Data-Driven Modeling of Brain Circuits and the SONATA Data Format

Eilif Muller, EPFL Blue Brain Project

Anton Arkhipov, Allen Institute for Brain Science

CNS*2018, Seattle, US

Motivation

- **New massive experimental datasets and increasing computing power enable a new generation of large-scale, data-driven, integrative models of brain circuits**

- **Being pursued at both the Allen Institute and the Blue Brain Project**

- **Need for efficient data formats for circuits and simulations**

- **Allen Institute and BBP have collaborated to develop:**
  SONATA Data Format (https://github.com/AllenInstitute/sonata)

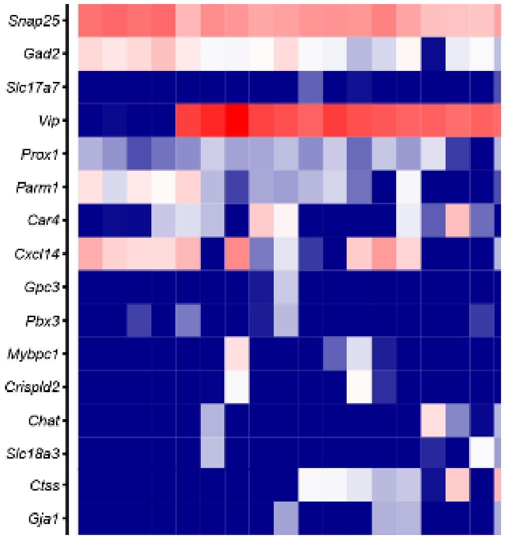
- **An open-source framework for representing neuronal circuits and sims**

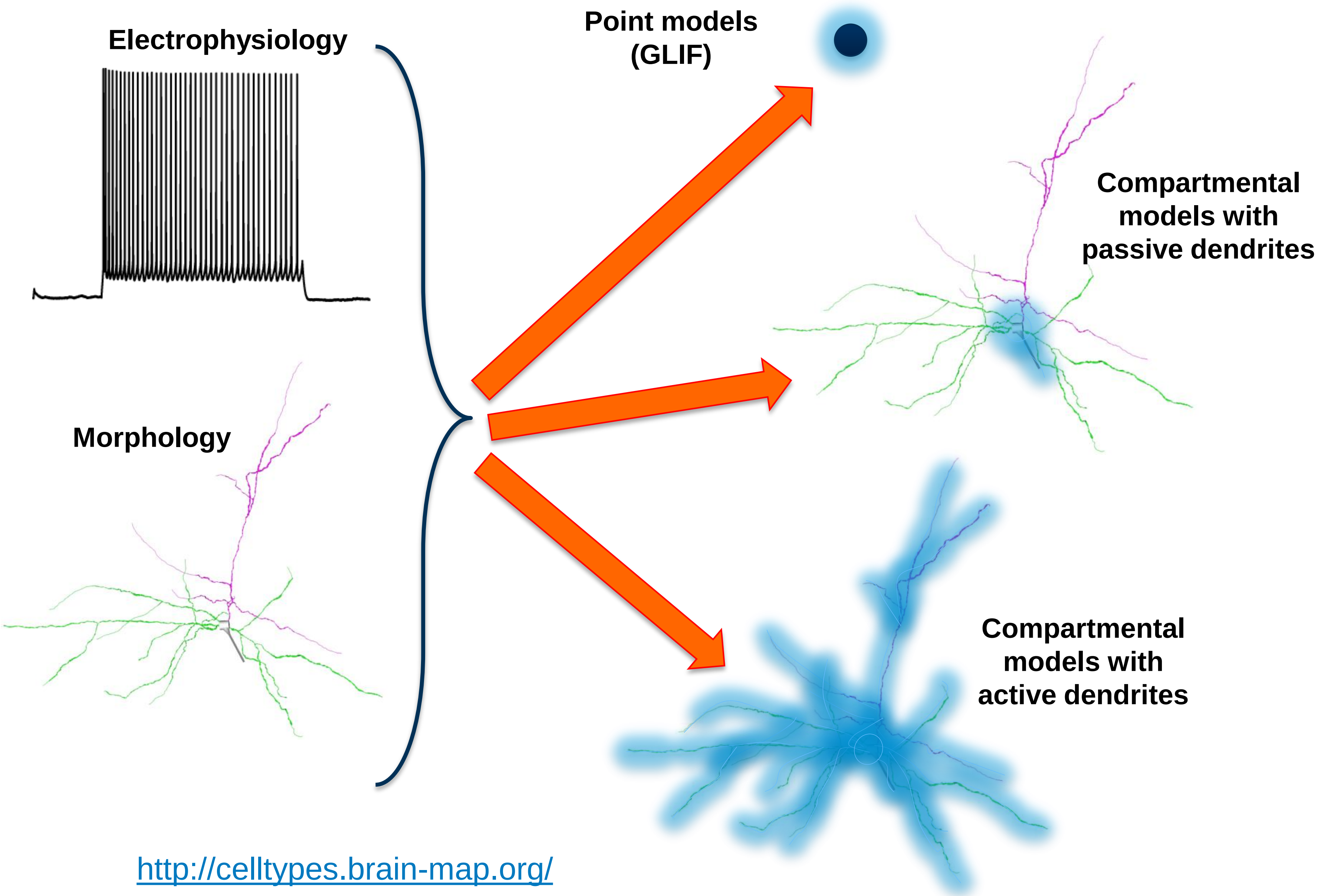
- **Designed to enable high performance: sim, viz, analysis**
  (Based on: HDF5, CSV, JSON, indexing layers)

Allen Institute’s Experiments and Models

Transcriptomics, Morphology, Electrophysiology, and Models of Single Neurons from Mouse and Human Cortical Tissue

