Computing @CSN5: applications and innovations at INFN
15/10/2024

Large-scale neuronal network models of memory and learning mediated by synaptic plasticity

UNIVERSITÀ DEGLI STUDI DI CAGLIARI

Istituto Nazionale di Fisica Nucleare
Sezione di Cagliari

Gianmarco Tiddia

gianmarco.tiddia@ca.infn.it

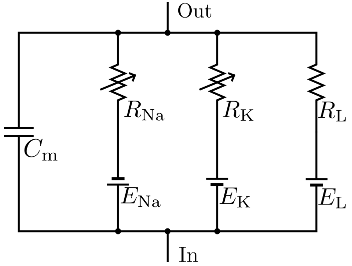
Outline

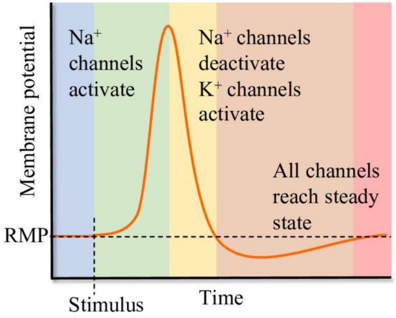
- Spiking Neural Networks (SNNs)
  - Our activities on simulation technology
- Synapse modeling and cognitive processes
  - Short-term synaptic plasticity and Working Memory
  - Structural synaptic plasticity and Learning

Single-neuron modelling

Hodgkin-Huxley neuron, a semiempirical neuron model which describes the membrane potential dynamics.

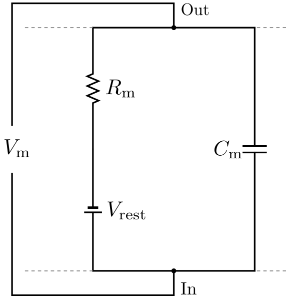
PROS optimal description of the neuronal dynamics
CONS large-scale networks simulations are not efficient

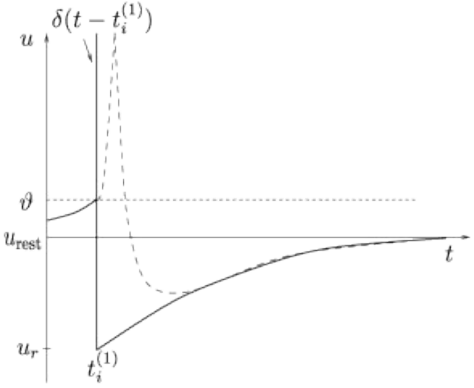




Spiking neuron, a point-like neuron which emits spike events.

PROS can be employed in large-scale network simulations
CONS needs a threshold to describe the spike emission

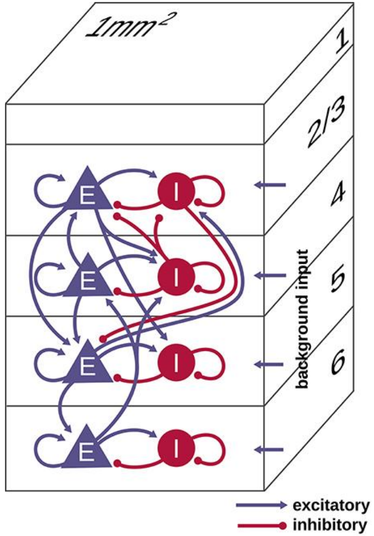




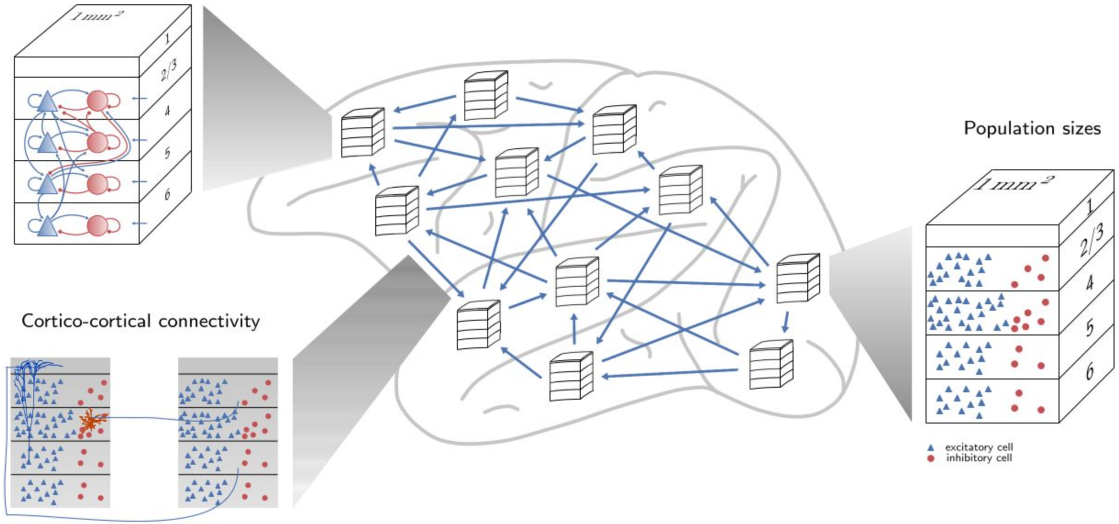
Large-scale models of spiking neurons

Cortical microcircuit: 77000 LIF neurons and 300M connections. It **represents 1mm^2 of cerebral cortex**.

