with better optics, at least 10 nm resolution is feasible. A fully operational correlated cryo-fluorescence and soft X-ray tomography system has been established[36].

Scanning x-ray microscopy techniques such as ptychography and X-ray fluorescence imaging have demonstrated local 3D imaging with high detection sensitivities, complementing the resolution gap between the optical and electron imaging techniques at the 10-100 nm range. Small-angle X-ray scattering tensor tomography (SAXS-TT) has been recently demonstrated to provide local myelin's nano-structural periodicity and orientation in whole mouse brain[37].

Lensless imaging (Ptychography) was discussed in more detail. The spatial resolution using this technique is not limited by optics. A resolution of <8 nm has been demonstrated on hard samples and <16 nm on soft samples (frozen-hydrated). Relaxing the resolution by 2x will speed up imaging by 10x. Staining is not required, but significantly improves contrast. The expected resolution limit for unstained, frozen-hydrated tissue is 10 nm. There is a tradeoff between energy and sample thickness. Higher energy can reduce required sectioning, but it has fewer photons and makes imaging more difficult (slower data acquisition). There will be a need to develop embedding approaches (i.e., resin), as well as protocols for frozen-hydrated samples, especially for thick sections.

X-ray fluorescence (XRF) microscopy can provide nanoscale resolution images with elemental sensitivity. The mapping of natural ions in the brain, e.g., potassium, calcium, iron, copper, and zinc, could provide chemical information on the health and activity of neurons probed. Moreover, genetically-encoded lanthanide binding tags have recently been applied to X-ray imaging for visualizing individual proteins in cells with <20 nm resolution[38]. For nanoscale XRF tomography, samples could be ~1 mm^3. XRF microscopy can readily be combined *simultaneously* with other imaging modalities such as epifluorescence light microscopy (to visualize, for example, GFP-tagged proteins), ptychography (for tissue morphology), and micro-SAXS (for myelin mapping). Since XRF microscopy uses raster-scanning, improvements in throughput are necessary. Current pixel dwell times are in the hundreds of microseconds, which could be improved by a factor of 100x or more. Better fluorescence detectors and brighter synchrotron beams will also increase throughput.

Sample preparation

Staining and Cooling. Although it is possible to image certain structures in unstained brain tissues, it has been indicated that staining with heavy metals is necessary to resolve fine structures using hard X-ray imaging and further EM imaging. Heavy metal staining is currently limited to around 1 mm^2 cross section. Thus, large samples are cut into small blocks and images are then registered. This requirement is magnified by several folds for EM where the imaging thickness is limited to about 30 nm. Cryogenic cooling to protect the samples from X-ray radiation damage is necessary in the sample scans. Both uniform staining and cryogenic cooling of large volume tissues are still challenging.

Sample thickness. There is a tradeoff based on the chosen X-ray energy. At lower energies, data acquisition is faster, but penetration is more limited, requiring the development of sectioning techniques. With an incident energy of 10 keV, sections can be effectively ~1 mm thick. With 25 keV, sections can be ~1 cm thick. However, as incident energy increases, coherent X-ray flux detector efficiencies decrease, making imaging more difficult and slower. Some of this loss can be recovered by noise tolerant reconstruction algorithms.

Approaches to increase resolution and throughput

With current technologies, at 50-100 nm resolution (with 30 nm pixel size), the time for imaging a mouse brain may take decades. However, the throughput can be improved by more than two orders of magnitudes with the fourth generation light sources currently in development and new detector technologies with higher speed and sensitivity, and larger frame size. In addition, multiple dedicated beamlines at DOE's light sources would need to be operating in parallel. The data to map a mouse brain is estimated to exceed 10 TB at 0.5^3 μm^3 voxel size, and 30 PB at 30^3 nm^3 voxel size. The amount of data for a human brain will be 1000 times larger. Image registration between different modalities at different scales is key to the multimodal multiscale approach.

The throughput of full-field microscopes could be significantly improved with better detectors, more efficient objective lenses for TXM (e.g., Fresnel zone plate whose efficiency can still increase by a factor >30), or the ability to use a TXM with an achromatic objective lens in order to use a larger bandwidth of the incoming X-ray beam. Forthcoming fourth generation storage ring light sources providing X-ray energies above 4 keV will exceed the capabilities of today's storage rings by 2 to 3 orders of magnitude in brightness, coherent flux, and nano-focused flux. The coherent flux will greatly enhance the sensitivity of the projection microscopes. Using cryo-stages compatible with computed tomography would enable increasing the power density tolerated by samples and therefore the acquisition speed.

To increase throughput in full-field soft X-ray tomography, better optics and algorithms for high-resolution optics (which have a very shallow depth of focus) will be needed for thick specimens. As for other cryogenic approaches, methods for obtaining thick brain sections (up to 20 µm) of cryo-fixed specimens, and for imaging cryo-fixed and embedded thick sections (up to 20 µm) of fully hydrated specimens (using water tolerant materials that slice well and are

radiation hard) will be required. Algorithms to tile individual images of large fields of view and align serial sections, as well as overlay correlative fluorescence and X-ray tomography of large fields of view in 3D will be required. A new undulator beamline will be needed to add spectral imaging capabilities to support imaging specific elements and heavy metal labels, as well as the algorithms to overlay spectral images on X-ray images. Robotics will be needed to automate data collection.