Transcriptomics

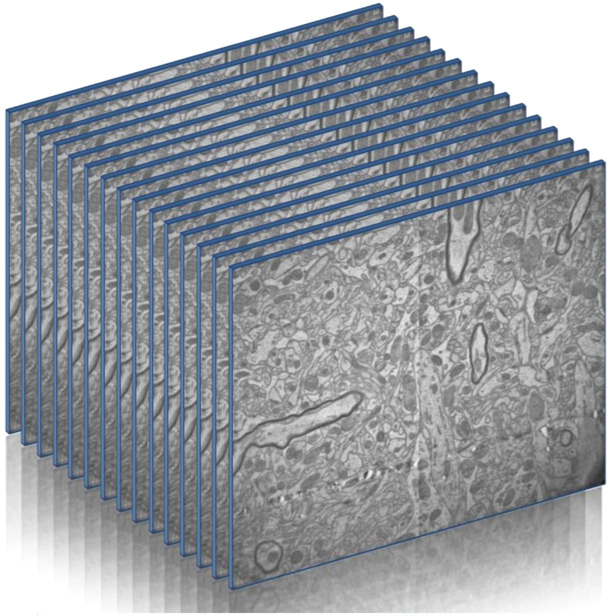


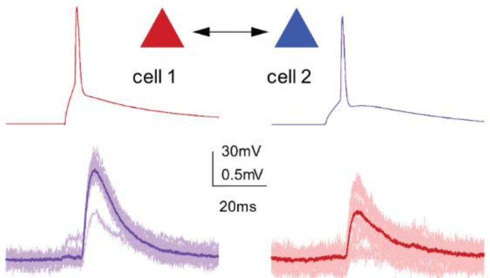


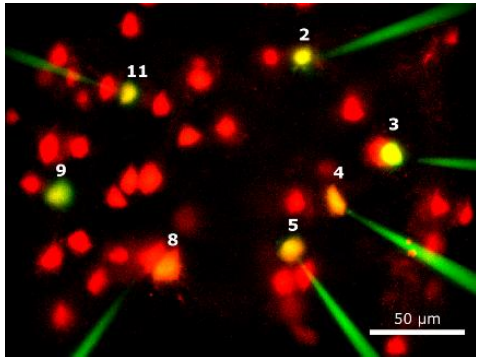
http://celltypes.brain-map.org/

Microscale and Mesoscale Connectivity in the Cortex, Reflected in Network Models

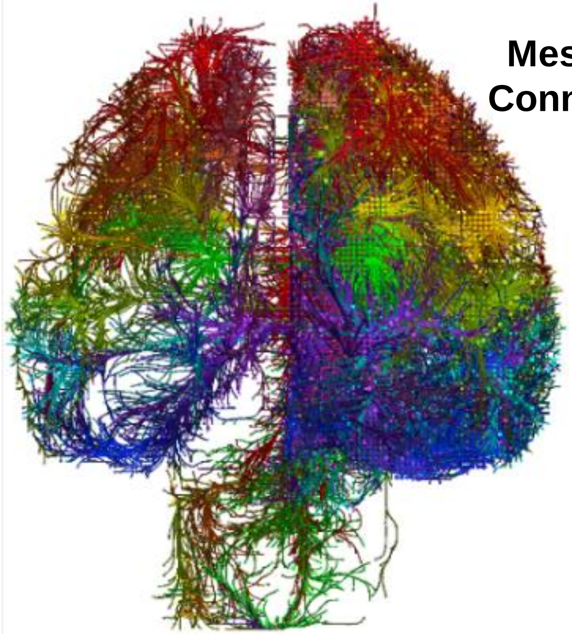
Electron Microscopy Connectomics







Multi-patch synaptic physiology



Mesoscale Connectome

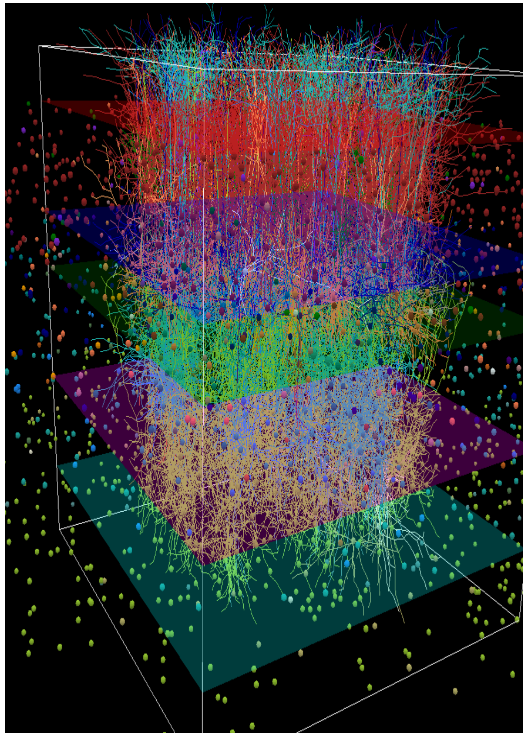
Connection probabilities

Target (columns) / *Source* (rows)