Multi-area model: 4M LIF neurons and ~24B connections. It **represents 32 areas of the macaque vision-related cortex.**



Potjans and Diesmann, ***Cerebral Cortex***, 24(3), 2014



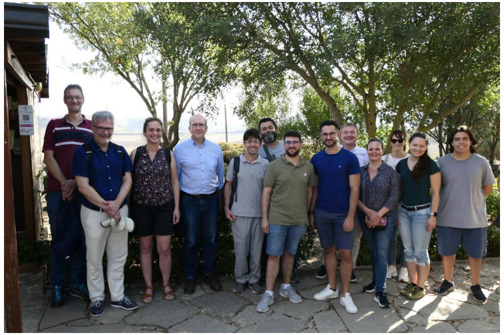
Tiddia et al., ***Front. Neuroinform.***, 16:883333, 2022

NEST GPU: a GPU-based simulator for SNN

NEST (NEural Simulation Tool) is one of the most reliable SNN simulators.
NEST GPU is the GPU-based simulator of the NEST Initiative.

We are able to exploit multi-GPU systems, with the possibility of simulating millions of neurons and billions of synapses in a relatively low simulation time.

Optimization in progress to take advantage of the modern supercomputers, like LEONARDO, with thousands of GPUs available.



NEST GPU is a result of active collaboration between:
Uni Cagliari & INFN, Sezione di Cagliari
INFN, Sezione di Roma 1 (APE Lab)
INM-6, Jülich Research Center

nest :: initiative

nest :: CPU-based simulator

nest :: gpu

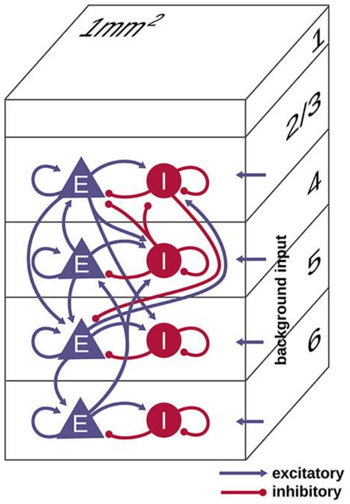
Golosio et al., ***Appl. Sci.***, *13*, 9598, 2023
Tiddia et al., ***Front. Neuroinform.***, 16:883333, 2022
Golosio et al., ***Front. Comput. Neurosci.***, 15:627620, 2021

NEST GPU: performance evaluation (1/2)

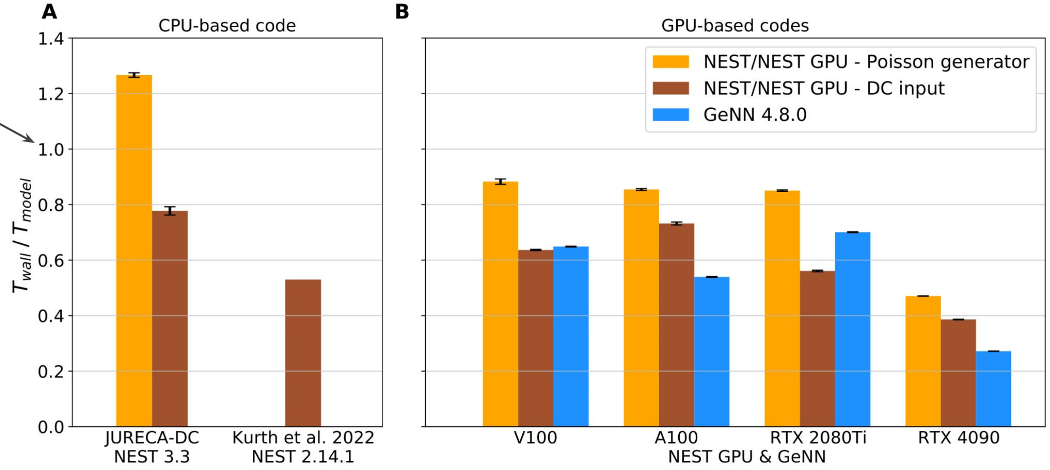
The cortical microcircuit model can be simulated on a single GPU.

IMPORTANT

NEST GPU achieves below real-time performances!



