The Brain Emulation Challenge: A Synthetic Data Platform to Evaluate Reconstruction Methods

Randal A. Koene*, Thomas Liao, John Jose, Trisha Mendoza, Aiden Behler, Jainil Shah, and Albert Huynh
Carboncopies Foundation for Brain Research – braingenix.carboncopies.org / rkoene@carboncopies.org

PSTR197.18 / YY5
SfN 2025

Introduction

Brain data collection techniques such as high-throughput electron microscopy provide within-specimen high resolution neural circuit insights, such as connectomes of Drosophila M. (1, 2), Mouse (3) and Zebrafish (4). Studies by Shiu et al (5) and Lappalainen et al (6) demonstrate attempts to reconstruct functional models based on such data. Typically implemented using leaky integrate-and-fire excitatory and inhibitory point neurons, resulting models were able to make useful predictions for future experimental verification despite ignoring many constraints and contributors (e.g. neuromodulation, gap junctions, individual receptor dynamics).
Still, validated working emulations at the neural circuit level, created by converting brain data are incomplete and rare. A key limitation is the **absence of ground-truth systems** that allow rigorous evaluation of candidate emulation methods.

In machine learning (AI), augmenting training data with synthetic data is frequently used to improve algorithms and models. We present the **Brain Emulation Challenge**, a community platform designed to accelerate progress in neural system identification through structured challenges based on virtual brain data from synthetic fully-known ground-truth cells and circuits.

We generate in-silico neural circuits with embedded structure-function mappings to be "discovered", as a cost-effective protocol to pre-test a system identification method's ability to identify, reconstruct and emulate the meaningful behavioral functions of neuronal systems, under constraints of data typically obtained. In addition to rapid iterative testing, we aim to support the study of neural coding, functional constraints and goal-specific separation of scales in the complex nonlinear dynamics of a brain.

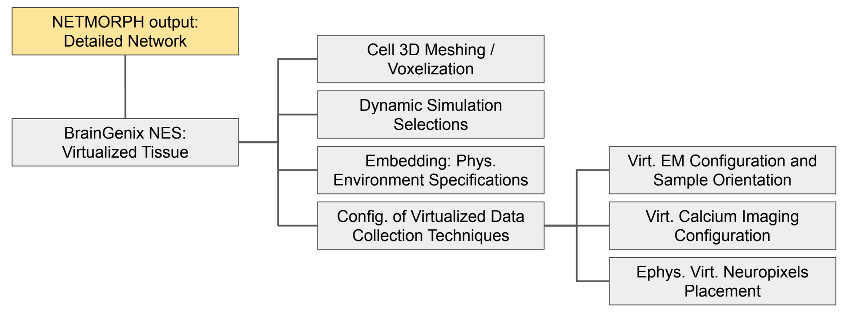
Example synthetic ground-truth challenge

Context-stimulus conditioned fear response:
The synthetic ground-truth system is to contain a hidden purposeful function that represents a brain sample preparation from a putative animal that has learned behavior represented by the following table.

Context	Stimulus	Response Behavior
no light input (in darkness)	no sound input (quiet)	remain motionless (at rest)
no light input (in darkness)	sound detected	move (hide)
light input (daylight)	no sound input (quiet)	remain motionless (freeze)
light input (daylight)	sound detected	move (forage)

Neural morphogenesis to conn. reservoir

An extended version of ***Netmorph*** (7) applies neural morphogenesis to generate the detailed synthetic neuronal structure, resulting in a connectome reservoir with candidate synaptic sites. The ***BrainGenix-NES*** platform (8) fleshes out the nanoscale morphology of this structure, selects synapses, and translates pairs of connected cell types and synapse area into proportions of specific ion channels and corresponding receptors with constrained maximum activated conductivity. An STDP model is used to entrain a behavior-specific circuit function for corresponding "in-domain" context-stimulus pairings. For the challenge, we consider this the *hidden* (unknown to the challenge taker) purposeful function of the synthetic ground-truth system.



The published version of Netmorph simulates growth-cone behavior, using stochastic and phenomenal descriptions of elongation, turning and branching to produce realistic axonal and dendritic morphologies, including neurite curvature, and to identify likely candidate sites for synapses. It has been shown to generate realistic cell-type specific morphologies (e.g. pyramidal neurons). Our extension of Netmorph adds chemo-attraction/repulsion so that morphogenesis of stereotyped network architectures is possible (regions, layers, cell type connectivity).