Lensless imaging (Ptychography) will gain from the development of fourth generation light sources. These advances will lead to an increase of two orders of magnitude imaging speed and provide the photons for imaging ~1 mm^3 in half a day at 10 nm isotropic resolution. Further improvements will require advanced detectors with much faster readout rates (3kHz -> 100kHz-1MHz). Current detectors for X-ray ptychography are typically limited in their sustained readout rate at full dynamic range to about 3kHz with single photon sensitivity. In order to maximize throughput, an ideal detector would have readout rates on the MHz level, while maintaining single photon sensitivity, and >1k x 1k sensors. In order to fully utilize higher energy x-rays (i.e., 30 or 50 keV instead of 10 keV), the sensor would need to be made out of a high-Z material (i.e., GaAs or CdTe). For one mouse brain per year scenario, a detector with 1k x 1k pixels might suffice. For one human brain per year scenario, a significantly increased detector size would be required, perhaps with modestly increased readout rates. Additional developments will be required in detection and imaging algorithms to overcome limitations in depth of field (to reconstruct the full 3D volume 'in focus') and motion blur during fast data acquisition, as well as to reduce the number of photons required to image a structure at a given resolution, including advanced iterative reconstruction techniques as well as machine learning.

III. Computation and Data

Main Opportunities. Exascale computers that DOE is deploying in the 2021/2022 timeframe have the compute capability and the storage resources to enable progress on the mouse connectome in any of the major imaging modalities. The current estimate is that the DOE machines planned for 2025 would be capable of storing and processing a full mouse connectome.

Major Challenges. The current estimate of mouse image raw data is roughly the scale of the main filesystems on Exascale computers (10^{18} bytes). The human connectome is approximately 1000x larger. Storage capacity and storage throughput is growing more slowly than computing capacity improvements. The current workflows have major human in the loop components that would need higher levels of automation to be practical.

Data storage needed for capturing image data and reconstruction workflows

Based on the roundtable participants' estimates, a synaptic resolving resolution of a mouse brain is approximately an exavoxel, and a human brain is approximately zettavoxel. The planned filesystems for the DOE exascale machines are approximately exabyte scale. With advances in lossless compression and image encoding, it is believed that these datasets could be encoded with order a byte per voxel or less, making it feasible in the near term for exascale machines to support the mouse connectome. Dedicated storage systems may be needed since the exascale machines are shared resources. Exabyte storage systems could be acquired for on the order of $50M. Current I/O rates are adequate for the mouse brain reconstruction. The compound annual growth rate for storage system capacity is currently 20% for hard disks and 30%+ for solid state devices. At these rates of growth, the zettavoxel (10^{21}) scale storage needed for capturing a whole human brain would not be possible at reasonable cost for another 15-20 years. However, it might be possible to do dynamic reconstruction from raw imaging data without the need for storing all of it by processing regions on the fly and storing only the reconstructed surfaces. This could potentially reduce data scale from order n^3 to something less, perhaps not quite n^2.

Algorithms and workflows for segmentation, registration, stitching

Current workflows from images to neuron identification and reconstruction go through steps of montage/stitching, alignment, segmentation, and reconstruction. Currently, each stage has considerable human inspection and correction. While there have been advances in the past decade, such as the flood filling networks for segmentation and improvements in conformal mapping for alignment, there remains a need for full end-to-end automation. Computation throughput of the pipeline is estimated at about 1 exaflop/s day per mm^3 based on current work on the mouse. Using these estimates and the expectations from DOE's computing roadmap, by 2030 it may be possible to reconstruct the mouse connectome in a few weeks of computing (100 exaflop/s systems and 1 exaflop/s day per mm^3). A major opportunity is the potential use of surrogate machine learning-based methods for accelerating the pipeline. While this has not be demonstrated in connectome reconstruction yet, in other fields[49], speedups of 1000x are common, and in some cases, over 50,000x have been achieved where physics-based

models are replaced with deep learning emulators. Current efforts developing connectome reconstruction pipelines[50] can provide training and validation datasets that will be extremely useful for development of future machine learning accelerated approaches to the connectome reconstruction problem.

Methods for automating 3D reconstruction of neural circuits

Research is ongoing to develop methods for automating the task of reconstructing neural circuits from annotated and segmented images. Once initial 3D traces have been generated, a proofreading step is needed. Proofreading has been called the major bottleneck in neuronal circuit reconstruction. Manual proofreading is not scalable (500K human hours estimated for proofreading a 1 mm^3 volume) and so methods are needed to leverage sparse manually proofread samples to produce machine learning-based automated proofreading able to scale to full brains. In addition to using machine learning methods to automate this process, there is work underway[39] to use virtual reality (VR) and advanced visualization methods to enable the accurate tracing of long-distance connections over distances of hundreds of voxels. By coupling VR methods with AI, active learning and uncertainty-based sampling can be used to choose those processes most informative for manual proofreading. The machine working together with the human curator could then improve the datasets for training. Hierarchical methods[40] that use "mesoscale" reconstructions to constrain the dense circuit reconstruction might offer some performance improvements. This approach might also leverage multimodal imaging if that becomes feasible, in particular using X-ray reconstruction-based process tracing to accelerate EM image-based processing. Finally, it is important to build a catalog of reconstructed neuronal circuits and make these available to the broader research community. There are efforts to extract neurons from reconstructions, to cluster and sort these into types based on morphology and topology, and to produce more abstract representations of circuits that can be the basis for functional modeling.