	i1HTR	E2/3	i2/3 PV	i2/3 SST	i2/3 HTR	E4	i4 PV	i4 SST	i4 HTR	E5	i5 PV	i5 SST	i5 HTR	E6	i6 PV	i6 SST	i6 HTR
i1HTR	0.656	0.356	0.093	0.068	0.464	0.148				0.148	0	0		0.148			
E2/3	0	0.16	0.395	0.182	0.105	0.016	0.083	0.083	0.083	0.083	0.081	0.102		0			
i2/3PV	0.024	0.411	0.451	0.03	0.22	0.05	0.05	0.05	0.05	0.07	0.073						
i2/3SST	0.279	0.424	0.857	0.082	0.77	0.05	0.05	0.05	0.05	0.021		0					
i2/3HTR	0	0.087	0.02	0.625	0.028	0.05	0.05	0.05	0.05	0							
E4		0.14	0.1	0.1	0.1	0.243	0.43	0.571	0.571	0.104	0.101	0.128	0.05	0.032			
i4PV		0.25	0.05	0.05	0.05	0.437	0.451	0.03	0.22	0.088	0.091	0.03	0.03				
i4SST	0.241	0.25	0.05	0.05	0.05	0.351	0.857	0.082	0.77	0.026	0.03	0	0.03				
i4HTR		0.25	0.05	0.05	0.05	0.351	0.02	0.625	0.028	0	0.03	0.03	0.03				
E5	0.017	0.021	0.05	0.05	0.05	0.007	0.05	0.05	0.05	0.116	0.083	0.063	0.105	0.047	0.03	0.03	0.03
i5PV	0	0	0.102			0	0.034	0.03	0.03	0.455	0.361	0.03	0.22	0.03	0.01	0.01	0.01
i5SST	0.203	0.169		0.017		0.056	0.03	0.006	0.03	0.317	0.857	0.04	0.77	0.03	0.01	0.01	0.01
i5HTR						0.03	0.03	0.03	0.03	0.125	0.02	0.625	0.02	0.03	0.01	0.01	0.01
E6		0				0				0.012	0.01	0.01	0.01	0.026	0.145	0.1	0.1
i6PV		0.1				0.1				0.1	0.03	0.03	0.03	0.1	0.08	0.1	0.08
i6SST										0.03	0.03	0.03	0.03	0.1	0.05	0.05	0.05
i6HTR										0.03	0.03	0.03	0.03	0.1	0.05	0.05	0.03

A detailed model of Mouse V1

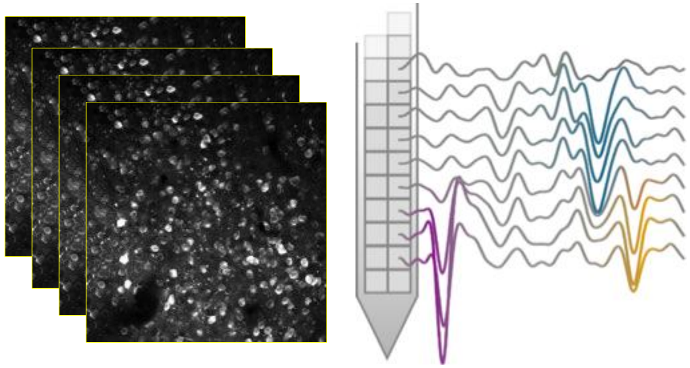
- 21 cell classes
- Cortical layers 1 to 6
- ~230,000 neurons

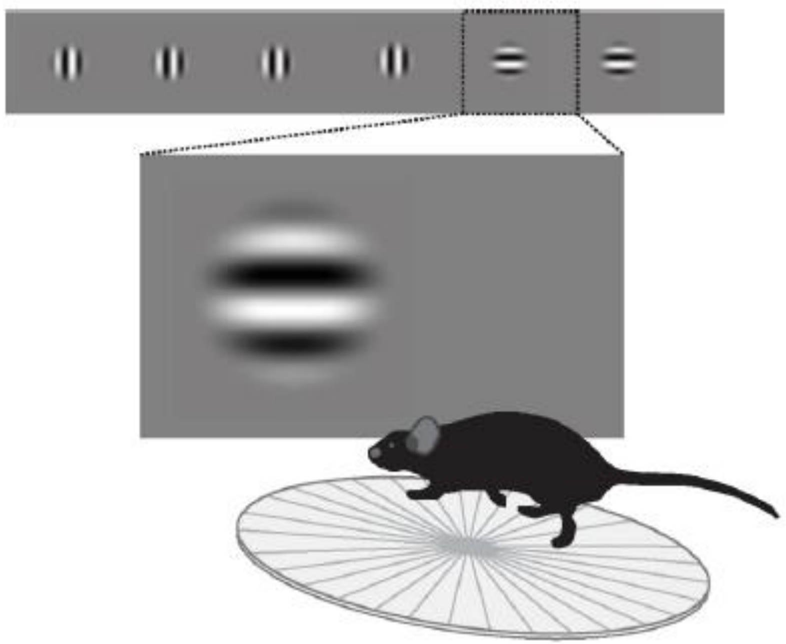


http://connectivity.brain-map.org/

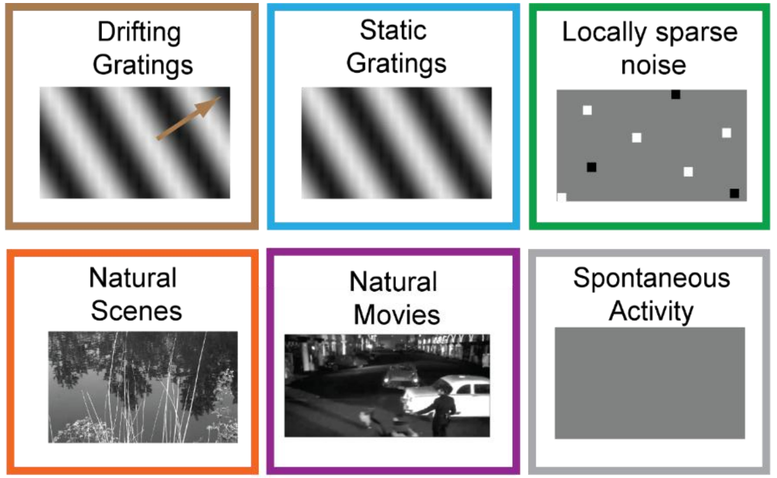
In Vivo Ophys and Ephys Brain Observatory