Reservoir phase: To build a synthetic system that is able to contain a target function, we specify cell population sizes, distribution and general architecture, along with cell type specific outgrowth parameters. We then run a batch of outgrowth simulations using different randomized initial states and retain those resulting networks that contain a sufficient number of candidate synaptic sites.

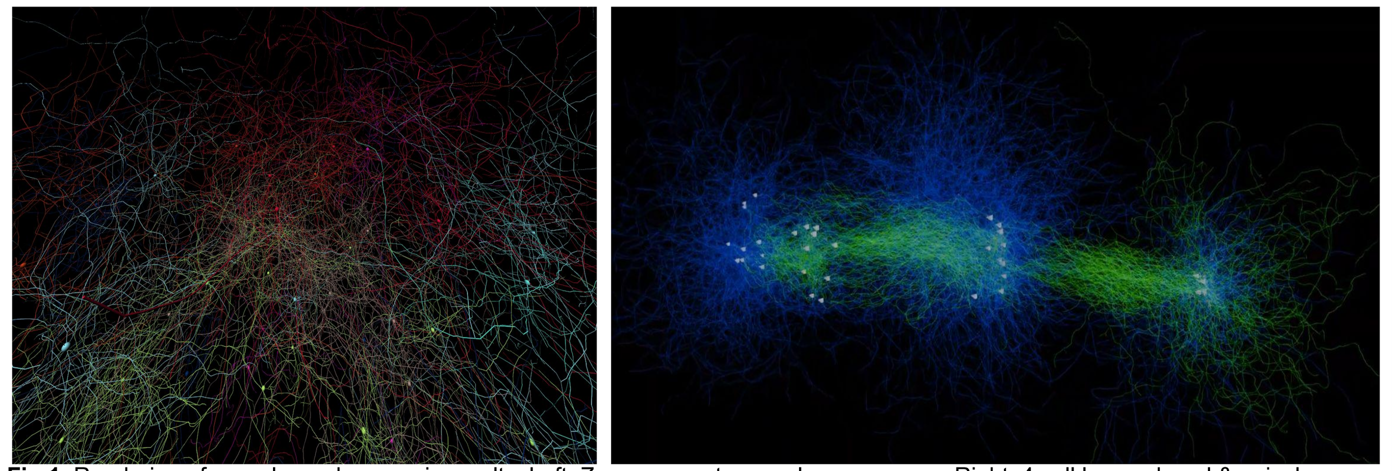


Fig.1: Rendering of neural morphogenesis results. Left: Zoom excerpt one color per neuron. Right: 4 cell layers, basal & apical dendrites blue, axons green.

Function embedding & virtual data acquisition

Connectome phase: The resulting structures produced by Netmorph are ingested by BrainGenix-NES. Realism features are selected and applied. Depending on the needs of a challenge, these features include characteristic variation in neurite compartments, dendritic spine morphology, and physical properties of extracellular materials that affect experimental measurements. In this phase, we also choose the dynamic models for neural response and conductance, and receptor dynamics. Candidate synapses are pruned and tuned while presenting training stimuli to the "connectome reservoir" until the desired target function is expressed (e.g. "associative memory" via attractor dynamics using recurrent excitatory and inhibitory cell connections). For a more difficult challenge, the training stimuli can contain confounders, so that the target function of the resulting network is embedded in and affected by a larger nonlinear dynamic system.

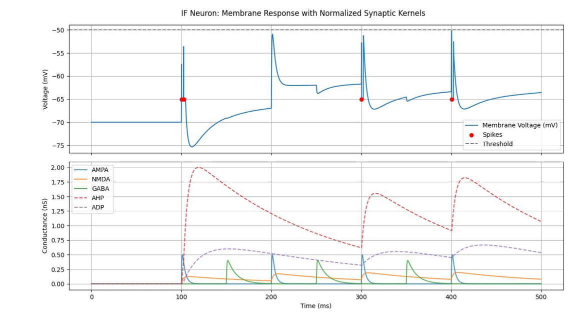


Fig. 2: Membrane potential of a pyramidal-like neuron (top) and contributing conductance kernel responses (bottom). Input received at AMPA and NMDA receptors (t = 100, 200, 300, 400 ms), at GABA receptors (t = 150, 250, 350 ms).