Computing hardware needed beyond the mouse connectome

Based on current approaches to data processing and reconstruction, it would take on order of a million exaflop/s days to process the imaging data from a human brain connectome. Doing that in one year (assuming massive progress in imaging that would pace the computing to produce a human connectome in one year) would require 10^4 exascale computing systems. It is expected that by 2030, it will be feasible to build classical computing systems that are roughly 100x faster than exascale systems being deployed in the 2022 timeframe (i.e., 100 exaflop/s). That leaves an additional factor of 100x improvement needed to accomplish the computing in one year. There are several paths to accomplishing this factor of 100x. Mixed precision (the use of multiple word sizes such as FP32 and FP16 at the same time) on next generation systems has been shown to accelerate some problems by factors of 10x. If processing and reconstruction methods are relatively stable, then applications specific integrated circuits (ASICs) could be developed to move those methods into hardware. Special hardware accelerators could probably achieve a factor of 10x-100x or more in throughput for the connectome problem. Earlier versions of these approaches could be explored with field programmable gate arrays

(FPGAs) that have demonstrated > 10x on many problems. In some cases, 100x or more acceleration is possible.

Another approach would be to explore machine learning-based surrogates as kernel accelerators. As previously mentioned in many areas[41-44], these approaches demonstrated 1000x-100,000x on a range of scientific applications. It is not known if surrogates could provide a path for accelerating the data processing for connectome. To the degree that machine learning is expected to play a large role in the future pipeline (e.g., segmentation and proofreading), the opportunity might be lower. New computer architectures that have some ability to do computing in memory, also called processor in memory (PIM) systems, could accelerate some aspects of the processing pipeline by doing as many local processing steps as possible closer to the data. In other applications domains, speedups of up to 100x have been achieved compared to systems with conventional memory[45; 46]. Given all of these possibilities, it seems likely, with sufficient investment and focus, that a computer system could be developed that would be optimized for human brain connectome computing tasks that would achieve the needed 10,000x over Exascale systems being designed today. How much beyond that is not clear. To satisfy the computing requirements for 10 or 100 human connectomes per year would entail sustaining the rate of development of capabilities for another one or two decades beyond the timeframe it becomes feasible for the first human connectome.

Analysis of connectomes and brain/circuit simulation

A major potential use of large-scale computing resources will be analysis and simulation of neuronal circuits produced from connectome efforts. Brain-scale neural network computation is becoming possible on today's largest machine and in the cloud. Recent natural language models based on artificial neural networks, such as GPT-3,[1] contain 175 billion parameters (compared to a mouse brain of 100 billion synapses). These deep neural network models are considerably simpler and more regular than biological brains, however they are providing use cases for "neural" style computation at scale that is a starting point. Other groups have developed more biologically-oriented neural network computational models and the scale of these simulations is also increasing.

Two general scenarios are emerging. The first is the use of DOE computing platforms to model, using computational neuroscience approaches, biological brains. These simulations provide insight to neuronal processes, circuit dynamics, functional dynamics, etc. To date, most of these simulations are based on estimations of brain circuits by using hierarchical models[47] that are parameterized to fit existing datasets. As experimental connectomics advances, simulations such as these will be increasingly constrained by experimental data. These modeling efforts will provide useful feedback to the structural and functional connectomics communities. The second approach is to develop computer architecture hardware based on the ideas from brains[48]. These neuromorphic systems replace conventional CPU designs with simplified models of neurons as their computational base. The most sophisticated of these systems have neuronal hardware that can be parametrized to approximate behavior of biological neurons. At sufficient scale, some neuromorphic systems can be used to test hypotheses derived from experimental data from connectomes[48]. Connectomics-based reconstructions of neurons and circuits may

also inform future computer hardware concepts, particularly if the underlying principles behind biological neural circuits can be understood. Simulating circuits, integrating function information, and ultimately whole brain level simulations will be needed to gain insights from brain scale datasets.

DOE Opportunities

Computation and Data

DOE is the only agency that combines state-of-the-art experimental facilities for high-throughput imaging and supercomputing at the same physical sites. This makes it possible to connect imaging systems and computing at multi-terabit bandwidths in the near term which is not cost effective over wide area networks. For example, by the end of the decade, Argonne National Laboratory (ANL) and Lawrence Berkeley National Laboratory (LBNL) will have upgraded synchrotron light-sources and electron microscopes on the same campus and on the same local networks as exascale-class supercomputers. Pacific Northwest National Laboratory (PNNL) has electron and fluorescence microscopes integrated with petascale supercomputing on the same local network in the same user facility building. Brookhaven National Laboratory (BNL) has a synchrotron light source, cryoEM facility, and petascale supercomputing close together. Oak Ridge National Laboratory (ORNL) has a neutron source, electron microscopes and fluorescence microscopes on the same campus as DOE's flagship supercomputer, Summit, and will have those capabilities on the same campus as DOE's first exascale supercomputer, Frontier. SLAC National Accelerator Laboratory (SLAC) has both a unique x-ray laser and third generation synchrotron light-source, as well as cryoEM facilities co-located on campus. The suite of imaging facilities at these multiprogram DOE laboratories are continually improving in support of the DOE mission. All of the DOE Office of Science (SC) laboratories have something to contribute to brain imaging challenges. DOE also operates ESnet, which is on track to provide terabit/s level connectivity between DOE sites. This type of bandwidth will be needed to implement a distributed connectome reconstruction and collaborations across the DOE laboratories and universities.

DOE could pioneer the physical integration of high-bandwidth data collection and imaging for connectomics with the real-time or near real-time processing of data at scale. Ideally, there would be multiple efforts to prototype the streaming pipelines that could ultimately provide real-time feedback on imaging quality, allowing adaptive sampling or re-imaging as needed to ensure high-quality reconstructions. DOE is uniquely capable of developing these pipelines at scale and to engineer the infrastructure needed to run them 24/7 for months or years.