Real- time

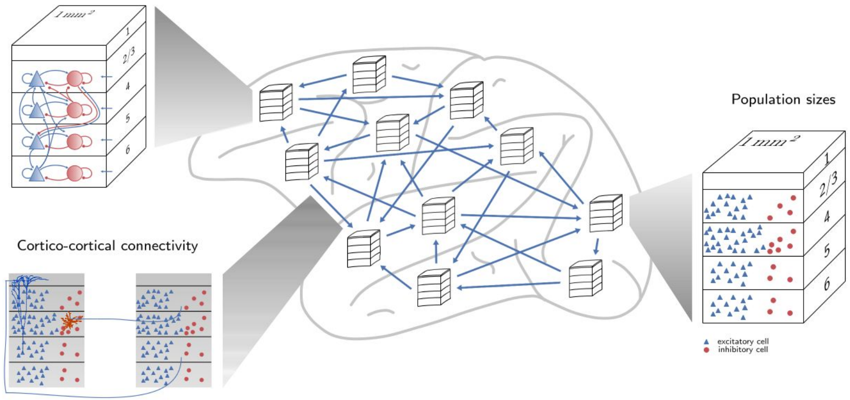


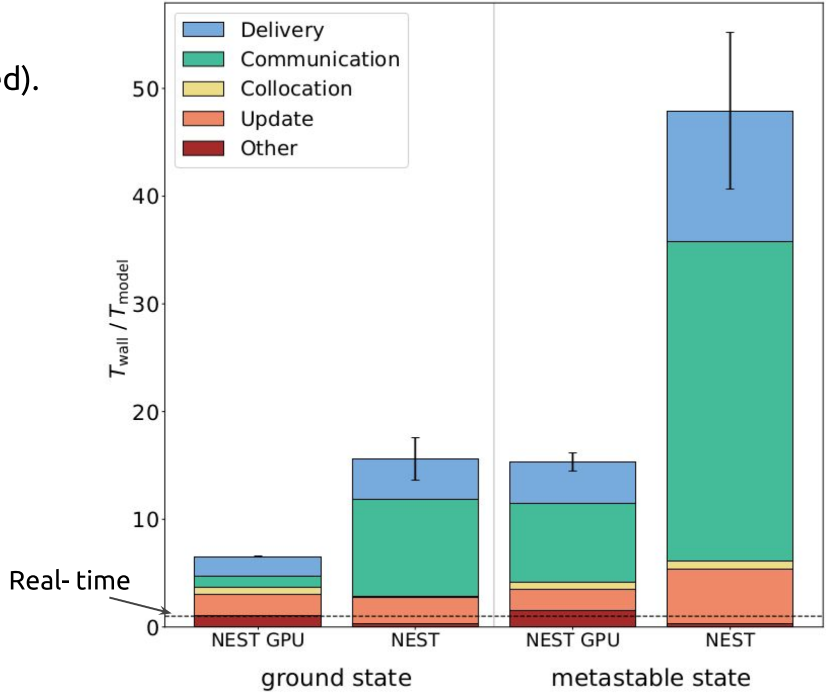
NEST GPU: performance evaluation (2/2)

The multi-area model has to be simulated on a MPI-GPU cluster.
Simulations performed with JUSUF@JSC (32 V100 GPUs employed).

IMPORTANT

NEST GPU achieves state-of-art performances!





Tiddia et al., ***Front. Neuroinform.***, 16:883333, 2022

Synapses and plasticity

Synapses may change over time. This changes are known under the name of **synaptic plasticity**.

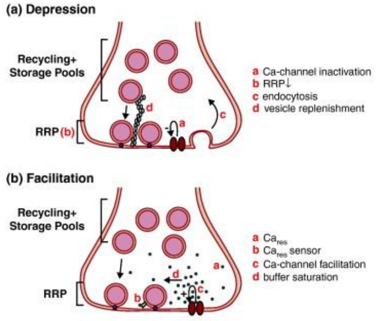
Synaptic plasticity

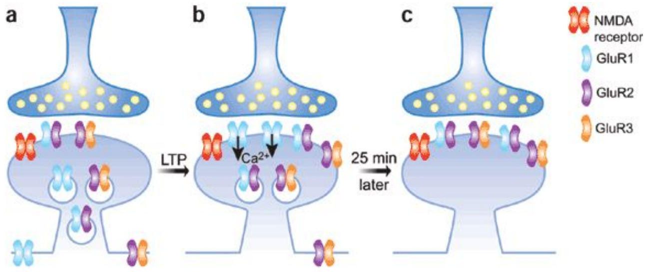
Functional plasticity (changes at the level of a single synapse)

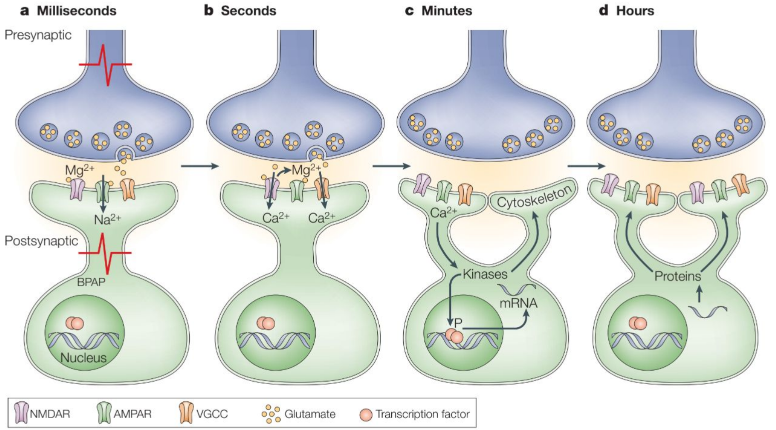
Structural plasticity (changes network topology)