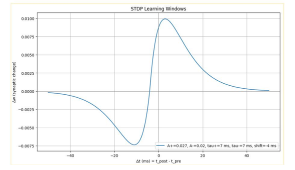


Fig. 3: STDP for pre-post time differences.

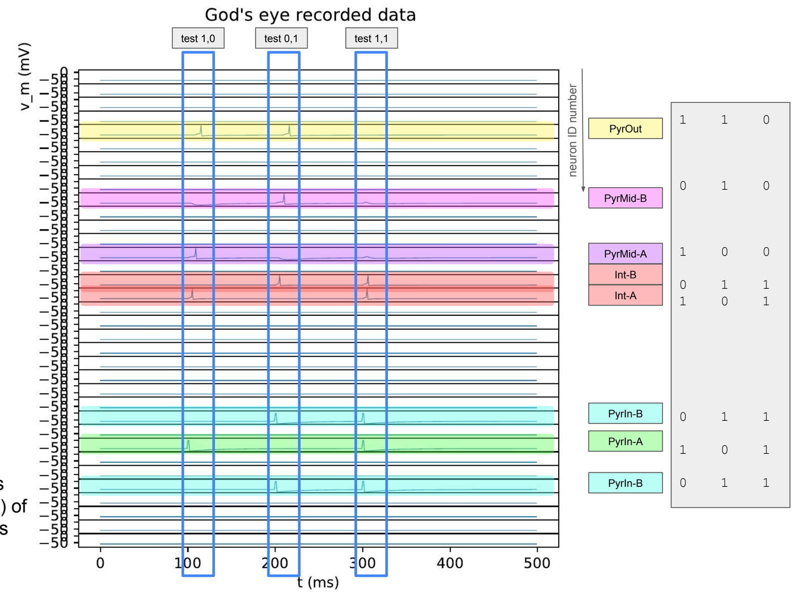


Fig. 4: In-domain synchronous context/stimulus pairing analysis (select cells in population layers) of synthetic ground-truth responses after STDP training. Successful target behavior.

Virtual data collection involves application of "in-domain" stimulation trials (simulated multi-frequency optogenetic or patch-clamp), followed by application of "out-of-domain" stimulation trials, during data acquisition subjected to simulated physics of device and medium. At present, the data acquisition modalities supported by BrainGenix are:

- ➢ High-throughput volume electron microscopy (simulating FIBSEM).
- ➢ Neuropixel-like extracellular electrophysiology & simulated patch-clamp.
- ➢ Calcium imaging. (Extending simulated optical techniques is presently ongoing.)

Acquisition phase: An experiment is defined and carried out via stimulation and measurement. Simulation of measurements begins by configuring devices. Orientation, placement, field of view, focus, depth for imaging. Placement of recording electrodes and recording sites on those electrodes. Cell types / volumes targeted by fluorescent indicators. Secondary configuration involves choosing type and degree of measurement artifacts (e.g. noise sources, focus aberration, sample prep errors such as folds/tears or the diffusion of contrast agents). The resulting multi-modal data stack can be augmented with expected post-processed data for researchers who do not intend to address that stage (e.g. connectome database with identified cell types and locations, Ephys post-processing or spike sorting). All of this data and the list of realism specifications applied – *but not knowledge of the hidden cognitive/behavioral function* – becomes the challenge data set.

A specific challenge consists of data sets from a batch of M similarly configured synthetic ground-truth examples. The same batch can be used to test, evaluate and contrast multiple approaches by multiple researchers.

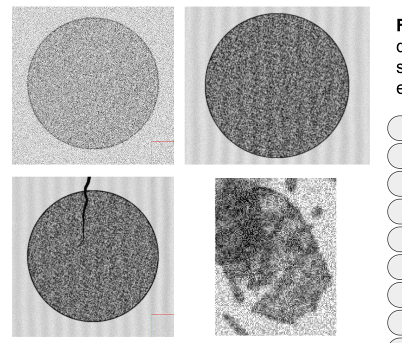
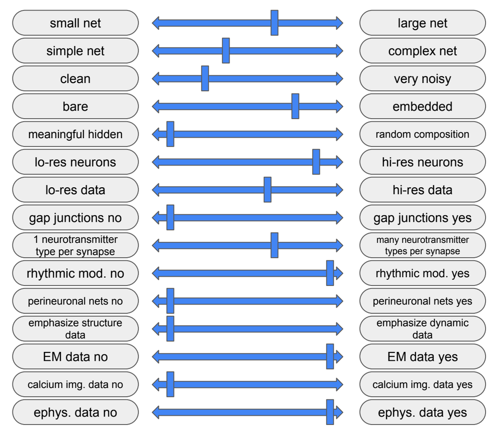


Fig. 5: Examples of acquisition artifacts applied to virtual EM voxelization & rendering (beam noise, focus, interference patterns, tears) and their composition.