DOE is well positioned to drive the development of new high-performance data storage technologies, on-the-fly fully automated quality control, advanced data management techniques, and novel data compression methods needed to host exabyte and zettabyte datasets that are the heart of any connectome program. Without progress in data storage and data management techniques, the practical feasibility of obtaining a high-quality mouse connectome (exascale datasets) is likely to take ten years. A human connectome (zettascale datasets) would likely not be feasible before 2040.

DOE is the only agency deploying Exascale systems in the next few years and the only agency positioned to continue the deployment of leadership-class computing over the next decade. DOE's expertise and computing roadmap position it to enable creation of the end-to-end software pipelines needed to address the neuroscience goals of full mouse and human

connectomes. These would likely be collaborative projects, with DOE laboratories collaborating with each other and with extramural NIH investigators as well as investigators from private research groups (e.g., Allen Institute, Janelia Research Campus, etc.) and universities. The goal would be to develop, debug, test, and scale-up complete computational pipelines that support stitching, segmentation, reconstruction, proofreading, and validation of connectomes at increasing scales and effectiveness over time. The development of two or three such pipelines would drive diversity of methods and diversity of imaging modalities. At least one of these pipelines could explore the strong use of AI methods for acceleration and improved reconstruction quality. Versions could also support multi-modal reconstruction and annotation.

DOE is well positioned to consider connectome requirements when developing its future supercomputer specifications for machines to be deployed in the 2025 and 2030 timeframes. DOE has a strong track record of driving computer architecture design to support DOE mission needs and the applications requirements for those missions. Factoring connectomics computational requirements into future procurements would need to be concomitant with a long-term investment in the algorithms and pipelines so that co-design principles could be used to optimize hardware and software over time. Co-design has been successfully used in DOE domains such as nuclear energy modeling and simulation, AI and data science, and many applications areas such as climate modeling and materials science as part of the Exascale applications development program.

DOE would be ideally positioned to drive the development of custom applications-specific accelerators for the connectome reconstruction pipeline once those pipelines have been validated on smaller brains. Reconstruction pipelines would be developed and optimized on mouse and other smaller brains and the most promising methods could drive the development of application-specific integrated circuits (ASICs) that could potentially accelerate the pipelines by one or two or more orders of magnitude over the more general-purpose supercomputers.

Bioimaging Technologies

DOE's network of Scientific National User Facilities presents a unique opportunity for a well-coordinated Distributed "Facility". In light of the strong recommendations for multimodal imaging approaches, a distributed, rather than centralized, "facility" needs to be one of the options explored to enable the connectomics effort. A distributed facility would require high-bandwidth interconnects to move large-scale datasets. ESnet has experience with supporting the world's largest scientific collaborations in high-energy physics and this experience could be leveraged to develop the requirements for a distributed connectome facility. Some of the DOE's User Facilities currently include multimodal bioimaging technologies integrated with advanced computing for real time data processing, including the infrastructure for multi-institutional project coordination and data sharing with the broader community. In addition, some User Facilities also include technologies indicated as complementary or supportive of the connectomics effort (i.e., multi-omics technologies under the same roof with microscopy and computing such as those found at PNNL's Environmental Molecular Sciences Laboratory). Such facilities are well positioned to make advances in massive parallelization of imaging technologies as described below.

All current imaging modalities would require massive parallelization and automation at scales beyond those likely to be supported by other agencies, presenting a unique opportunity for DOE. While tens of light and/or electron microscopy systems working in parallel will be required to reconstruct a microscale connectome of a mouse brain, a few thousand of such systems will be required to reconstruct a microscale connectome of a human brain within about 5 years (details in the Summary and below). DOE strengths in imaging technology development, as well as current support for the development of automation for scientific discovery, could be leveraged to increase throughput of current technologies (examples in the Summary and below) and to develop the large-scale automation needed throughout the whole pipeline.

One of the roundtable main conclusions has been the need for a combination of imaging modalities, presenting an opportunity for technology development. While a limited number of studies using combined modalities have been demonstrated, these have been mostly done sequentially, rather than using one correlative system. Such integrated systems will be required for the large scale connectomics project. An opportunity for DOE would be the co-locating and integrating of microscopies into correlative systems (i.e., EM and light microscopy, X-ray and EM microscopy, X-ray and light microscopy), as well as the development of the computational tools necessary for correlative image analyses.

Specific opportunities for DOE under light microscopy-based connectomics. The lattice light sheet microscope, in combination with expansion microscopy and multiplex labeling (i.e., Brainbow), has been a major light microscopy technology discussed for connectomics. It has been estimated that a *mesoscale* (~100 nm resolution) human brain projectome could be achieved in 1 year using this approach, requiring about 100 light sheet microscopes running in parallel. A few thousand microscopes (at current technology) would be required to reconstruct a human brain with a resolution approaching *microscale* connectome (~40 nm with expansion) in a reasonable timeframe. Although light microscopes are relatively inexpensive compared with electron microscopes (EM), such large-scale parallelization presents a unique opportunity for DOE. In addition to parallelization and automation, an opportunity for DOE would be to develop the technology to increase throughput of the lattice light sheet microscope (i.e., increasing the field-of-view and the size, speed, and sensitivity of the cameras).

Opportunity for large-scale parallelization for functional imaging. Roundtable participants strongly recommended that functional information using cortex-wide imaging of neural activity be included in a program to develop a structural connectome of the mouse brain. Although several optical microscopy techniques have been established for recording neural activity, a large-scale parallelization of such systems to enable simultaneous recording of millions of cortical neurons (over multiple timepoints), including automation, would be an opportunity for DOE. Furthermore, the development of computational tools for registering neural activity and structural images at large scales will be required, presenting an additional opportunity to tap into DOE's strengths.