Short-term plasticity

Long-term plasticity





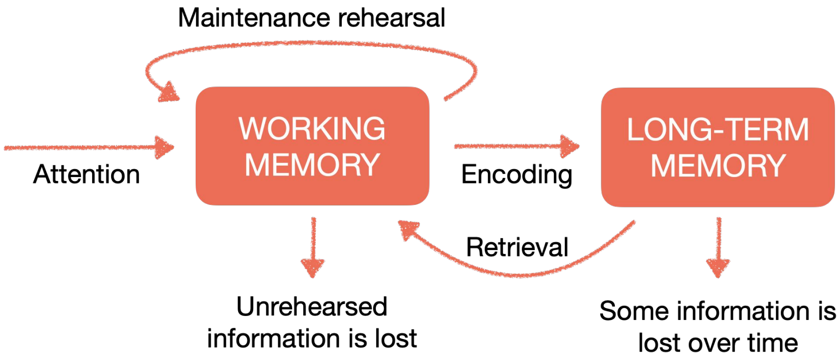


Synaptic theory of working memory

Working Memory (WM) is a cognitive process able to hold and manipulate information for a short time. It is fundamental for speech, visual and spatial processing.
It is observed in the prefrontal cortex (PFC) during *delay response tasks.*

The **Synaptic Theory of Working Memory** posits that a mechanism of short-term synaptic facilitation leads to <u>information maintenance in both synaptic and spiking form</u>, with spiking activity functional for synaptic facilitation upkeep.

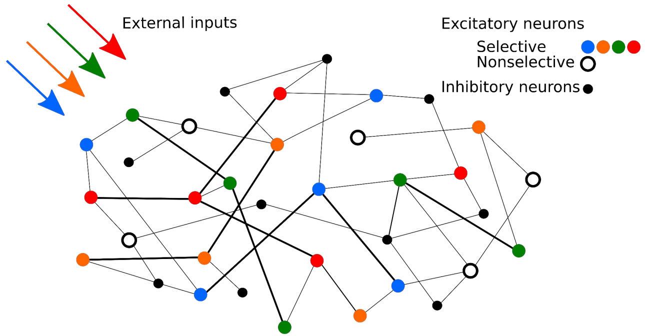
Mongillo et al., ***Science***, 319, 2008



Working Memory → spiking activity + activity-silent mechanism

→ short-term plasticity?

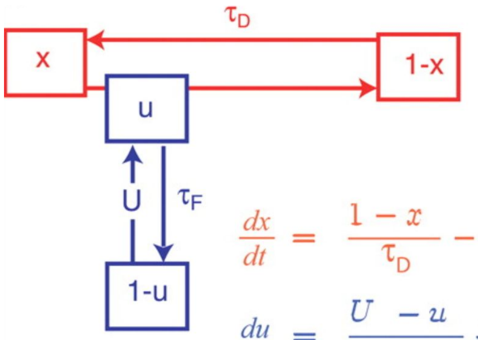
Spiking model of working memory



Raster plots of two selective populations (green and black dots=spikes).

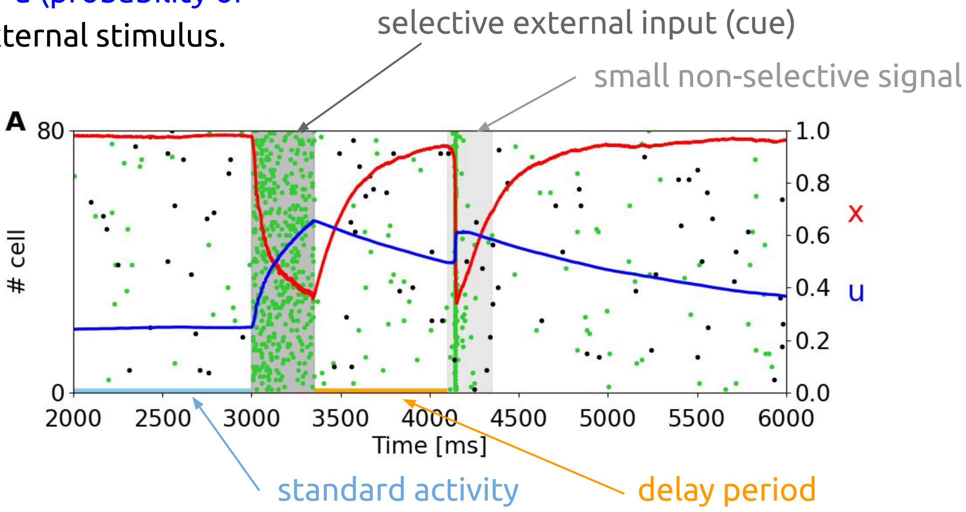
Spiking activity is modulated by the background current (increases from panel A to panel B).

STP variables — x (amount of neurotransmitters) and — u (probability of release) averaged over the population targeted by the external stimulus.