Fig. 6: A subset of the many configuration parameters that determine the realism and difficulty of a generated synthetic ground-truth challenge. A virtual laboratory with exploratory control, one variable at a time.



simplified representation, e.g. "noisy" data is actually of multiple types

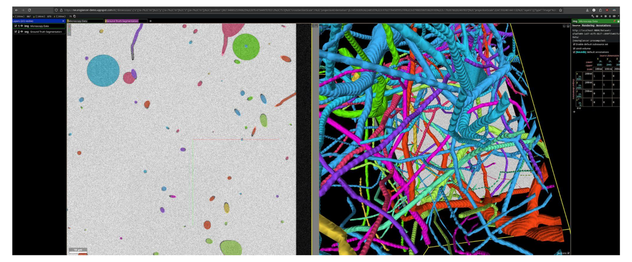


Fig. 7: Exploring segmentation & meshing of synthetic ground-truth aligned EM volume stacks in Neuroglancer.

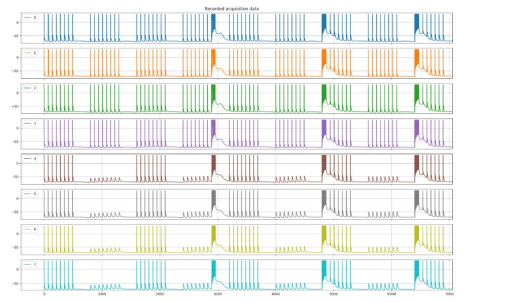


Fig. 8: Post-processed virtual patch-clamp if selected cells in synthetic ground-truth during in-domain context-stimulus testing.

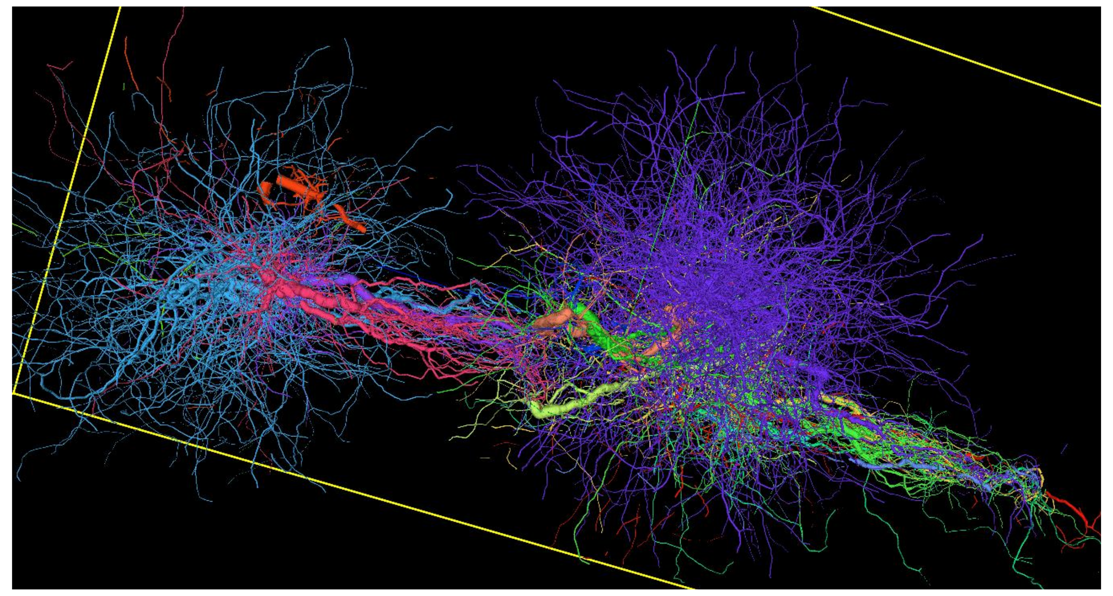


Fig. 9: Neuroglancer inspection highlighting subset of segmented & meshed cells following volume EM reconstruction from virtual data generated for the synthetic ground-truth system.