Specific opportunities for DOE under electron microscopy-based connectomics. The throughput in EM technologies is ultimately limited by the need to image thin tissue sections and thus, a few thousand systems working in parallel would be required to reconstruct a human brain in a reasonable timeframe. One example is the GCIB-SEM's broad ion milling + MultiSEM approach, where it has been estimated that 20 MultiSEMs will be required for reconstructing one mouse brain, and 3000 for reconstructing one human brain, in 5 years. Such large-scale operation presents a unique opportunity for DOE. In addition, DOE could support the development of advanced optics and large-scale and sensitive imaging detectors to increase throughput, as well as multispectral detection approaches.

X-ray microscopy and tomography technologies show a potential for playing key roles in connectomics – a unique opportunity for DOE. While X-ray techniques are capable of nondestructive imaging of large volumes (mm^3) with high spatial resolution, including frozen-hydrated tissue requiring minimal sample processing, they are yet to demonstrate the ability to resolve synapses. The development of methods and protocols (i.e., cryogenic cooling of large volume tissues to prevent radiation damage), and approaches to increase throughput (i.e., larger detectors with higher speed and sensitivity, more efficient objective lenses) will be critical. Given that these technologies are part of the DOE's Synchrotron Light Source systems, such effort naturally presents a unique opportunity for DOE. The throughput can be improved by more than two orders of magnitudes with the 4^{th} generation light sources, and multiple dedicated beamlines operating in parallel will be required.

Applications to the DOE Mission

The DOE Office of Science operates scientific user facilities for imaging and computation that could play a major role in accelerating a national effort in connectomics. However, the needs of connectomics could also accelerate the application of advanced bioimaging and computation to key imaging problems in the science, energy, national security, and environmental missions of DOE.

The large-scale image based 3D reconstruction of complex nanoscale structures required for the connectome will fill current gaps between molecular and cellular level studies and field studies at scales required for accurate, high resolution modeling of the environment. It will advance our understanding of the structure and function of plant and microbial community interactions and their impact on the environment. The new capabilities will advance large scale data generation and analysis of omics data and assignments of functions to genomes to inform large scale systems biology models for engineering more resilient biological systems and the production of biofuel and bioproducts. These will also support the development of complex biomaterials, including nanostructured composites that have active functions, such as self-repairing materials and materials that respond to the environment.

Similar applications exist in large-scale analysis of semiconductors and advanced manufacturing processes. Furthermore, discoveries that will come from building the mouse brain connectome are expected to yield new ideas for artificial intelligence (AI) models, new approaches to

automated experimental control, and new insights into resilient, low-power neuromorphic computing. Advances in each of these areas will benefit DOE's diverse mission space.

Finally, advances in computing driven by DOE mission applications and the large-scale application of AI and data science to DOE's portfolio of scientific, energy, and national security missions will continue to push computing architectures, data management, storage technology, and advanced workflow systems that are also needed to scale advanced imaging techniques to the level required for microscale connectome reconstruction. The use cases of the connectome are aggressive and complement the existing drivers for DOE. The tight integration of DOE imaging facilities with leadership scale computing is already in progress in the DOE laboratories. This supercomputing enabled imaging ecosystem would be ideal to address the scientific challenges presented by the national-scale connectome project the BRAIN Initiative will pursue.

AA. Roundtable Participants

First Name	Last Name	Institution
Frank	Alexander	Brookhaven National Laboratory
Tammie	Borders	Idaho National Laboratory
Kris	Bouchard	Lawrence Berkeley National Laboratory
Ed	Boyden	MIT H. H. Uhlig Corrosion Laboratory
Wah	Chiu	SLAC National Accelerator Laboratory
Vincent	De Andrade	Argonne National Laboratory
Mark	Ellisman	University of California San Diego
Nicola	Ferrier	Argonne National Laboratory
Ken	Hayworth	Janelia Research Campus
Na	Ji	Lawrence Berkeley National Laboratory
Bobby	Kasthuri	Argonne National Laboratory
Garrett	Kenyon	Los Alamos National Laboratory
Carolyn	Larabell	Lawrence Berkeley National Laboratory
Scott	Lea	Pacific Northwest National Laboratory
Jeff	Lichtman	Harvard University
Tianyi	Mao	Oregon Health & Science University
Lisa	Miller	Brookhaven National Laboratory
Galya	Orr	Pacific Northwest National Laboratory
Clay	Reid	Allen Institute
Scott	Retterer	Oak Ridge National Laboratory
Mark	Schnitzer	Stanford University
Catherine (Katie)	Schuman	Oak Ridge National Laboratory
Rick	Stevens	Argonne National Laboratory
Tamas	Varga	Pacific Northwest National Laboratory
Stefan	Vogt	Argonne National Laboratory
Bobbie-Jo	Webb-Robertson	Pacific Northwest National Laboratory
Jim	Werner	Los Alamos National Laboratory
Zhuhao	Wu	Mount Sinai
Xianghui	Xiao	Brookhaven National Laboratory
Hongkui	Zeng	Allen Institute

Bioimaging Capabilities to Enable Mapping of the Neural Connections in a Complex Brain
Virtual Roundtable
November 2 - 4, 2020
10:00am – 3:00pm CT

Background

The Department of Energy Office of Science (SC), through its BRAIN Initiative Working Group, is convening a virtual DOE Roundtable to develop a conceptual framework to identify the DOE capabilities needed to enable mapping the neural connections in the mouse brain and ultimately the human brain.

