Cortex: High Performance Computing based Large Scale Brain Circuit Simulation Framework

Tianxiang Lyu *1 and Zhe Sun †2

[1]Department of Mechanics, Sun Yat-sen University, 510275, Guangzhou, China
[2]Computational Engineering Applications Unit, RIKEN, Japan

Sep 2022

1 Abstract

Computational neuroscience seems to be the current methodology for systematically approaching the multi-scale brain with ever-increasing and more detailed neural networks. Moreover, the large spatial and temporal scale of the brain simulation with complex connectivity in brains under study mandate the utilization of a simulation framework running on supercomputers, for the purpose of resulting in a reasonable time. Here, we present Cortex, a novel and user-friendly framework. In the HPC Benchmark on mainstream scientific computing environments, Cortex with precise spike times runs more than 4 orders-of-magnitude faster than NEST, however, not in precise spiking times mode, and reduce 80% memory consumption in maximum.

2 Design Concepts

The main idea of Cortex is to formulate an abstract description of the approach for users, and apply it to specific simulations in general neuroscience. A serial "program" written by user could be parallelized by Cortex using its libraries functions as "metaprogramming".
The main loop of neuroscientific simulation could be the following:
1. Spike Broadcast
2. Synaptic Delay Polling
3. Synapses-Neurons Interaction

*lvtx@mail2.sysu.edu.cn
†zhe.sun.vk@riken.jp

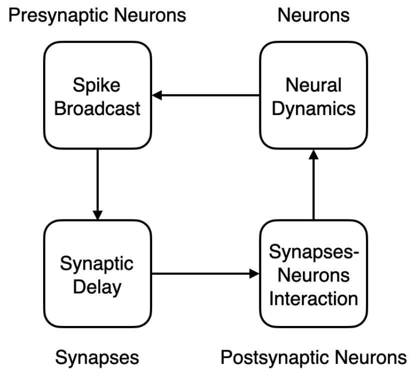


Figure 1: The Main Loop of Cortex

4. Neurons' Dynamics Calculation
The procedures mentioned above are the different stages on which Cortex dealing with the user-defined cases. These program interfaces are clear for users to know what the framework is processing and maintain the related parameters, data types and functions to build a specific simulation.

3 Scalability and Performance

In this section, a simulation will be used for study the scalability and performance of Cortex in terms of run time and memory usage running on a distributed-memory high-performance computing system with 2 Intel Xeon Scalable Gold 6148 Processors and 189GB RAM inside each node.
For the measurements of memory usage and run time, Balanced random networks[1, 2] consist of 80% excitatory and 20% inhibitory intergrate-and-fire neurons with alpha-shape post-synaptic currents, the model iaf_psc_alpha[3], is used. All excitatory to excitatory connections exhibite spike timing dependent plasticity while all other connections are static.
K_{in} = 11250 ($K_{in,\mathcal{T}E}$ = 9000, $K_{in,\mathcal{T}I}$ = 2250) is the total number of incoming connections per neuron. Normal distribution with $\mu = 5.7mV$ and $\sigma = 7.2mV$ drawn the initial membrane potentials. The initial synaptic weights of excitatory are set to J_E = 45.61 pA. J_I = -228.05 pA are the weights for inhibitory synapses.
Excitatory external poissonian inputs are received by all neurons, causing a mean membrane potential of $\eta V_{th} = \tau_{syn} J_E \frac{\tau_m}{C_m} \nu_{ext}$. With η = 1.685, V_{th} = 20 mV, τ_m = 10 ms, C_m = 250 pF, and τ_{syn} = 0.3258 ms, the input spike rate is $\nu_{ext} = \eta \frac{V_{th}}{\tau_{syn} J_E \frac{\tau_m}{C_m}} \simeq 20856$ spikes per second for all external inputs to one neuron. Time for simulation is fixed to 500 ms for 0.1 ms per time-step.

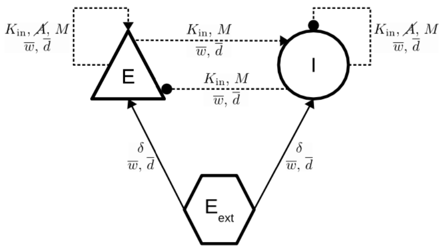
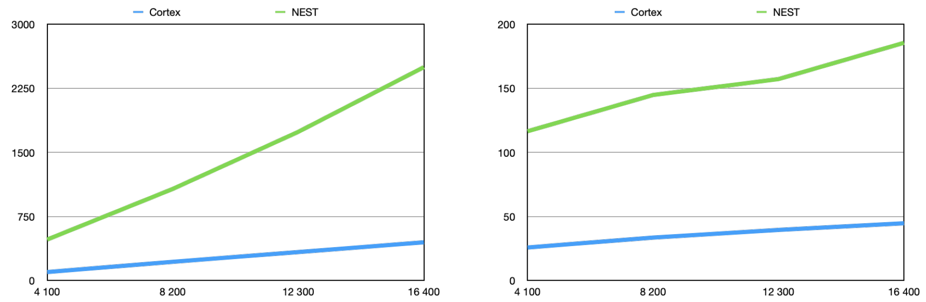


