Research · Report NR-06 · 13–27 September 2026

Millions of neurons with free tools

How large a simulated brain can be on a free GPU, and how the neuron-based organism went from 236 to 27.5 hours for its safety test.

Gioele Chiarolanza Nadirion

In progress free Colab and Kaggle · Tesla T4 GPUs · no spending

4 M	236 → 27.5	4/4	€0
neurons and 1.4 billion synapses on a free GPU	estimated hours for the T-OP test on the neuron-based organism	equivalence criteria of the fused-kernel engine on Colab	spent on computing

SUMMARY

The long-term goal is to test the theory on millions of neurons with free tools. A first benchmark on a Tesla T4 GPU in Google Colab measured memory and speed at 2, 3 and 4 million neurons: the infrastructure holds up to 4 million with 1.4 billion synapses, at about 25 real seconds per biological second. The benchmark does not use validated circuits and demonstrates capacity, not readiness for experiments. The second step is concrete: the organism on published neurons (NR-05) cost 63 ms per organism per step in Brian2, and the T-OP safety test would have taken 236 hours. Rewritten in PyTorch with the same equations and then with fused kernels, it repeated the verdicts already recorded (criteria P1–P5 passed) and the test is now estimated at 27.5 hours on Kaggle's two free T4s. Along the way a duration estimate turned out to be wrong, because it measured throughput instead of latency, and it was corrected before the test was launched. The first session started on 27 September 2026.

1 The scale benchmark

A network of neurons with asymmetric topology, in PyTorch, on a T4 with 15.5 GB of free memory. The event count was checked against a full pass, with zero error; firing rates do not change between 2 and 4 million neurons.

neurons	synapses	peak memory	ms per step	events per second
2 million	700 million	6.36 GB	5.63	618 million
3 million	1.05 billion	9.22 GB	8.72	—
4 million	1.40 billion	12.08 GB	12.55	555 million

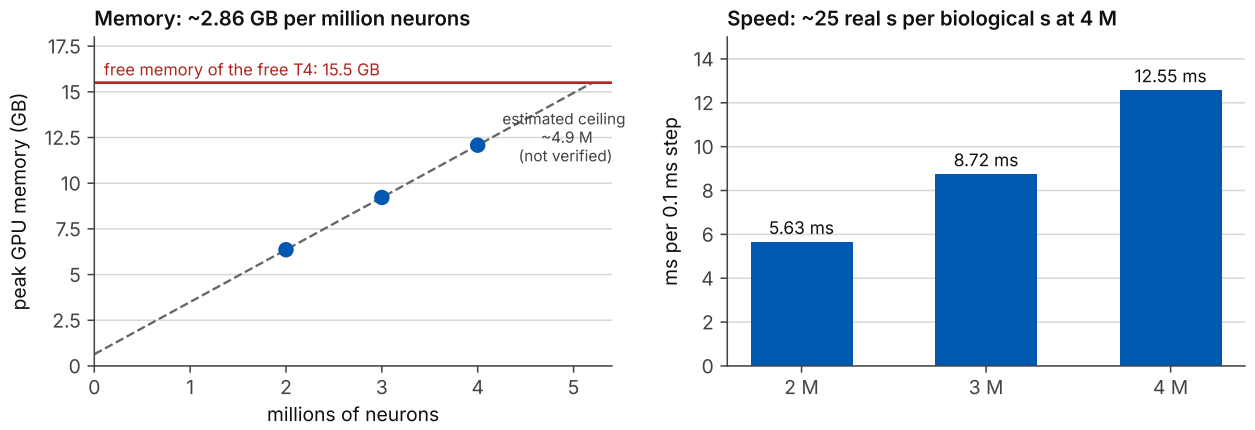


Figure 1. Left, the peak GPU memory at 2, 3 and 4 million neurons, with the line through them: about 2.86 GB per million neurons, that is about 8.2 bytes per synapse. The extrapolated ceiling, about 4.9 million, is not verified because the memory reserved by the allocator is not measured. Right, the time per simulation step. Reference on CPU (Ryzen 3 2200G): 219 ms per step at 2 million.