Roundtable

Participants will explore the three main topics listed below, highlighting where there is a clear mutual benefit to DOE's missions in science, energy, security, and environmental stewardship.

1. **Connectome Requirements**: From a Neuroscience perspective, what are the neuronal properties, and the imaging and computing requirements to fully map the connections in the brains of a mouse, a human and other relevant mammals? We will address a progression of challenge problems and associated scaling issues ranging from 1 mouse brain in one year to 100 human brains in one year.
2. **Bioimaging Technology Roadmap**: Focusing on imaging modalities, including optical, electron and x-ray technologies, what are the technical R&D hurdles preventing the achievement of the goals defined under the first topic (i.e., automation, sample preparation and data acquisition in high throughput and parallel configurations)? Are there other modalities not listed that should be considered?
3. **Computing and Data Roadmap**: What is the R&D needed to address the computing and data environment needed to host and process connectome related datasets and the algorithms and software stacks needed to automate the handling and transformation of datasets generated?

Output

The output of this Roundtable will be a report that outlines a roadmap for how DOE science-at-scale capability in imaging and computing through suitable investments and R&D could support a national connectome effort. The report will be due by the end of the year. This report will inform a multi-agency workshop on the connectome that is planned for early 2021.

Document Directory

https://anl.app.box.com/folder/123719894360

Video Conferencing

https://bluejeans.com/629617487

Monday, November 2, 2020

10:00am	**Introduction and Overview** *Rick Stevens and Galya Orr*
10:20am	**Connectome Requirements** *Ed Boyden*
10:30am	**Bioimaging Technology** *Stefan Vogt*
10:40am	**Bioimaging Technology** *Tianyi Mao*
10:50am	**Computing and Data** *Garrett Kenyon*
11:00am	**Panel Discussion: 1** *Boyden, Vogt, Mao, Kenyon*
12:00pm	**Break for Lunch**
1:00pm	**Connectome Requirements** *Bobby Kasthuri*
1:10pm	**Bioimaging Technology** *Na Ji*
1:20pm	**Bioimaging Technology** *Wah Chiu*
1:30pm	**Bioimaging Technology** Xianghui Xiao
1:40pm	**Computing and Data** *Frank Alexander*
2:00pm	**Panel Discussion: 2** *Kasthuri, Ji, Chiu, Xiao, Alexander*
3:00pm	**Adjourn**

Tuesday, November 3, 2020

10:00am	**Greetings from the Director** Dr. Chris Fall
10:10am	**Connectome Requirements** *Jeff Lichtman*
10:20am	**Bioimaging Technology** *Carolyn Larabell*
10:30am	**Bioimaging Technology** *Mark Ellisman*
10:40am	**Bioimaging Technology** *Jim Werner*
10:50am	**Computing and Data** *Nicola Ferrier*
11:00am	**Panel Discussion: 3** *Lichtman, Larabell, Ellisman, Werner, Ferrier*
12:00pm	**Break for Lunch**
1:00pm	**Connectome Requirements** *Mark Schnitzer*
1:10pm	**Bioimaging Technology** *Lisa Miller*
1:20pm	**Bioimaging Technology** *Tamas Varga*
1:30pm	**Bioimaging Technology** *Scott Lea*
1:40pm	**Computing and Data** *Kris Bouchard*
2:00pm	**Panel Discussion: 4** *Schnitzer, Miller, Varga, Lea, Bouchard*
3:00pm	**Adjourn**

Wednesday, November 4, 2020

10:00am	**Charge for Day 3** *Rick Stevens and Galya Orr*
10:10am	**Connectome Requirements** *Hongkui Zeng*
10:20am	**Bioimaging Technology** *Zhuhao Wu*
10:30am	**Bioimaging Technology** *Vincent De Andrade*
10:40am	**Computing and Data** *Bobbie-Jo Webb-Robertson*
11:00am	**Panel Discussion: 5** *Zeng, Wu, De Andrade, Webb-Robertson*
12:00pm	**Break for Lunch**
1:00pm	**Connectome Requirements** *Clay Reid*
1:10pm	**Bioimaging Technology** *Ken Hayworth*
1:20pm	**Bioimaging Technology** *Scott Retterer*
1:30pm	**Computing and Data** *Katie Schuman*
1:40pm	**Computing and Data** *Tamie Borders*
2:00pm	**Panel Discussion: 6** *Reid, Hayworth, Retterer, Schuman, Borders*
2:50pm	**Concluding Remarks** *Rick Stevens and Galya Orr*
3:00pm	**Adjourn**