$$\frac{dx}{dt} = \frac{1-x}{\tau_D} - u\,x\,\delta(t-t_{sp})$$

$$\frac{du}{dt} = \frac{U-u}{\tau_F} + U\,(1-u)\,\delta(t-t_{sp})$$



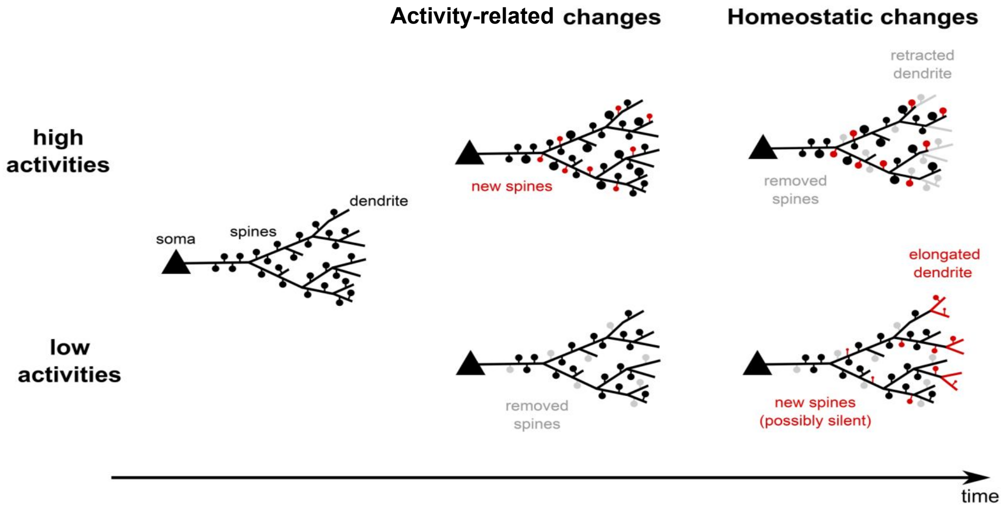
Structural synaptic plasticity

Structural plasticity describes the mechanisms of connection creation, consolidation and erasure (or pruning).

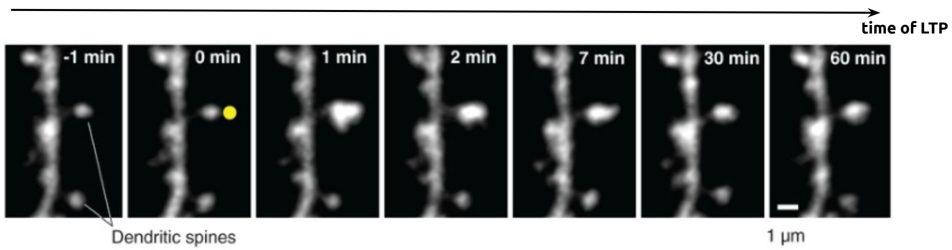
It can be activity-related or homeostatic. Fauth and Tetzlaff, ***Front. Neuroanat.***, 10:75, 2016

In particular, synaptic pruning and connection reorganization are fundamental mechanisms for learning and neural circuits optimization.

Question: can we estimate the impact of structural plasticity in learning?



Fauth and Tetzlaff, ***Front. Neuroanat.***, 10:75, 2016



The firing rate-based model

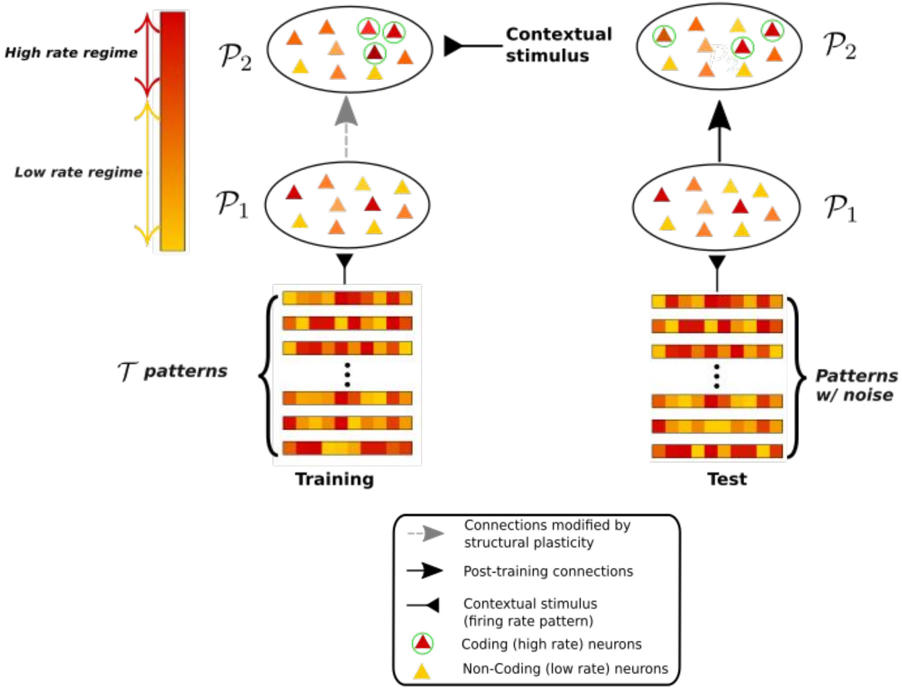
Feed-forward rate-based network with two neuron populations.

Training

- a pattern and a contextual stimuli are injected into the network
- structural plasticity stabilizes connections between neurons at high rate (**stabilization**)
- periodically, non-stabilized synapses are removed and created again randomly (**rewiring**)

Test

- a pattern is injected into the network without the contextual stimulus



Tiddia et al., **Phys. Rev. E (accepted)**, 2024

Impact of synaptic rewiring

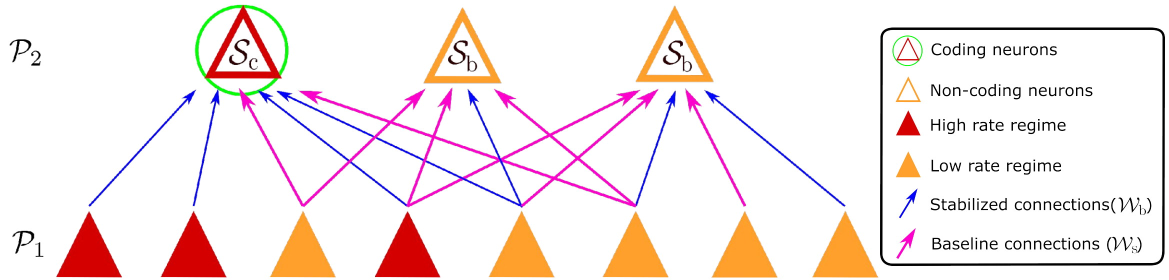
During test, we estimate the signal-difference-to-noise-ratio to evaluate the capability of the network of recognizing the pattern using both **C++ simulations** and a **theoretical framework**.