With one plasticity trace per synapse the ceiling would fall to about 3.3 million. The benchmark says **nothing** about the theory: random connectivity, a single seed, 500 biological milliseconds, no body and no validated circuit.

2 From the benchmark to the organism: the engine for D5

D5 is the step in which T-OP and tests F1–F4 are repeated on the organism on published neurons. With Brian2 on the CPU the whole organism cost 63 ms per organism per step: the T-OP protocol (12 seeds, 750 episodes of 1,200 steps, 5 variants) would have taken 236 hours, and free Kaggle sessions last at most 12.

The **same published equations**, with the same connections and constants, were rewritten in PyTorch for the GPU: many organisms in the same simulation, CUDA graphs, basal-ganglia decisions and amygdala probes collected in batches. The engine had to earn its adoption by repeating the verdicts already recorded with Brian2, with criteria written in advance.

criterion	what it repeats	first version (25 September)	fused kernels (27 September)
P1 insula	D4 bands	passed	passed (Colab)
P2 LC	rate ≤ 2 Hz	**failed at the margin**: 2.01 Hz	passed: error 1.89 Hz, correlation 0.990 (Colab)
P3 workspace	criterion L3	passed	passed (Colab)
P4 organism	criterion L5, means within 9% of Brian2	passed	passed (Colab)
P5 cost of T-OP	under 30 hours	**failed at the margin**: 30.0 hours	passed: 27.5 hours (Kaggle)
P6 basal ganglia	C0–C2 and choice curve	passed	untouched by the fused kernels
P7 amygdala	calibration and A0–A5	passed	untouched by the fused kernels

The first version was adopted by the author with the two marginal failures on record, on condition that the cost be reduced before T-OP. An important observation: the engine reproduces Brian2 **even in its defects** (LC gain, bistable workspace), a sign that the change of simulator neither adds nor removes properties of the model. On criterion P2, the same model passes or fails depending on the realisation of the noise: the 2 Hz threshold is at the edge of its variability, and the earlier failure stays on record.

A wrong duration estimate

On 26 September a cost reduction had brought the T-OP estimate down to 19.8 hours. Before writing the bench we realised that **the estimate was wrong**: it measured throughput, that is how many organism-steps the GPU executes per second with 256 organisms together. But T-OP has only 60 runs, each of 900,000 steps **in sequence**: 256 organisms cannot be run in parallel. With few organisms per process the time of a step is dominated by latency, about 0.15–0.2 s, and the true duration was 40–50 hours. No run was launched with the wrong estimate. Since then duration estimates use the **latency rule**: number of steps times the slowest step measured.

To reduce latency, the second version of the engine rewrites insula, LC and workspace as a single computation per window on the GPU (fused kernels), with the same equations and the same order of operations. The latency of a step with 15 organisms falls from 187 to 66 ms, and one T-OP run from 47 to 16.4 hours. The engine repeats the criteria: all passed, with P5 measured by the latency rule.