References

1. Brown T, Herbert-Voss and Gretchen Krueger and Tom Henighan and Rewon Child and Aditya Ramesh and Daniel M. Ziegler and Jeffrey Wu and Clemens Winter and Christopher Hesse and Mark Chen and Eric Sigler and Mateusz Litwin and Scott Gray and Benjamin Chess and Jack Clark and Christopher Berner and Sam McCandlish and Alec Radford and Ilya Sutskever and Dario Amodei BMaNRaMSaJKaPDaANaPSaGSaAAaSAaA. 2020. Language models are few-shot learners. arXiv: 2005.14165.
2. Dana H, Sun Y, Mohar B, Hulse BK, Kerlin AM, Hasseman JP, Tsegaye G, Tsang A, Wong A, Patel R et al. 2019. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. Nat Methods. 16(7):649-657.
3. Kim TH, Zhang Y, Lecoq J, Jung JC, Li J, Zeng H, Niell CM, Schnitzer MJ. 2016. Long-term optical access to an estimated one million neurons in the live mouse cortex. Cell Rep. 17(12):3385-3394.
4. Lipovsek M, Bardy C, Cadwell CR, Hadley K, Kobak D, Tripathy SJ. 2021. Patch-seq: Past, present, and future. J Neurosci. 41(5):937-946.
5. Fang T, Lu X, Berger D, Gmeiner C, Cho J, Schalek R, Ploegh H, Lichtman J. 2018. Nanobody immunostaining for correlated light and electron microscopy with preservation of ultrastructure. Nat Methods. 15(12):1029-1032.
6. Wu GH, Mitchell PG, Galaz-Montoya JG, Hecksel CW, Sontag EM, Gangadharan V, Marshman J, Mankus D, Bisher ME, Lytton-Jean AKR et al. 2020. Multi-scale 3d cryo-correlative microscopy for vitrified cells. Structure. 28(11):1231-1237.e1233.
7. Kuan AT, Phelps JS, Thomas LA, Nguyen TM, Han J, Chen CL, Azevedo AW, Tuthill JC, Funke J, Cloetens P et al. 2020. Dense neuronal reconstruction through x-ray holographic nano-tomography. Nat Neurosci. 23(12):1637-1643.
8. Chen F, Tillberg PW, Boyden ES. 2015. Optical imaging. Expansion microscopy. Science. 347(6221):543-548.
9. Chang JB, Chen F, Yoon YG, Jung EE, Babcock H, Kang JS, Asano S, Suk HJ, Pak N, Tillberg PW et al. 2017. Iterative expansion microscopy. Nat Methods. 14(6):593-599.
10. Alon S, Goodwin DR, Sinha A, Wassie AT, Chen F, Daugharthy ER, Bando Y, Kajita A, Xue AG, Marrett K et al. 2020. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. bioRxiv.2020.2005.2013.094268.
11. Chen BC, Legant WR, Wang K, Shao L, Milkie DE, Davidson MW, Janetopoulos C, Wu XS, Hammer JA, Liu Z et al. 2014. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. Science. 346(6208):1257998.
12. Gao R, Asano SM, Upadhyayula S, Pisarev I, Milkie DE, Liu TL, Singh V, Graves A, Huynh GH, Zhao Y et al. 2019. Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. Science. 363(6424).
13. Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW. 2007. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature. 450(7166):56-62.
14. Lichtman JW, Livet J, Sanes JR. 2008. A technicolour approach to the connectome. Nat Rev Neurosci. 9(6):417-422.
15. Cai D, Cohen KB, Luo T, Lichtman JW, Sanes JR. 2013. Improved tools for the brainbow toolbox. Nat Methods. 10(6):540-547.
16. Tillberg PW, Chen F, Piatkevich KD, Zhao Y, Yu CC, English BP, Gao L, Martorell A, Suk HJ, Yoshida F et al. 2016. Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. Nat Biotechnol. 34(9):987-992.

17. Shen FY, Harrington MM, Walker LA, Cheng HPJ, Boyden ES, Cai D. 2020. Light microscopy based approach for mapping connectivity with molecular specificity. Nat Commun. 11(1):4632.
18. Sun W, Tan Z, Mensh BD, Ji N. 2016. Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs. Nat Neurosci. 19(2):308-315.
19. Lu R, Liang Y, Meng G, Zhou P, Svoboda K, Paninski L, Ji N. 2020. Rapid mesoscale volumetric imaging of neural activity with synaptic resolution. Nat Methods. 17(3):291-294.
20. Rumyantsev OI, Lecoq JA, Hernandez O, Zhang Y, Savall J, Chrapkiewicz R, Li J, Zeng H, Ganguli S, Schnitzer MJ. 2020. Fundamental bounds on the fidelity of sensory cortical coding. Nature. 580(7801):100-105.
21. Scheffer LK, Xu CS, Januszewski M, Lu Z, Takemura SY, Hayworth KJ, Huang GB, Shinomiya K, Maitlin-Shepard J, Berg S et al. 2020. A connectome and analysis of the adult. Elife. 9.
22. Schalek R, Lee D, Kasthuri N, Peleg A, Jones T, Kaynig V, Haehn D, Pfister H, Cox D, Lichtman JW. 2016. Imaging a 1 mm 3 volume of rat cortex using a multibeam sem. Microscopy and Microanalysis. 22(S3):582-583.
23. Yin W, Brittain D, Borseth J, Scott ME, Williams D, Perkins J, Own CS, Murfitt M, Torres RM, Kapner D et al. 2020. A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy. Nat Commun. 11(1):4949.
24. Xu CS, Hayworth KJ, Lu Z, Grob P, Hassan AM, García-Cerdán JG, Niyogi KK, Nogales E, Weinberg RJ, Hess HF. 2017. Enhanced fib-sem systems for large-volume 3d imaging. Elife. 6.
25. Hayworth KJ, Peale D, Januszewski M, Knott GW, Lu Z, Xu CS, Hess HF. 2020. Gas cluster ion beam sem for imaging of large tissue samples with 10 nm isotropic resolution. Nat Methods. 17(1):68-71.
26. Beck M, Baumeister W. 2016. Cryo-electron tomography: Can it reveal the molecular sociology of cells in atomic detail? Trends Cell Biol. 26(11):825-837.
27. Gorelick S, Buckley G, Gervinskas G, Johnson TK, Handley A, Caggiano MP, Whisstock JC, Pocock R, de Marco A. 2019. Pie-scope, integrated cryo-correlative light and fib/sem microscopy. Elife. 8.
28. Strunk KM, Wang K, Ke D, Gray JL, Zhang P. 2012. Thinning of large mammalian cells for cryo-tem characterization by cryo-fib milling. J Microsc. 247(3):220-227.
29. A.F. L, S. P, M. E. 2012. Electron tomography and multiscale biology. In: M. A, S.B. C, A. L, editors. Theory and Applications of Models of Computation: Lecture Notes in Computer Science.
30. Phan S, Boassa D, Nguyen P, Wan X, Lanman J, Lawrence A, Ellisman MH. 2017. 3d reconstruction of biological structures: Automated procedures for alignment and reconstruction of multiple tilt series in electron tomography. Adv Struct Chem Imaging. 2:8.
31. E.A. B, S. P, M.H. E. 2020. Using x-ray microscopy to increase targeting accuracy in serial block-face scanning electron microscopy. In: I. W, E. H, S. B, R. S, editors. Volume Microscopy: Neuromethods.
32. Bouwer JC, Mackey MR, Lawrence A, Deerinck TJ, Jones YZ, Terada M, Martone ME, Peltier S, Ellisman MH. 2004. Automated most-probable loss tomography of thick selectively stained biological specimens with quantitative measurement of resolution improvement. J Struct Biol. 148(3):297-306.
33. Martell JD, Deerinck TJ, Lam SS, Ellisman MH, Ting AY. 2017. Electron microscopy using the genetically encoded apex2 tag in cultured mammalian cells. Nat Protoc. 12(9):1792-1816.
34. Adams SR, Mackey MR, Ramachandra R, Palida Lemieux SF, Steinbach P, Bushong EA, Butko MT, Giepmans BNG, Ellisman MH, Tsien RY. 2016. Multicolor electron microscopy for simultaneous visualization of multiple molecular species. Cell Chem Biol. 23(11):1417-1427.
35. Hua Y, Laserstein P, Helmstaedter M. 2015. Large-volume en-bloc staining for electron microscopy-based connectomics. Nat Commun. 6:7923.