Ophys and Ephys measurements in awake mice

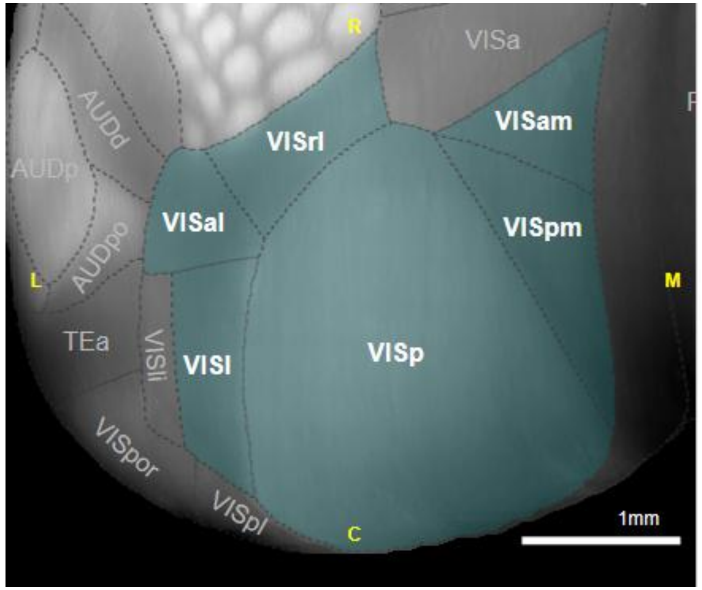


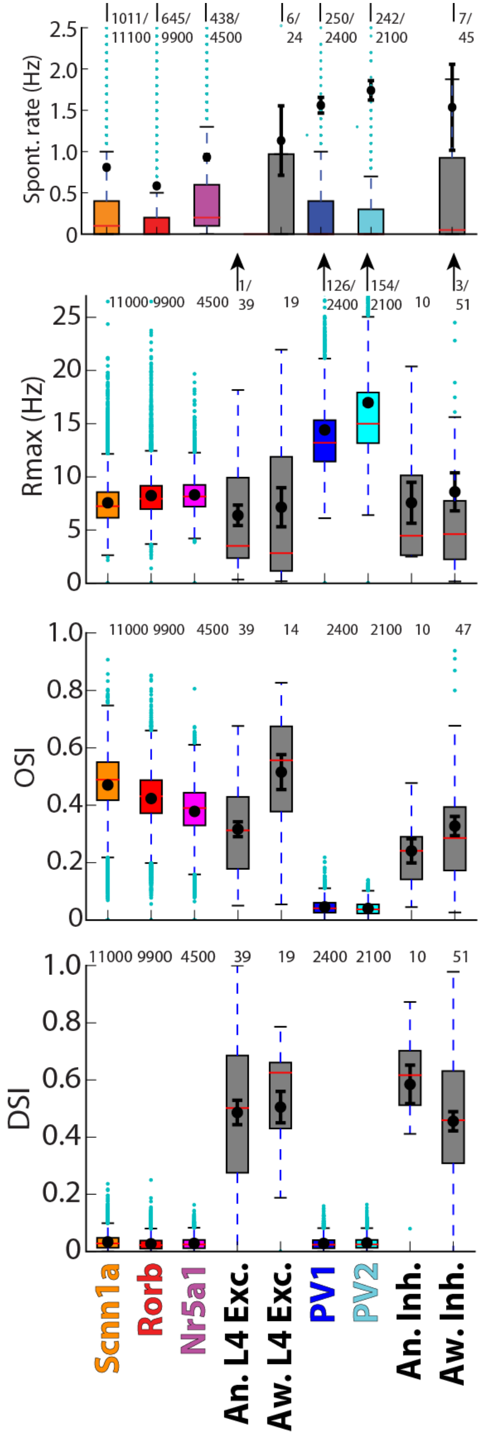


Diverse visual stimuli



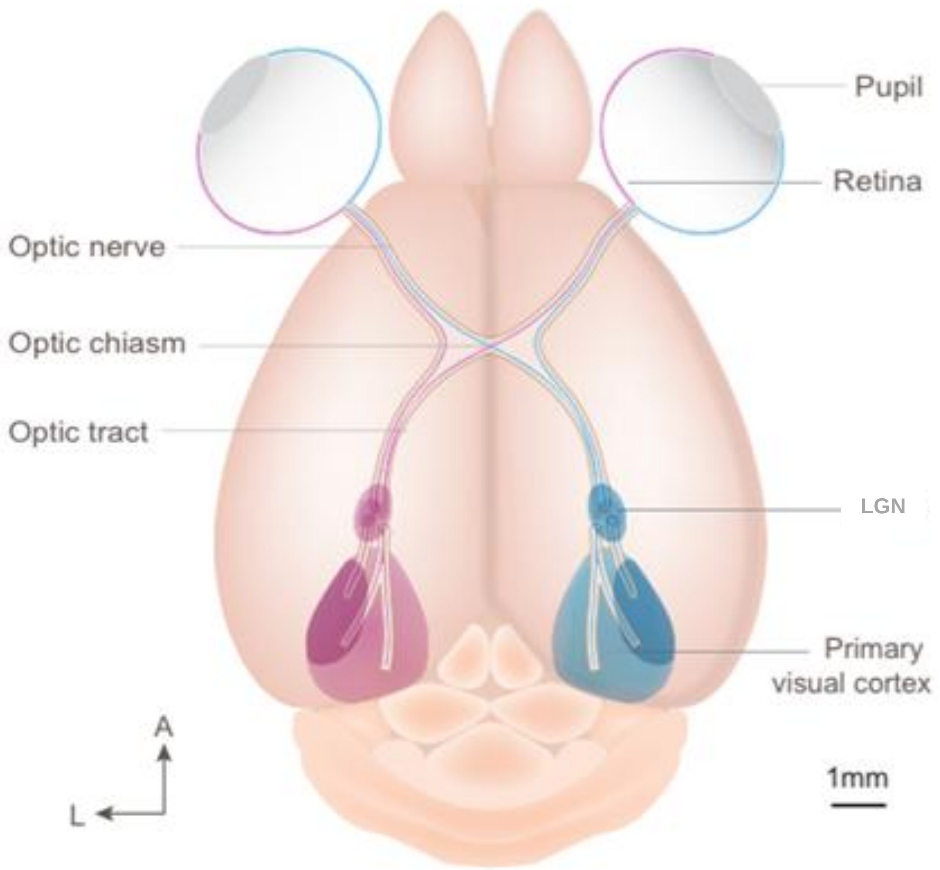
Various brain regions and cell populations

