IBM Research Report

Towards Real-Time, Mouse-Scale Cortical Simulations

James Frye*, Rajagopal Ananthanarayanan, Dharmendra S. Modha
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

*Also with University of Nevada, Reno

Towards real-time, mouse-scale cortical simulations

James Frye[1,2]**, Rajagopal Ananthanarayanan**[1]**, and Dharmendra S. Modha**[1]
[1]IBM Almaden Research center,   [2]University of Nevada, Reno

Neurobiologically realistic, large-scale cortical and sub-cortical simulations are bound to play a key role in computational neuroscience and its applications to cognitive computing [1]. The mouse cortex has roughly 8×10^6 neurons and 8000 synapses per neuron. Modeling at this scale imposes tremendous constraints on computation, communication, and memory capacity of any computing platform. For example, assuming an average firing rate of 1 Hz, the entire memory must be refreshed every second, each neuron must be updated at every simulation time step, and each neuron communicates to each of its targets at least once a second.

We have designed and implemented a massively parallel cortical simulator. The simulator is designed for low to moderate complexity spiking neurons, for example, [2]. The state of each neuron is updated synchronously at a user-specified resolution, say, 1ms. Each neuron can be presynaptic to a realistic number of target neurons. Each synapse can implement STDP [3]. The simulator also incorporates axonal conductance delays ranging from a low of one time step to a model-defined maximum. To realize mouse-scale simulations, we have made several key novel technical contributions as briefly discussed below.

1. In previous simulators (e.g., NCS [1]), each neuron, upon its firing, sends an individual message to each of its target neurons. This becomes a bottleneck. We observed that, in a massively parallel implementation each compute node will house and process several neurons, and, hence, each neuron can group messages to all targets that reside on a single compute node into a single message. This idea drastically reduces the communication complexity from order of synapses to order of neurons.

2. In STDP, when a neuron fires, each excitatory synapse is examined for LTP, which appears to require a computation proportional to the number of synapses in that neuron. Instead, we maintain all synapses that were activated since the last neuronal firing on a recently-fired list. Since, typically, the size of this list is far smaller than the number of synapses, this leads to a significant computational benefit.

We deployed the simulator on a 4096-processor BlueGene/L supercomputer with 256 MB per CPU. Using spiking neurons [2], we were able to represent 8×10^6 neurons (80% excitatory) and 6300 synapses per neuron in the 1 TB main memory of the system. Using a synthetic pattern of neuronal interconnections, at a 1 ms resolution and an average firing rate of 1 Hz, we were able to run 1s of model time in 10s of real time. On other (smaller) models, we have observed biologically consistent dynamical properties such as spontaneous formation of neuronal groups and synchronous/asynchronous firing patterns. Our goal is to continually refine the computational architecture and to add neurobiological details such as macro and micro neuroanatomy, neuromodulators, and more detailed neuron and synapse dynamical equations.

References

[1] R. Brette et al., "Simulation of networks of spiking neurons: A review of tools and strategies," *Journal of Computational Neuroscience (submitted)*, 2006.

[2] E. M. Izhikevich, J. A. Gally, and G. M. Edelman, "Spike-timing dynamics of neuronal groups," *Cerebral Cortex*, vol. 14, pp. 933–944, 2004.

[3] H. Markram, J. Lubke, M. Frotscher, and B. Sakmann, "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs," *Science*, vol. 275, pp. 213–215, 1997.