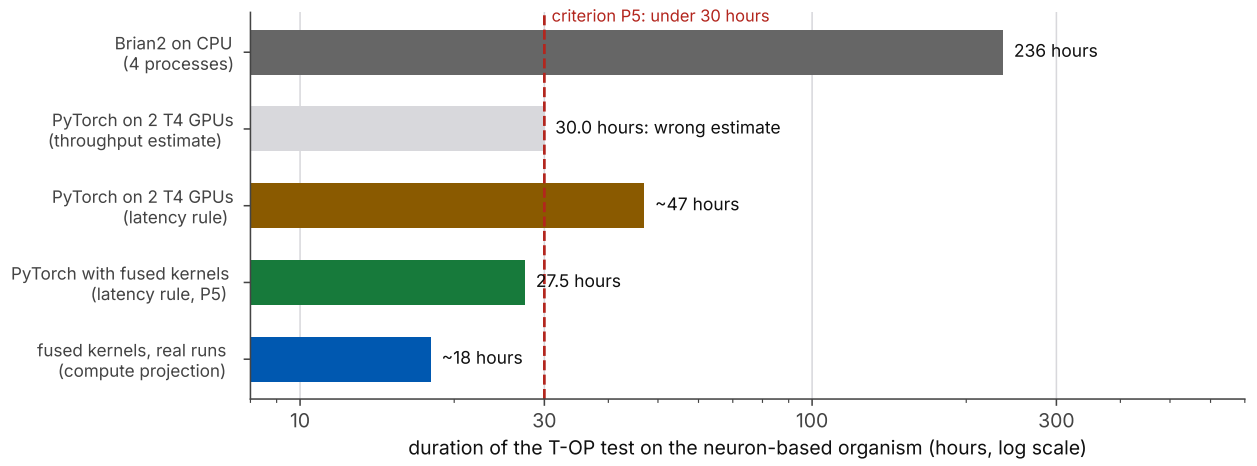


Figure 2. Duration of the T-OP test on the organism on published neurons. The PyTorch engine's estimate made by measuring throughput (30.0 hours) turned out to be wrong; the latency rule, adopted on 26 September, gives about 47 hours without fused kernels and 27.5 with them (criterion P5: under 30). In the first real runs a step takes 55–83 ms, and the computation required projects to about 18 hours, that is three free sessions. Log scale.

3 Status of the test

The first T-OP session on the neuron-based organism (notebook version 45, two free T4s on Kaggle) was launched on 27 September 2026, with protocol 3 of the test on the integrated organism: seeds, thresholds and variants unchanged. Two decisions of the author were taken before the run, both for measured reasons: the organism is the one used in the F tests, and the adversarial condition of O1 uses the mean vigilance before the command instead of that of every step, because with the neuron LC vigilance fluctuates between 11 and 18.5 Hz due to the noise of the window. After 4.9 hours of the first session the four series of runs were at episode 190 of 750; later sessions resume from the saved state. Outcome: protocol 3 stopped at 3 of 7 criteria because the instruments were not sensitive enough; after a diagnosis and a calibration, protocol 4 passed 7 of 7 criteria on 5 October 2026 (report NR-11).

4 What it shows and what it does not

It shows that the scale of millions of neurons can be reached for free, with a time constraint, and that a validated organism can change simulator without changing behaviour, if this is checked with criteria written in advance.

It does not show that the organism with millions of neurons behaves like the small one: scaling requires calibration first, and a workspace architecture that switches off (NR-05).

Limitations. The scale benchmark uses random connectivity and a single seed. The T-OP duration estimates are based on the slowest step measured and must be checked in the real runs. Free resources have session and weekly quota limits that lengthen real times.

References

- Paszke, A. and colleagues (2019). PyTorch: an imperative style, high-performance deep learning library. *NeurIPS* 32.
- Brunel, N. and Wang, X.-J. (2001). *J. Comput. Neurosci.* 11, 63.
- Stimberg, M., Brette, R. and Goodman, D. F. M. (2019). Brian 2. *eLife* 8, e47314.
- Internal documents: whitepapers WP-04 and WP-14; `benchmarks/colab_1m_nadirion.py`; `docs/FASE2_PIANO.md` (decisions G-pp, G-qq, G-rr, G-ss, G-tt); transcripts in `digital-brain/kaggle/risultati/`.

How to cite

Chiarolanza, G. (2026). Millions of neurons with free tools. Nadirion, research report NR-06, 13–27 September 2026. https://nadirion.eu/en/research/free-gpu-computing/

Nadirion technical report, not peer-reviewed. The criteria of every test were written before running it; negative results are reported like positive ones. The tests measure functional roles.