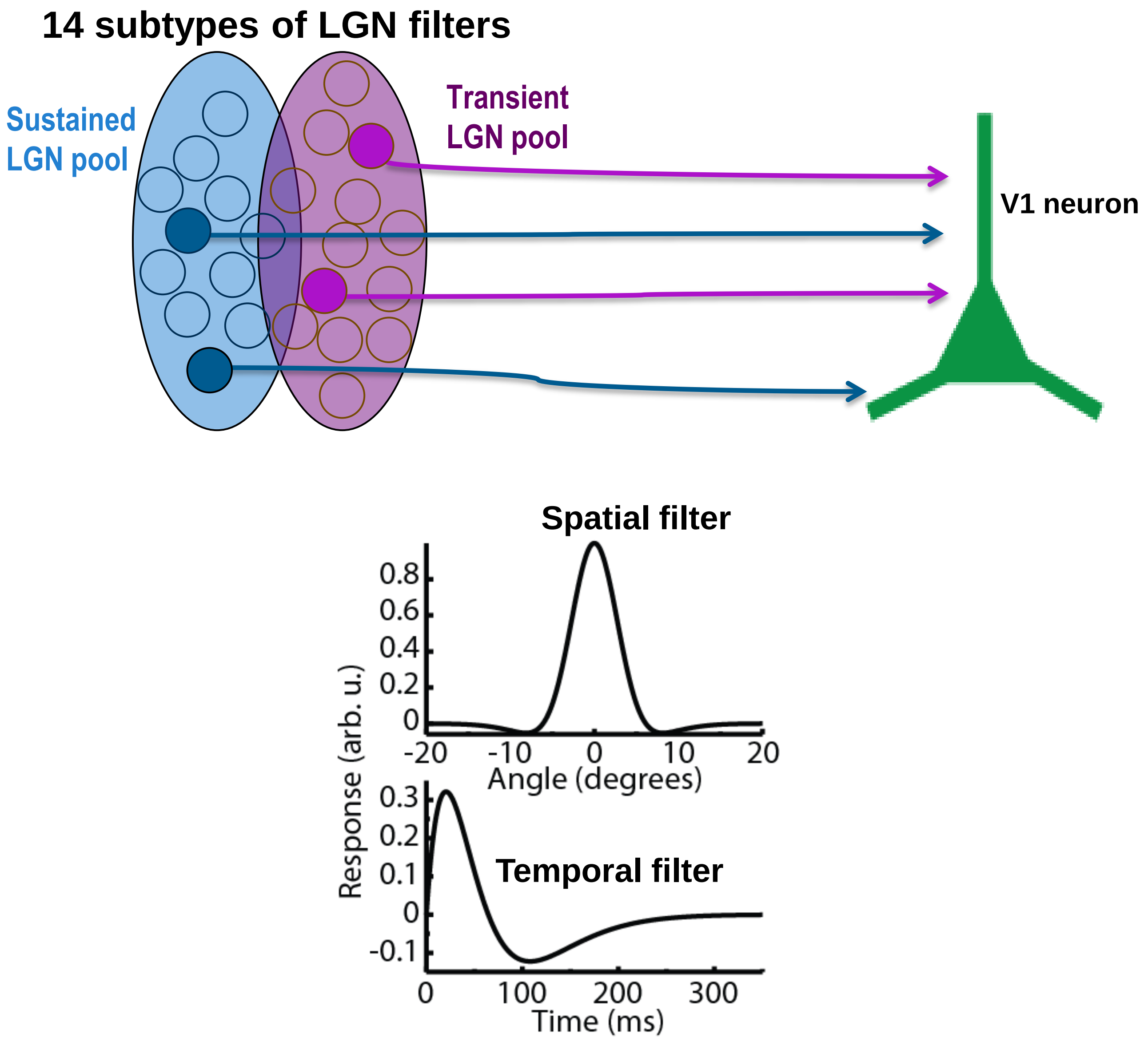




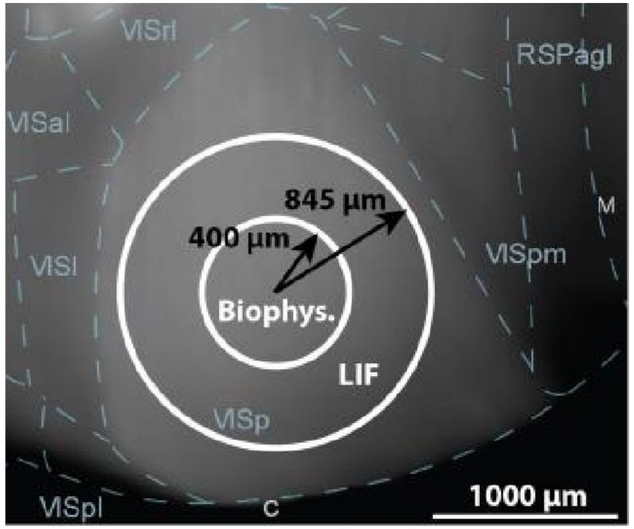
http://observatory.brain-map.org/visualcoding