Figure 2: Random Balanced Network



(a) Memory Consumption

(b) Runtime

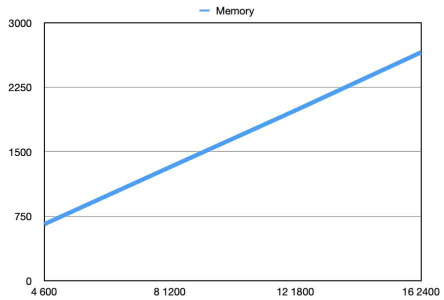
Figure 3: Cortex vs NEST

Table 1: Connectivity

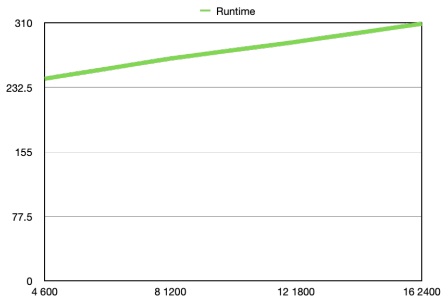
Name	Source	Target	Pattern
E to E	E	E	In-degree $K_{in,\mathcal{T}E}$ with multapses (No autapses)
I to E	E	I	In-degree $K_{in,\mathcal{T}E}$ with multapses
E to I	I	E	In-degree $K_{in,\mathcal{T}I}$ with multapses
I to I	I	I	In-degree $K_{in,\mathcal{T}I}$ with multapses (No autapses)
Ext	E_{ext}	E $\cup$ I	One-to-one

Table 2: Cortex vs NEST

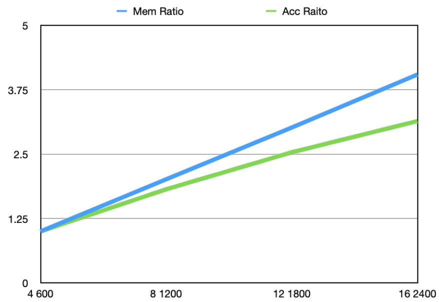
Nodes	Populations $\times 10^6$	Connections $\times 10^{10}$	Cortex Memory(GB)	Cortex Runtime(s)	NEST Memory(GB)	NEST Runtime(s)
4	1.125	1.26	100	25.812	480	116.4
8	2.25	2.52	220	33.553	1072	144.7
12	3.375	3.78	334	39.516	1740	157.1
16	4.5	5.04	448	44.725	2496	185.4



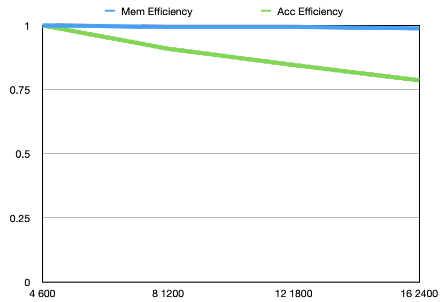
(a) Memory Consumption



(b) Runtime



(c) Ratio



(d) Efficiency

Figure 4: Weak Scaling Results

Table 3: Cortex Weak Scaling Results

Nodes	Populations $\times 10^6$	Connections $\times 10^{10}$	Memory(GB)	Runtime(s)
4	6.75	7.56	656	242.915
8	13.5	15.12	1320	267.243
12	20.25	22.68	1980	287.257
16	27	30.24	2656	309.116

Table 4: Cortex Weak Scaling Results

Nodes	Acc. Ratio	Memory Ratio	Acc. Efficiency	Memory Efficiency
4	1	1	1	1
8	1.82	2.01	0.90	0.99
12	2.54	3.02	0.85	0.99
16	3.14	4.05	0.79	0.98

References

[1] Moritz Helias et al. "Supercomputers Ready for Use as Discovery Machines for Neuroscience". In: *Frontiers in neuroinformatics* 6 (Nov. 2012), p. 26. DOI: `10.3389/fninf.2012.00026`.

[2] Susanne Kunkel et al. "Spiking network simulation code for petascale computers". In: *Frontiers in neuroinformatics* 8 (Oct. 2014), p. 78. DOI: `10.3389/fninf.2014.00078`.

[3] Abigail Morrison, Ad Aertsen, and Markus Diesmann. "Spike-Timing-Dependent Plasticity in Balanced Random Networks". In: *Neural computation* 19 (July 2007), pp. 1437–67. DOI: `10.1162/neco.2007.19.6.1437`.