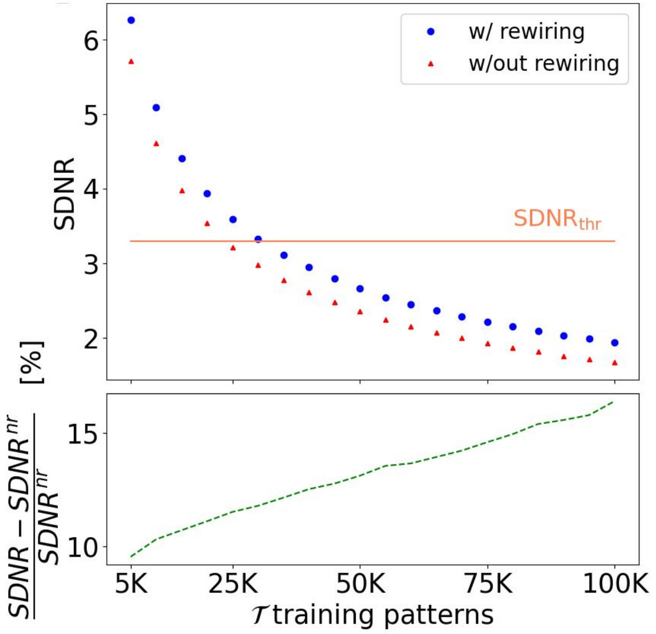
$$\mathrm{SDNR} = \frac{|\langle \mathcal{S}_c \rangle - \langle S_b \rangle|}{\sigma_b}$$

average coding signal

average non-coding
(i.e., background) signal

background standard deviation





Outlook

Simulation technology for neuroscience is advancing to fully exploit next era supercomputers

- With NEST GPU, we can exploit clusters such as LEONARDO to perform very-large-scale simulations
  E.g.: using 96 GPUs, we simulate ~22M neurons and ~240B synapses with RTF ~15

- We are working towards the development of large-scale models able to investigate the role of synaptic processes in high-level cognitive processes

Thank you for your attention!

Bibliography

- Tiddia, Sergi and Golosio, *Theoretical framework for learning through structural plasticity*, Physical Review E (accepted for publication). Preprint doi: arXiv:2307.11735 [q-bio.NC]
- Tiddia, *Large-scale neuronal networks: from simulation technology to the study of plasticity-driven cognitive processes*, 2024. Zenodo. doi: 10.5281/zenodo.10245798
- Golosio, Villamar, Tiddia, Pastorelli, Stapmanns, Fanti, Paolucci, Morrison and Senk, *Runtime Construction of Large-Scale Spiking Neuronal Network Models on GPU Devices*, Applied Sciences, 2023
- Tiddia, Golosio, Fanti and Paolucci, *Simulations of Working Memory spiking network driven by short-term plasticity,* Frontiers in Integrative Neuroscience, 2022
- Tiddia, Golosio, Albers, Senk, Simula, Pronold, Fanti, Pastorelli, Paolucci and van Albada, *Fast simulation of a multi-area spiking network model of macaque cortex on an MPI-GPU cluster,* Frontiers in Neuroinformatics, 2022
- Golosio, Tiddia, De Luca, Pastorelli, Simula, Paolucci, *Fast Simulations of Highly-Connected Spiking Cortical Models Using GPUs*, Frontiers in Computational Neuroscience, 2021

Backup

NEST GPU: scaling test

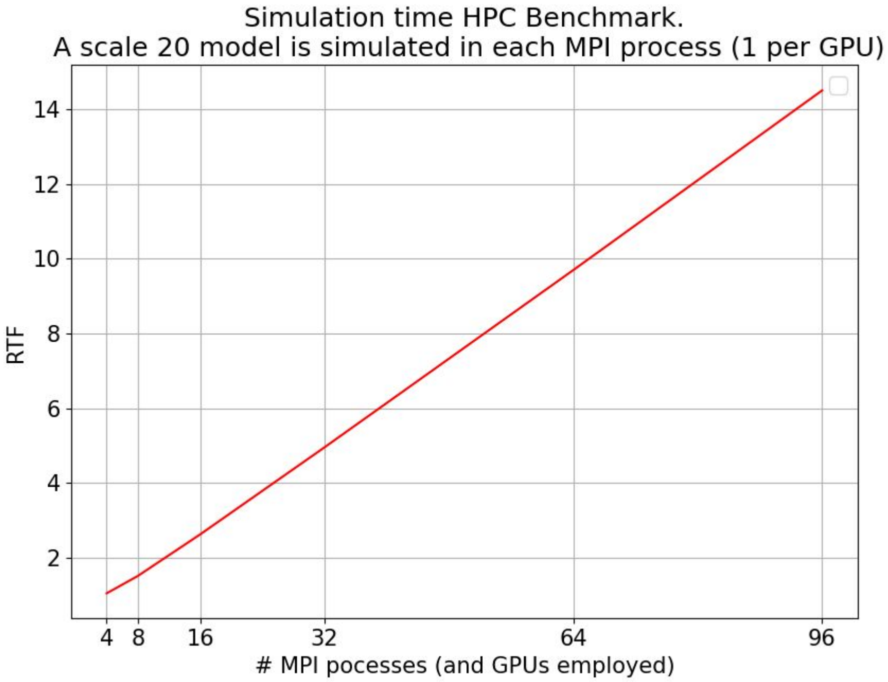
Scaling test of NEST GPU with LEONARDO: network with 225k neurons and 2.5B connections per GPU employed.