36. Smith EA, McDermott G, Do M, Leung K, Panning B, Le Gros MA, Larabell CA. 2014. Quantitatively imaging chromosomes by correlated cryo-fluorescence and soft x-ray tomographies. Biophys J. 107(8):1988-1996.
37. Georgiadis M, Gao Z, Zingariello D, Zerbi V, Liebi M, Sommer S, Augath M, Bunk O, Guizar-Sicairos M, Schroeter A et al. 2017. Mri vs. X-ray scattering: Comparative study of myelin distribution, fiber direction and white matter tracts in mouse brain. Proc. Intl. Soc. Mag. Reson. Med. 25.
38. Victor TW, O'Toole KH, Easthon LM, Ge M, Smith RJ, Huang X, Yan H, Chu YS, Chen S, Gursoy D et al. 2020. Lanthanide-binding tags for 3d x-ray imaging of proteins in cells at nanoscale resolution. J Am Chem Soc. 142(5):2145-2149.
39. Wang Y, Li Q, Liu L, Zhou Z, Ruan Z, Kong L, Li Y, Zhong N, Chai R, Luo X et al. 2019. Teravr empowers precise reconstruction of complete 3-d neuronal morphology in the whole brain. Nat Commun. 10(1):3474.
40. Hidayetoglu M, Bicer T, Gonzalo S, Ren B, Andrade V, Gursoy D, Kettimuthu R, Foster I, Hwu W. 2020. 2020 sc20: International conference for high performance computing, networking, storage and analysis (sc). Petascale XCT: 3D Image Reconstruction with Hierarchical Communications on Multi-GPU Nodes. Los Alamitos, CA, USA: IEEE Computer Society. p. 510-522.
41. Lehe R, Hanuka A, Edelen A, Huang X, Mayes C, Cook N, Huebl A, Emma C, Djordjevi B, Kemp A et al. September 1, 2020. Machine learning and surrogate models for simulation-based optimization of accelerator design.
42. Nyshadham C, Rupp M, Bekker B. 2019. Machine-learned multi-system surrogate models for materials prediction. npj Comput Mater.
43. Eisenbach M, Zhang J, Pei Z, Lupo Pasini M, Li YW, Junqi. 2019. Machine learning surrogate models to accelerate monte-carlo calculation. APS March Meeting Abstracts. @INPROCEEDINGS: APS Meeting Abstracts.
44. Jeon K, Yang S, Kang D, Na J, Lee WB. 2019. Development of surrogate model using cfd and deep neural networks to optimize gas detector layout. *Korean J. Chem. Eng*.
45. Ghose S, HSIEH K, BOROUMAND A, AUSAVARUNGNIRUN R, MUTLU O. 2018. Enabling the adoption of processing-in-memory: Challenges, mechanisms, future research directions. arXiv:1802.00320v1.
46. Gupta S. 2018. Processing in memory using emerging memory technologies. UC San Diego.
47. Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, Ailamaki A, Alonso-Nanclares L, Antille N, Arsever S et al. 2015. Reconstruction and simulation of neocortical microcircuitry. Cell. 163(2):456-492.
48. Roy K, Jaiswal A, Panda P. 2019. Towards spike-based machine intelligence with neuromorphic computing. Nature. 575(7784):607-617.
49. Kasim MF, Watson-Parris D, Deaconu L, Oliver S, Hatfield D, et al. 2020. Up to two billion times acceleration of scientific simulations with deep neural architecture search. arXiv:2001.08055v1.
50. Mulcahy B, Witvliet D, Holmyard D, Mitchell J, Chisholm A, Meirovitch Y, Samuel A, Zhen M. 2018. A Pipeline for Volume Electron Microscopy of the *Caenorhabditis elegans* Nervous System. Front. Neural Circuits, 21 November 2018