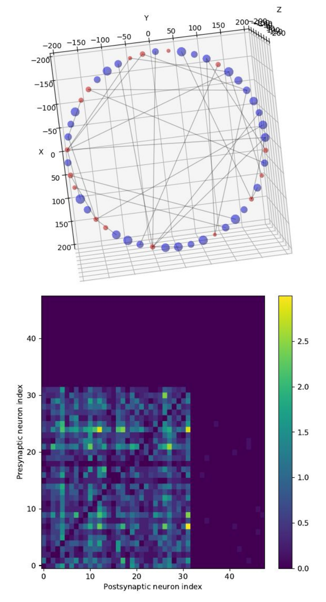


Fig. 10: Displays of connectivity (top) & median summed connection strengths (bottom) from exported connectome database.

Evaluation of attempted emulation

Our goals for emulation prioritize discovery and replication of cognitive or behavioral function. Behavioral success metrics are the primary benchmark for the evaluation and score of a result achieved by any attempt to reconstruct, functionalize and emulate based on available brain data. ***Can your approach discover the correct purposeful function hidden in the data acquired from a synthetic ground-truth system?***

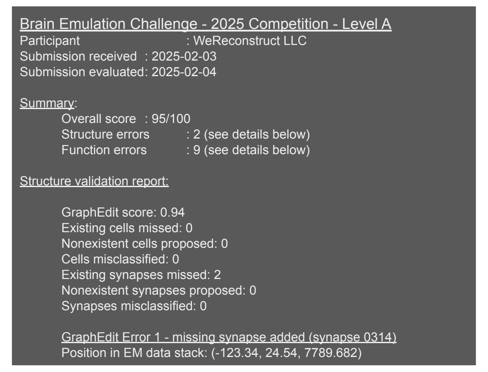
Dynamic similarity overall and at the component level is evaluated and scored via secondary benchmark metrics.

Consider an attempt at system identification that involves a model initialization constrained by an identified connectome followed by a process of functionalization, e.g. Lappalainen et al (6). Researchers can submit their attempts to our evaluation through comparison with the actual synthetic ground-truth either by connecting to our ***BrainGenix-EVM*** server API, or by delivering the resulting model.

Evaluation includes these principal benchmarks:

- Structural similarity metrics (Edit distance, Topological similarity, Synaptic matching, Neuron Type Distribution).
- Primary behavioral success metrics (In-domain context/stimulus response, Out-of-domain response).
- Secondary dynamic similarity metrics (+ auto-insights generated):
  - ➢ Peri Spike Time Histogram (PSTH) differences between ground-truth and submitted.
  - ➢ KS ground-truth vs submitted & context/stimulus KS p-value & Wasserstein.
  - ➢ Inter-spike interval (ISI) per neuron.
  - ➢ Fano-factor per neuron.
  - ➢ Membrane potential auto-excerpted difference highlights.
  - ➢ EPSP/IPSP count per context/stimulus per neuron.
  - ➢ Cross-correlograms neuron-to-neuron for GT & SUB spikes.
  - ➢ Coincidence neuron-to-neuron for GT & SUB spikes.
  - ➢ Van Rossum spikes for each context/stimulus pair.
  - ➢ Multi-scale spike correlation for each context/stimulus pair.
  - ➢ Multi-scale spike correlation per neuron.
  - ➢ Schreiber similarity for each context/stimulus pair.
  - ➢ Granger causality of neuron spike trains for each context/stimulus pair.

Fig. 11: Example of structural metrics report.


Brain Emulation Challenge - 2025 Competition - Level A
Participant : WeReconstruct LLC
Submission received : 2025-02-03
Submission evaluated: 2025-02-04

Summary:
Overall score : 95/100
Structure errors : 2 (see details below)
Function errors : 9 (see details below)

Structure validation report:

GraphEdit score: 0.94
Existing cells missed: 0
Nonexistent cells proposed: 0
Cells misclassified: 0
Existing synapses missed: 2
Nonexistent synapses proposed: 0
Synapses misclassified: 0

GraphEdit Error 1 - missing synapse added (synapse 0314)
Position in EM data stack: (-123.34, 24.54, 7789.682)