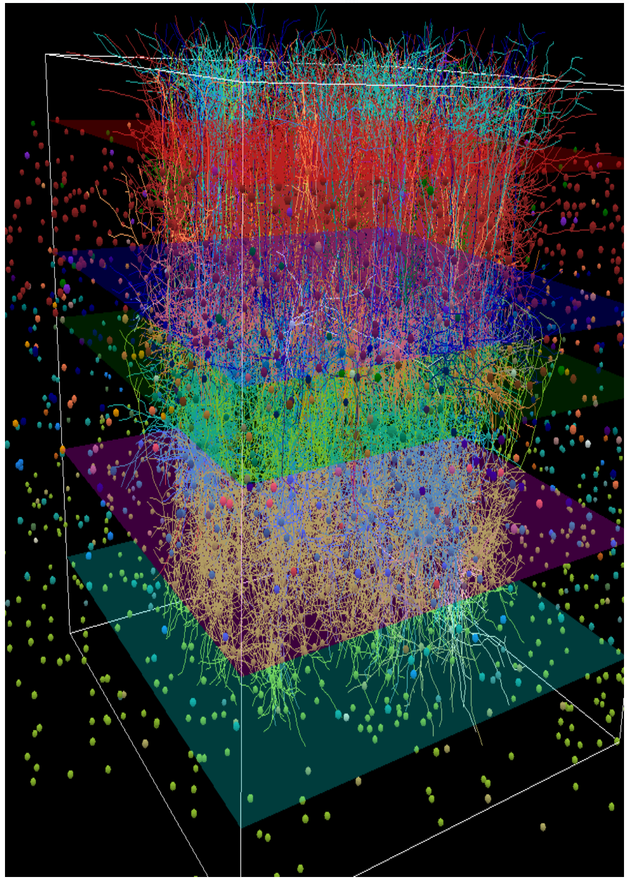
Modeling Visual Inputs into Cortical Network