Connections are instantiated between neurons that mainly belong to different blocks of the models (i.e., GPUs).

Average neuron firing rate: 8 Hz

Using 96 GPUs:

- 22M neurons
- 243B synapses
- RTF ~15 (i.e., a second of neural activity is simulated in around 15 seconds)

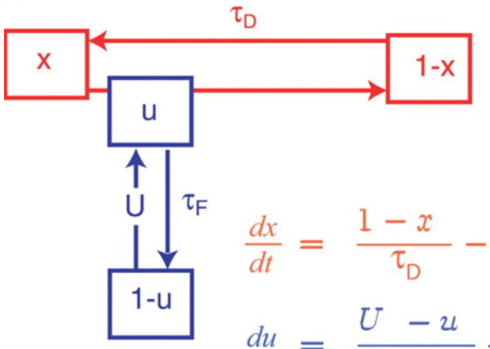


How to model short-term plasticity

How can we model a synaptic mechanism?

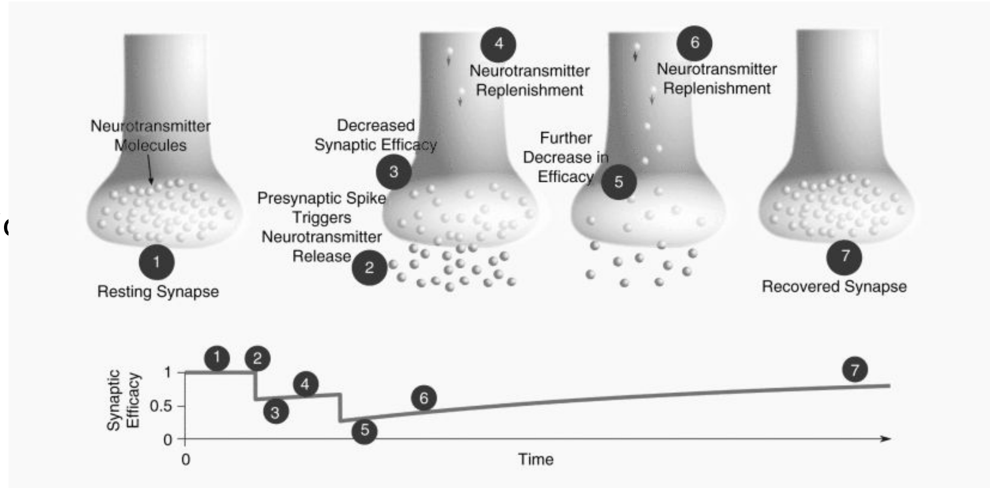
An example we worked on: short-term plasticity

Short-term plasticity model modulated the synaptic efficacy taking into account the dynamics of neurotransmitters and presynaptic calcium.



$$\frac{dx}{dt} = \frac{1-x}{\tau_D} - u\,x\,\overbrace{\delta(t-t_{sp})}^{\text{Spike at time } t_{sp}}$$

$$\frac{du}{dt} = \frac{U-u}{\tau_F} + U\,(1-u)\,\delta(t-t_{sp})$$



x : (normalized) amount of neurotransmitters ready to be released from the synaptic vesicles

u : (normalized) presynaptic calcium concentration
U: baseline value of u

Synaptic efficacy modulated by u(t)x(t)

Effects of STP on postsynaptic potential (PSP)