InputPattern	TP	FN	TN	FP	Accuracy	Sensitivity	Specificity	Events
Pattern 1 (A=0,B=0)	0	0	100	0	1.00	0.0	1.0	100
Pattern 2 (A=1,B=0)	100	0	0	0	1.00	1.0	0.0	100
Pattern 3 (A=0,B=1)	100	0	0	0	1.00	1.0	0.0	100
Pattern 4 (A=1,B=1)	0	0	40	60	0.40	0.0	0.4	100
Total	200	0	140	60	0.85	1.0	0.7	400

InputPattern	TP	FN	TN	FP	Accuracy	Sensitivity	Specificity	Events
Pattern 1 (A=0,B=0)	0	0	100	0	1.00	0.0	1.0	100
Pattern 2 (A=1,B=0)	100	0	0	0	1.00	1.0	0.0	100
Pattern 3 (A=0,B=1)	100	0	0	0	1.00	1.0	0.0	100
Pattern 4 (A=1,B=1)	0	0	0	100	0.00	0.0	0.0	100
Total	200	0	100	100	0.75	1.0	0.5	400

Fig. 12: Example of in-domain context-stimulus paired trials behavior report with 100 trials of each pairing showing true positives, false negatives, true negatives, false positives and overall accuracy score. Comparing ground-truth behavior (top) and an attempted emulation (reconstruction & functionalization, bottom).

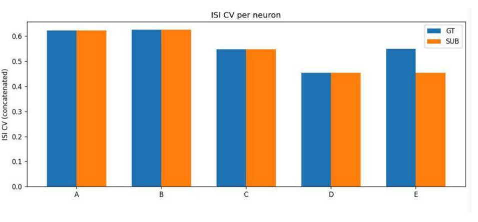


Fig. 13: Example of interspike interval scores for 5 different cell groups (A context sensitive, B stimulus sensitive, E behavior driving). GT=ground-truth.

Fig. 15: Deep inspection of PSPs can be particularly informative about causes of divergence.

pattern	neuron	n_trials	EPSP_GT	EPSP_SUB	IPSP_GT	IPSP_SUB
00	A	100	0	0	0	0
00	B	100	0	0	0	0
00	C	100	0	0	0	0
00	D	100	0	0	0	0
00	E	100	0	0	0	0
01	A	100	0	0	0	0
01	B	100	0	0	0	0
01	C	100	100	100	0	0
01	D	100	100	100	100	100
01	E	100	100	100	100	100
10	A	100	0	0	0	0
10	B	100	0	0	0	0
10	C	100	100	100	0	0
10	D	100	100	100	100	100
10	E	100	100	100	100	100
11	A	100	46	46	50	50
11	B	100	32	32	35	35
11	C	100	149	149	115	115
11	D	100	167	167	100	100
11	E	100	128	159	72	100

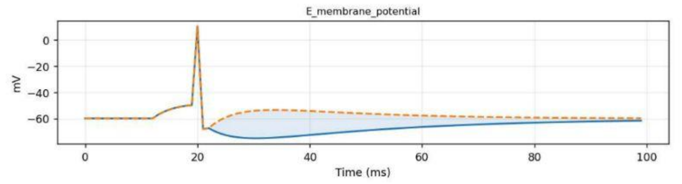


Fig. 14: Zones exhibiting significant dynamic difference are automatically extracted for inspection at suspect cells. (GT in blue.)

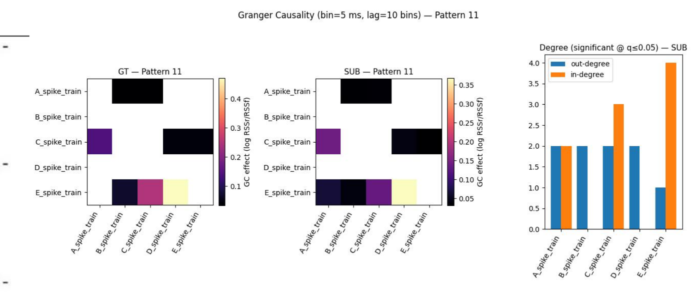
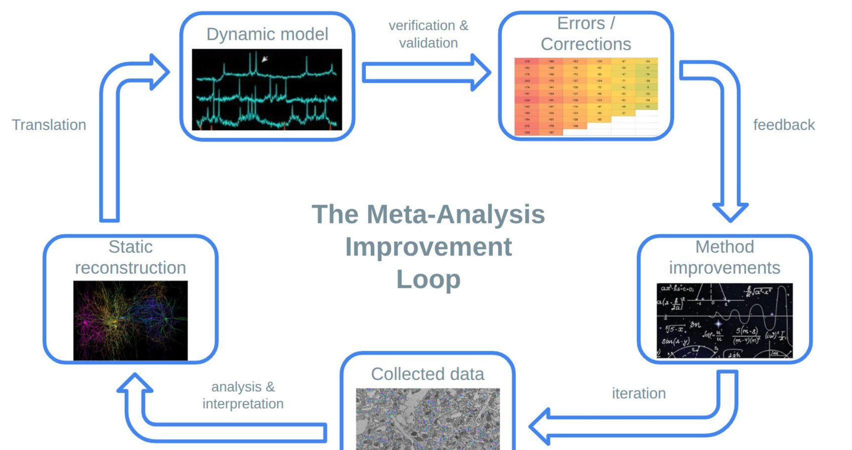