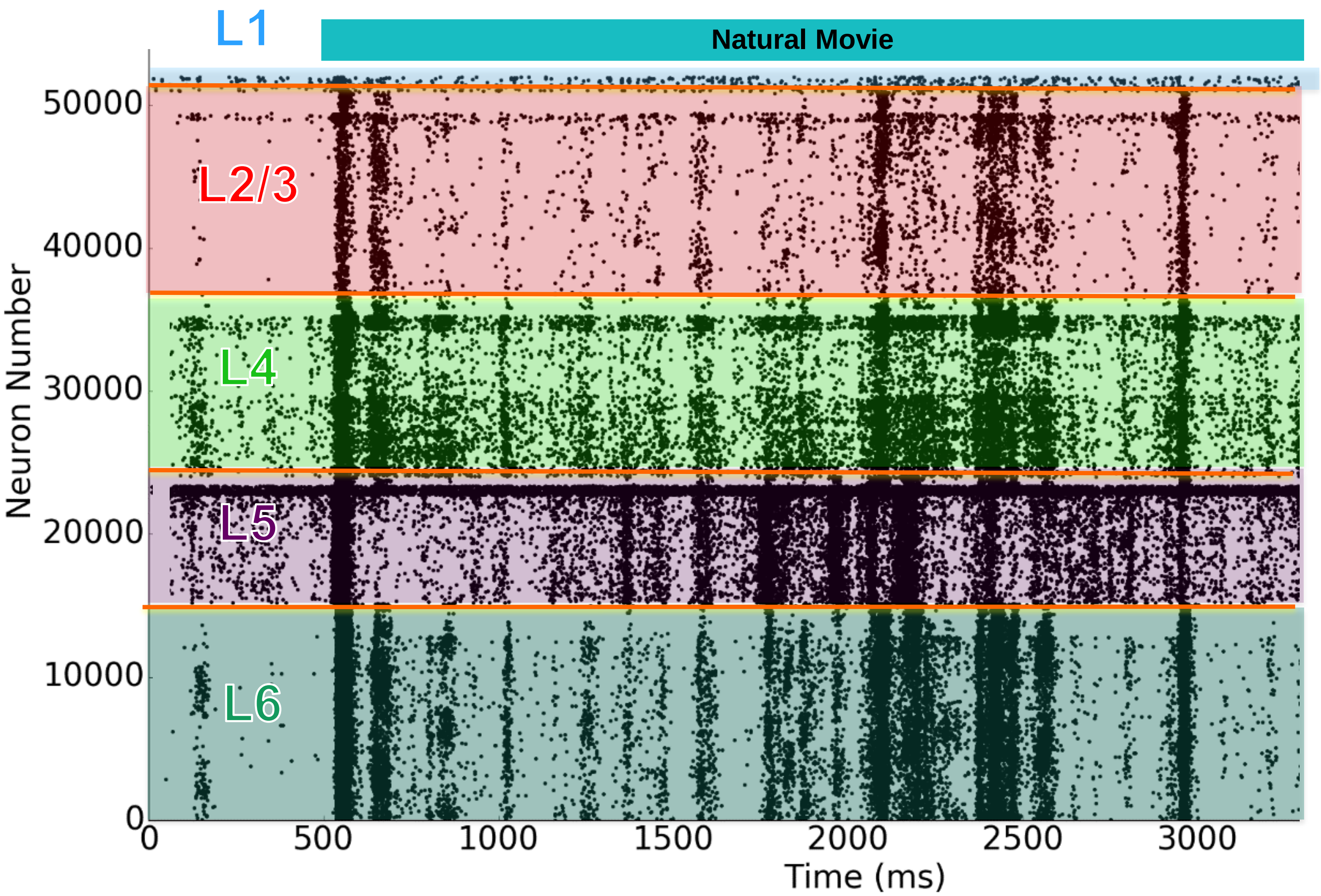




Output from Large-Scale Network Simulations

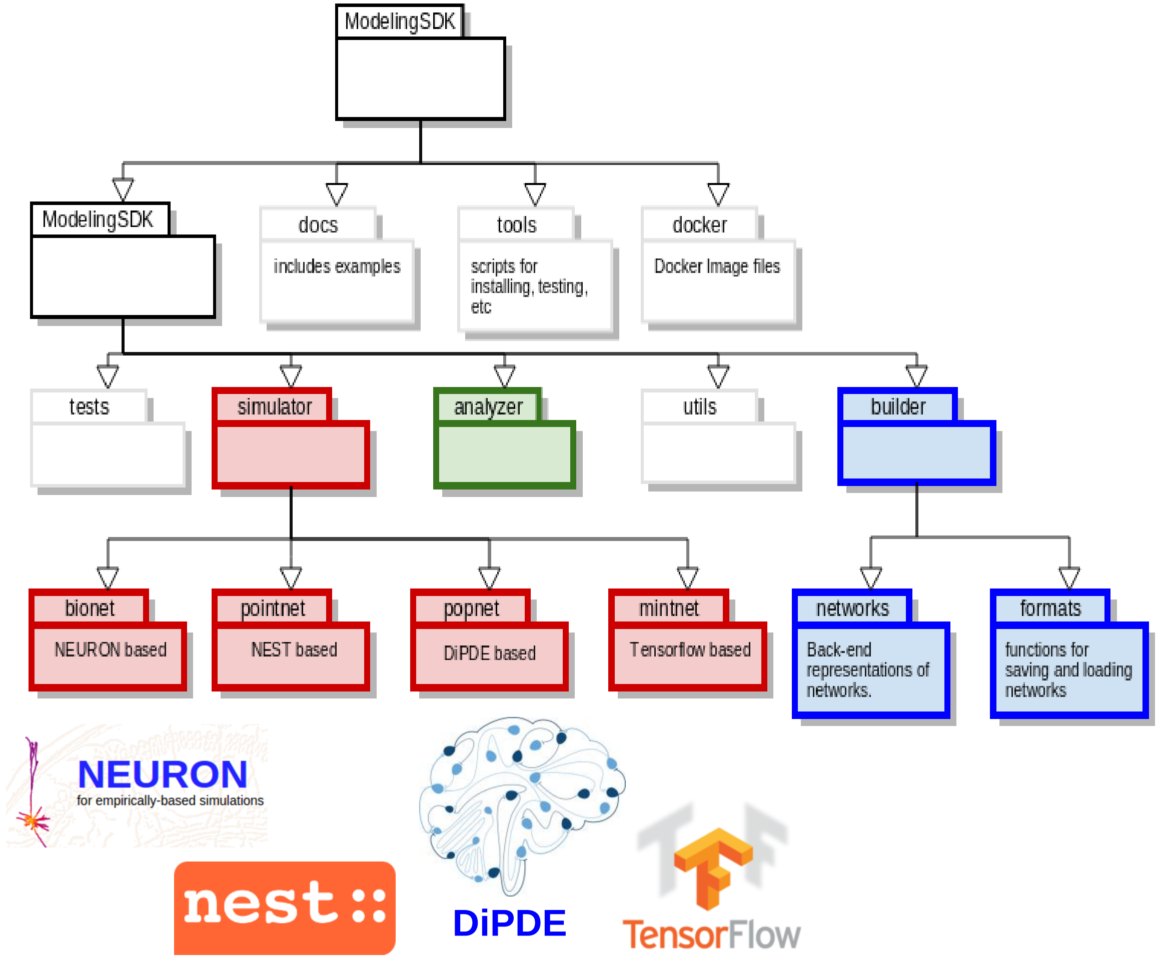






Brain Modeling ToolKit (BMTK)

BMTK is a software suite for building, simulating and analyzing models at different levels of resolution (biophysical, point neuron, filter, population statistics, machine intelligence).



The backbone of this ecosystem is a file format for model standardization, storage, and exchange.

- Cells
- Cables
- Morphologies
- Mechanisms and parameters
- Synapses
- Connection matrices
- Input spike trains
- Simulation parameters
- Simulation output

…

SONATA: github.com/allenInstitute/sonata

https://alleninstitute.github.io/bmtk/

BBP Projects and Use Cases

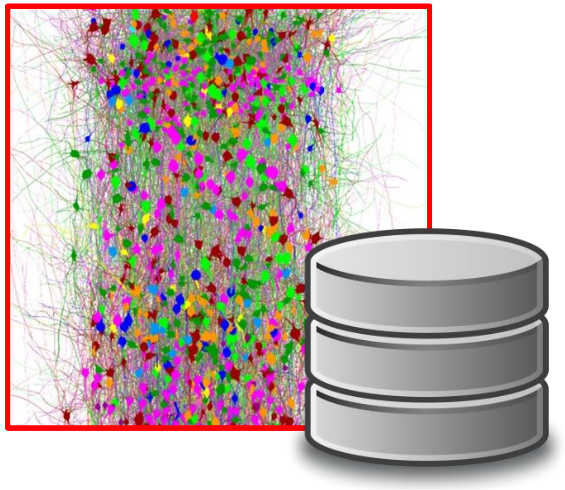
Data-driven, Integrative, Predictive, Collaborative

Blue Brain Project

Anatomy
Physiology
Cellular
Synaptic
Microcircuit

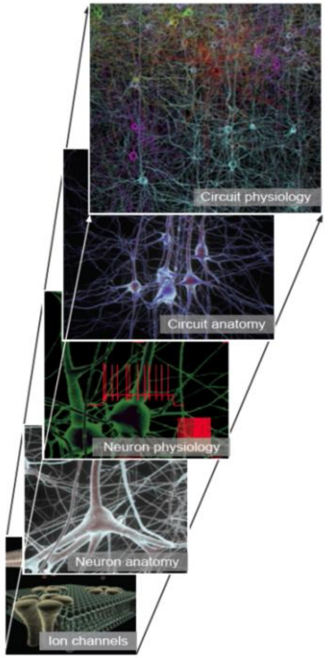
Sparse Data → **Inductive Principles, Algorithms** → **Structured Workflow** → **Dense Model**

