Multiscale Computational Modeling for the US BRAIN initiative

Bill Lytton, Raj Vadigepalli, Mark Kramer, IMAG BRAIN theme group, IMAG Computational Neuroscience group

How will we understand the brain? To start, we must observe the brain's dynamic activity, which consists of rich patterns propagating across the brain's spatial and temporal scales, within the context of the brain's complex anatomy. A fundamental component of the BRAIN Initiative is the development of new technologies. Here we present suggestions for new technologies in simulation and mathematical modeling – spanning spatial scales from molecules to large neuronal populations – leveraging the strengths of the Multiscale Modeling (MSM) community.

Although statistical approaches and data mining are critical to understanding the brain, these representations typically lack a mechanistic, biological interpretation of brain function, and do not directly connect phenomenology across scales. To understand the brain's multiscale activity will require combined approaches, linking physical and conceptual reasoning - i.e., computational experiments - directly to neuronal data, whether raw or processed - and ideally both. Although new technologies continue to facilitate the collection of novel brain data, we will never completely observe all features of the brain's dynamics. An important role for MSM is to provide insight into the hidden biological mechanisms underlying the brain's activity that are not directly observed. This becomes critical in: 1. **Connecting spatial and temporal scales** where information gathered with different techniques is initially incommensurable; 2. **Identifying gaps** in datasets where critical parameters have not yet been gathered; 3. Making **explicit predictions** that not only test the adequacy of the model but also test the adequacy of the underlying data in providing explanations for an emergent brain phenomenon of interest, e.g., dynamical, informational, behavioral, or cognitive.

The following new technologies are needed in simulation and mathematical analysis:

1. Directed high-dimensional nonlinear dynamical analysis methods. Detailed biological models necessarily reside in high-dimensional state space. However, methods for model analysis generally require dimensional reduction to 2 or perhaps 3 dimensions; in these dimensions, important dynamical system features (e.g., manifolds of fixed points and limit cycles, and their stability) can be visualized. Computer simulations allow us to "see" all state variables simultaneously in high-dimensional space (e.g., thousands or millions of state variables). Rather than projecting these dynamics down to convenient dimensions for visual inspection (a data-mining dimensional reduction), we need to further develop mathematical tools that permit both a wide angle, global perspective of the full system space, and high resolution, focused perspectives on individual system components. A goal is to bring our "virtual eyes" - our manipulations and our evaluation process - into the system's native space. In this context, large scale pattern analysis tools that have been successful in the so called 'omics' studies could be adapted for analyzing high dimensional simulation data.

2. Development of new measures to assess fundamental features in large nonlinear systems. Characterization of a nonlinear system often requires determination of important features such as stability, identification of manifolds and dimensionality, analysis of sensitivity and robustness to perturbations, and detection of stable and unstable orbits. We require comparable methods for large nonlinear systems, and for dynamics based on mixed systems (e.g., systems consisting of ordinary differential equations, combined with delay differential equations, combined with partial differential equations, combined with event-driven effects). Of relevance to this, we need to define **functionally relevant large-scale input/perturbation patterns** that could be used to evaluate new measures of stability and robustness in realistic contexts and would be useful in predicting results of new experiments. The cellular activity and molecular phenotyping data from the BRAIN Initiative could provide initial estimates of these high-dimensional input patterns.

3. New simulation methods for multiscale models. Simultaneous simulation of activity at different spatial, temporal and abstraction scales requires different techniques that must be made to work together. For example, at the subcellular scale this requires different stochastic simulators to couple with 1D and 3D diffusion, as well as with reaction schemes defined deterministically or stochastically. At the neuronal network level, this requires hybrid simulations that combine detailed multi-compartmental cells with more basic integrate-and-fire cells. At even larger spatial scales, this requires methods for connecting neuronal networks to simulate brain areas, connecting brain areas to simulate systems, and finally connecting systems to simulate full-brain models. How to connect these models across scales - from the subcellular to the full-brain - and integrate additional features (e.g., neurovascular coupling) remains a challenge.

4. Extend multiscale models up to behavior and representations Current multiscale methods focus largely on bottom-up structural constraints. Explicit top-down models of behaviors are expressed by engineers in control-theory simulations that describe movements, and by linguists and cognitive scientists in a variety of symbolic representations and symbol sequencers. It would be desirable to identify explicit correspondences between phenomenological top-down models and high-dimensional bottom-up models. An ideal would be to then incorporate the phenomenological model as the top scale of a multiscale model.

5. Develop methods to link data and models. Most computational models possess a large number of parameters, which typically remain experimentally unconstrained. A common procedure for estimating these parameters is “hand-tuning” to produce simulated model dynamics that match qualitatively the desired neuronal activity. Hand-tuning approaches usually require a great deal of time and expertise. Moreover, once a set of suitable parameters is found, it is often unclear whether the solution is unique or whether other model formulations compatible with the data exist. New techniques that go beyond hand-tuning, and instead rigorously estimate model features from neuronal data are required.

6. Simulator interoperability. Multiple simulators have been moving towards each other by

virtue of all adopting Python as a *lingua franca*. However, many difficulties remain in developing reliable interfaces between complex simulators. Standard interfacing must be developed via application programming interfaces (APIs) that define what information needs to be sent from one simulator to the other, both within and between spatial scales and scales of abstraction. This must be done in the context of High Performance Computers (HPCs), which place different segments of a simulation on different processors.

7. Development of databases for data and model sharing. Accessible, shared resources need to be developed at multiple levels. Databases should be defined both to suit the needs of experimentalists and data sharing, as well as to work readily with current and future simulators. A good example of the difficulties of this can be seen in existing anatomical databases, which have largely been developed for visualization by neuroanatomists but are often unfriendly for direct automated access by a simulator trying to extract needed information. Moreover, as computational models become increasing complex, continuing efforts are required to develop collaborative and community-driven platforms that support model verification.

8. Continuous co-system data-mining. Increased automation will support continuous data from two processes: Ongoing (computer-directed automated) experiments will be accompanied by ongoing model simulations, with both being probed through ongoing continuous data-mining. This integrated process can be effected at at least four levels of complexity.
Level 1: Data-miner detects faults or errors in the experiment or in a companion simulation.
Level 2: Data-miner identifies activity of interest in either set; they then perform more detailed post-simulation comparisons between model and experimental results, identifying additional similarities, model features, or experimental features.
Level 3: Co-adaptation: 3a: Simulations are updated based on parameters just now inferred from experiment. 3b: Based on knowledge of the model, an automated experimental apparatus is redirected to obtain new parameters, different locations or different aspects of a parameter being measured in an optimally informative way.
Level 4: Simulations automatically adapt to more closely match data via newly developed inductive or selective algorithms. Ideally these algorithms will closely mirror actual biological processes so that the system is modeling development and learning as well as modeling dynamics.