Fig. 16: Example of Granger Causality inspection. Here, highlighting context-stimulus pattern light context + sound stimulus.

Scores are weighted within categories first (behavioral, dynamic, structural). Total score is a weighted combination of category scores. In-domain behavioral weighting >> dynamic weighting to emphasize hidden function discovery over robustness comparisons and out-of-domain behavior. Evaluation intends to score performance, but also to offer insights that allow identification of the cause of significant differences. Our metrics are applied in a manner that suits the researcher's approach to emulation. While there are many possibilities, we presently focus on (a) neuron-by-neuron reconstructions, and (b) black-box I/O reconstructions.

Method testing and method development is an iterative process in which a researcher must isolate and identify systematic problems, edge cases and limitations of their approach. We aim to facilitate the generation of tiered synthetic ground-truth examples and corresponding data sets to enable this process of isolation and identification. Challenges will be offered on scales of size & sophistication.



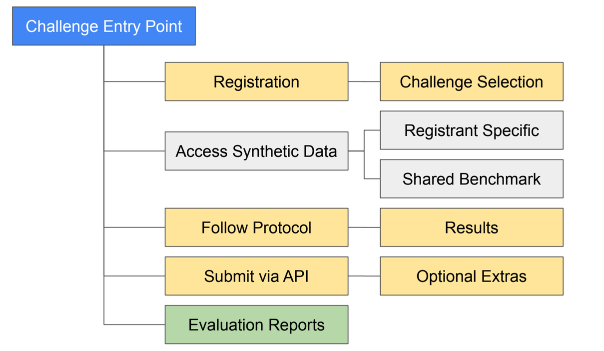


Fig. 17: Representation of the protocol for challenge participation. Yellow are activities expected of the participating researcher.

Conclusions & Future work

Testing approaches on data from synthetic ground-truth challenges can identify, describe and help overcome limitations. An approach that performs well may inspire greater confidence when applied to neuroscience data. Compared with experiment, it is low-cost, fast, and systematic.

Neural coding of function depends on architecture, dynamic characteristics of cell types, and constraining factors. Any combination of these can affect the performance of a particular approach to system identification. As we build out challenges, we intend to explore a terrain of possibilities:

- Basic nonlinear behavior (as in the example shown).
- Auto- and heteroassociative attractors formed by populations with recurrent connectivity.
- Short-term buffers that sustain activity for one-shot learning.
- Dendritic spikes and single-neuron dendritic computation (e.g. for high-speed binaural sound localization).
- Spontaneous recruitment as reservoir computing in an interconnected population of neurons.
- Feature detector composition through simple- to complex-cell layering (e.g. early visual system).
- Cell-specific receptive fields (e.g. place/grid cell expression).
- Intrinsic effects in cortical pyramidal cells (e.g. sustained activity by after-depolarization).
- Columnar or lattice processing.
- Synchronized phase-dependent rhythmic modulation and interaction within neural circuits and populations.

Primary risks & some remedies:

Misdirected performance optimization	*Drawing conclusions based on an "alien" ground-truth*
ill-defined or insufficient success criteria	assumptions used to generate synthetic ground-truth & data
emergent criteria not surrounded by adequate set of validation metrics	
insufficient / insufficiently broad test data	when bio. ground-truth cannot be obtained or is not understood: iterative learning (improve analysis methods, improve synthetic ground-truth)
scoring can lead to algorithm optimization that is misleading	when bio. ground-truth is merely costly/rare: test best-performing method on that, as available

Extension, testing and improvement of the platform is ongoing. Principal axes for improvement are:

- Fine-tuning of virtual EM (e.g. morphology of dendritic spines & synapses, soma morphology and intracellular detailing).
- Additional data acquisition modalities.
- AI-driven configuration & optimization of Netmorph specifications for reservoir generation.
- Performance improvements for EM voxelization & rendering and long duration Ephys in systems with >10k cells.