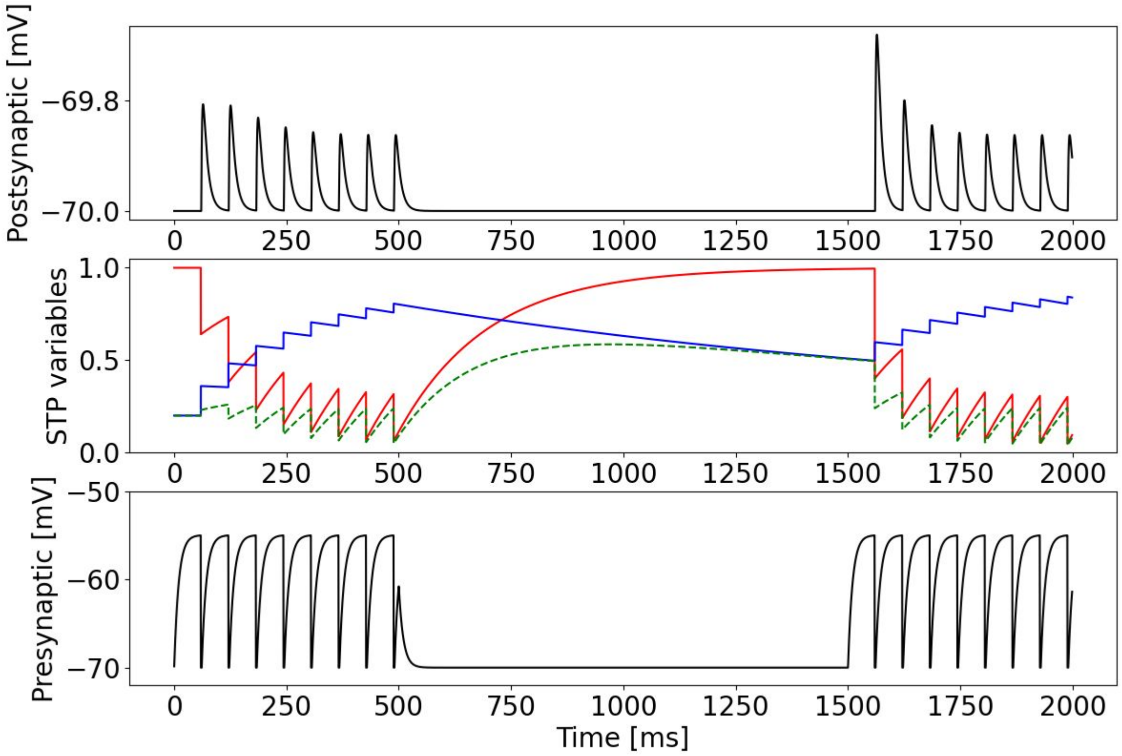
An example of synaptic facilitation (i.e. $\tau_f \gg \tau_d$). Neurotransmitters recovers faster, whereas calcium concentration decrease slowly.

STP variables: — u — x --- ux

$$\frac{du_{i,j}}{dt} = -\frac{u_{i,j} - U}{\tau_f} + U(1 - u_{i,j}) \sum_s \delta(t - t_s^{(i)})$$

$$\frac{dx_{i,j}}{dt} = \frac{1 - x_{i,j}}{\tau_d} - u_{i,j} x_{i,j} \sum_s \delta(t - t_s^{(i)})$$

$$J_{i,j}(t) = J_{i,j}^{(abs)} u_{i,j}(t - \hat{\delta}_{i,j}) x_{i,j}(t - \hat{\delta}_{i,j})$$



Working Memory Spiking Network Model

Network of 10000 LIF neurons with exponential postsynaptic currents.
Neuronal (subthreshold) dynamics for a neuron *j* :

$$\tau_m \frac{dV_j}{dt} = -V_j + R_m(I_j^{exc} + I_j^{inh} + I_{ext,j})$$

Contribution driven by synaptic connections:

$$\tau_{exc} \frac{dI_j^{exc}}{dt} = -I_j^{exc} + \sum_i \alpha J_{i,j}(t) \sum_s \delta(t - t_s^{(i)} - \hat{\delta}_{i,j})$$

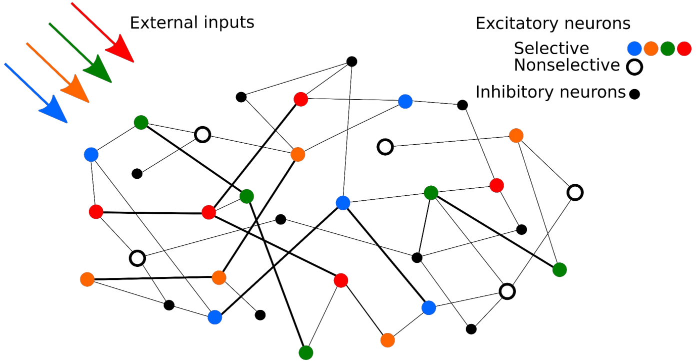
Short-Term Plasticity

$$J_{i,j}(t) = J_{i,j}^{(abs)} u_{i,j}(t - \hat{\delta}_{i,j}) x_{i,j}(t - \hat{\delta}_{i,j})$$

$$\tau_{inh} \frac{dI_j^{inh}}{dt} = -I_j^{inh} + \sum_i \alpha J_{i,j} \sum_s \delta(t - t_s^{(i)} - \hat{\delta}_{i,j})$$

External background current (Gaussian white noise):

$$I_{ext,j}(t - \hat{\delta}_j) = \mu_{ext} + \sigma_{ext} G_k$$



Excitatory neurons organized in 5 selective populations. Neurons of the same selective population have stronger $J^{(abs)}$.

Inhibitory neurons have non-specific connectivity.

SDNR and memory capacity

Why do we calculate SDNR?

SDNR can be linked to the memory capacity by defining the probability of correct recall of a learned pattern (P_{C})!

Setting a threshold probability (e.g., 95%), we can derive $\mathrm{SDNR}_{\mathrm{thr}}$

From the SDNR threshold that we can estimate the **memory capacity** as the maximum amount of patterns that can be stored in the network having at least a SDNR equal to $\mathrm{SDNR}_{\mathrm{thr}}$

$$\mathrm{SDNR} = \frac{|\langle \mathcal{S}_{\mathrm{c}} \rangle - \langle \mathcal{S}_{\mathrm{b}} \rangle|}{\sigma_b}$$

$$P_{\mathrm{C}} \simeq \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\mathrm{SDNR}}{\sqrt{8}}\right)\right]$$