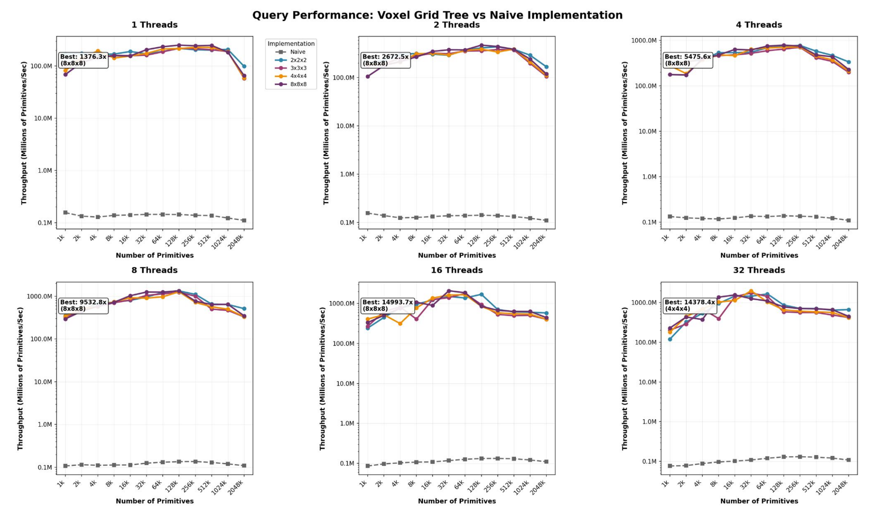


Fig. 18: Ongoing performance improvements. VoxelGridTree reads from the data structure with an adjustable number of threads, primitives, and subdivision sizes. VoxelGridTree is a flat-array spatial tree that pre-allocates its entire node hierarchy in a single contiguous memory block, unlike traditional pointer-based octrees that incur cache misses from pointer chasing. Parent-child relationships are encoded through a deterministic mathematical indexing scheme to eliminates memory fragmentation. Tree navigation becomes simple integer arithmetic.

References

1. Dorkenwald, S., Matsliah, A., Sterling, A.R. et al. (2024). Neuronal wiring diagram of an adult brain. Nature 634, 124–138. https://doi.org/10.1038/s41586-024-07558-y
2. Berg, S. et al. (2025). Sexual dimorphism in the complete connectome of the Drosophila male central nervous system. bioRxiv 2025.10.09.680999; https://doi.org/10.1101/2025.10.09.680999
3. Ding, Z. et al. (2025). Functional connectomics reveals general wiring rule in mouse visual cortex. Nature, 640(8058), 459-469. https://www.nature.com/articles/s41586-025-08840-3
4. Petkova, M. D. et al. (2025). A connectomic resource for neural cataloguing and circuit dissection of the larval zebrafish brain. bioRxiv, 2025.06.10.658982. https://doi.org/10.1101/2025.06.10.658982
5. Shiu, P. K. et al. (2024). A Drosophila computational brain model reveals sensorimotor processing. Nature, 634(8032), 210–219. https://doi.org/10.1038/s41586-024-07763-9
6. Lappalainen, J. K. et al. (2024). Connectome-constrained networks predict neural activity across the fly visual system. Nature, 1–9. https://doi.org/10.1038/s41586-024-07939-3
7. Koene, R. A. et al. (2009). NETMORPH: A framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. Neuroinformatics, 7(3), 195–210. https://doi.org/10.1007/s12021-009-9052-3
8. BrainGenix-NES: https://gitlab.braingenix.org/carboncopies/BrainGenix-NES

For full details and references, please contact us at rkoene@carboncopies.org to receive **our paper preprint.**

Acknowledgements

Work supported by grants from the Silicon Valley Community Foundation (2024) and the Brain Preservation Foundation (2025).

The Carboncopies Foundation is a 501(c)(3) non-profit: **https://carboncopies.org**

All code is available under the AGPL-3.0 license: **https://carboncopies.org/Research/BrainGenix/About/**