Inverse Problem



Blue Brain Nexus

“Massively Collaborative Modeling”

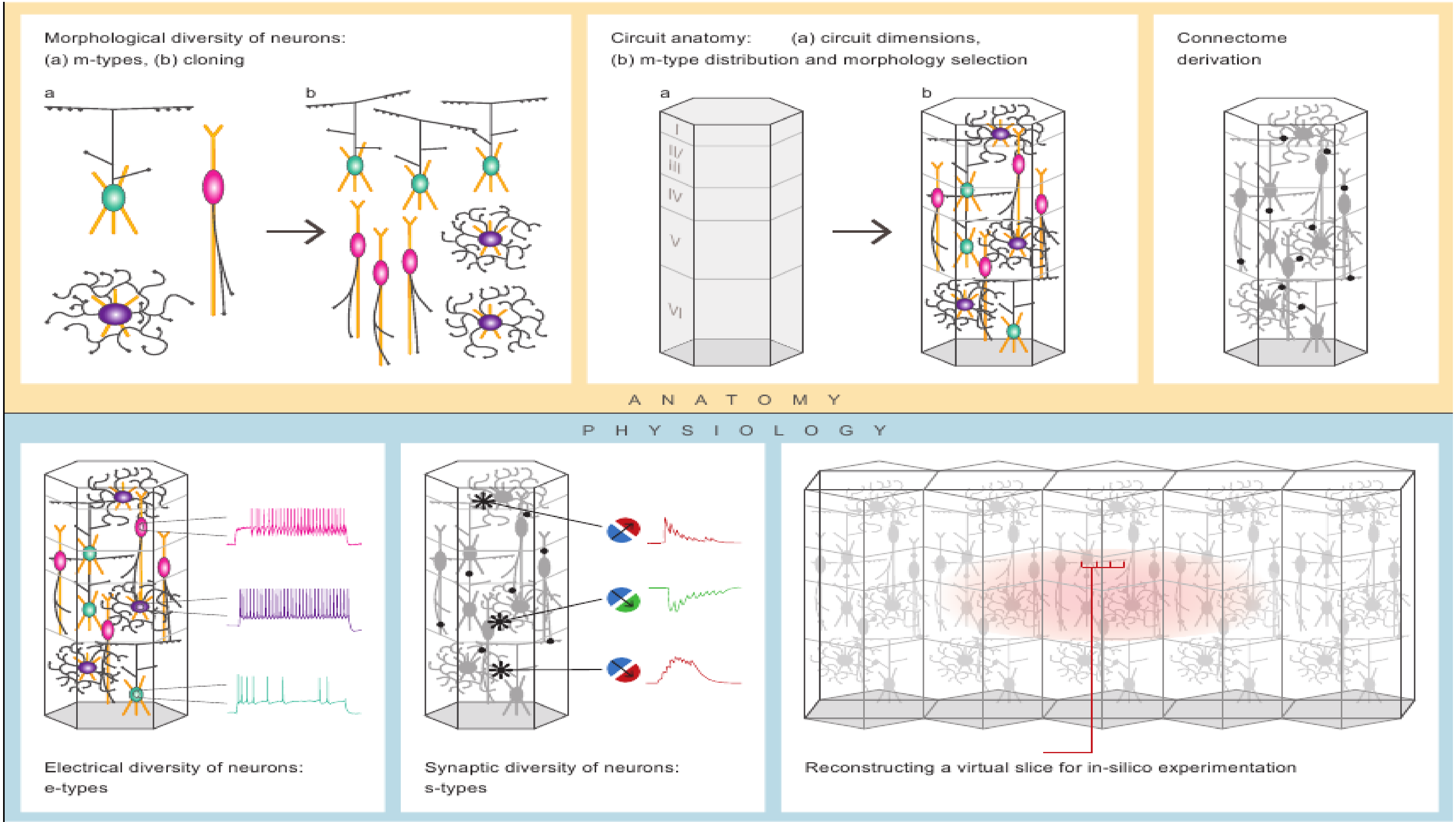


Multi-scale integration by domain experts

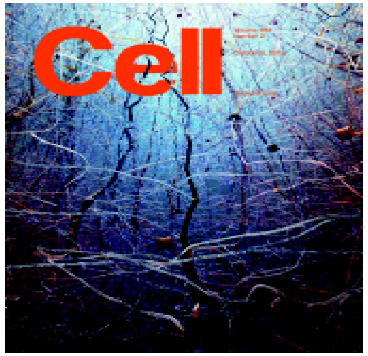
An analogy …

Photosynth 3D - Blaise Aguera y Arcas, Microsoft Corp.

Reconstruction Workflow

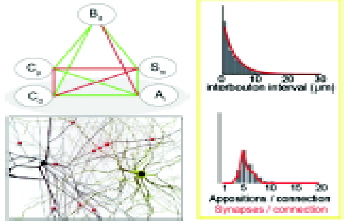


Resources



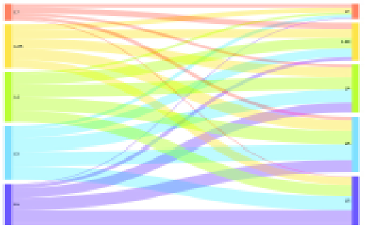
Reconstruction and simulation of neocortical microcircuitry

Markram H*, Muller E*, Ramaswamy S*, Reimann MW*, …
DeFelipe J, Hill SL, Segev I, Schuermann F
Cell 163:2, p456-492, 8 October 2015
* - indicates co-first author



An algorithm to predict the connectome of neural microcircuits

Reimann et al.
Front. Comp. Neurosci., 8 October 2015



The neocortical microcircuit collaboration portal: a resource for rat somatosensory cortex

Ramaswamy S, Courcol J-D, et al.
Frontiers in Neural Circuits, 8 October 2015

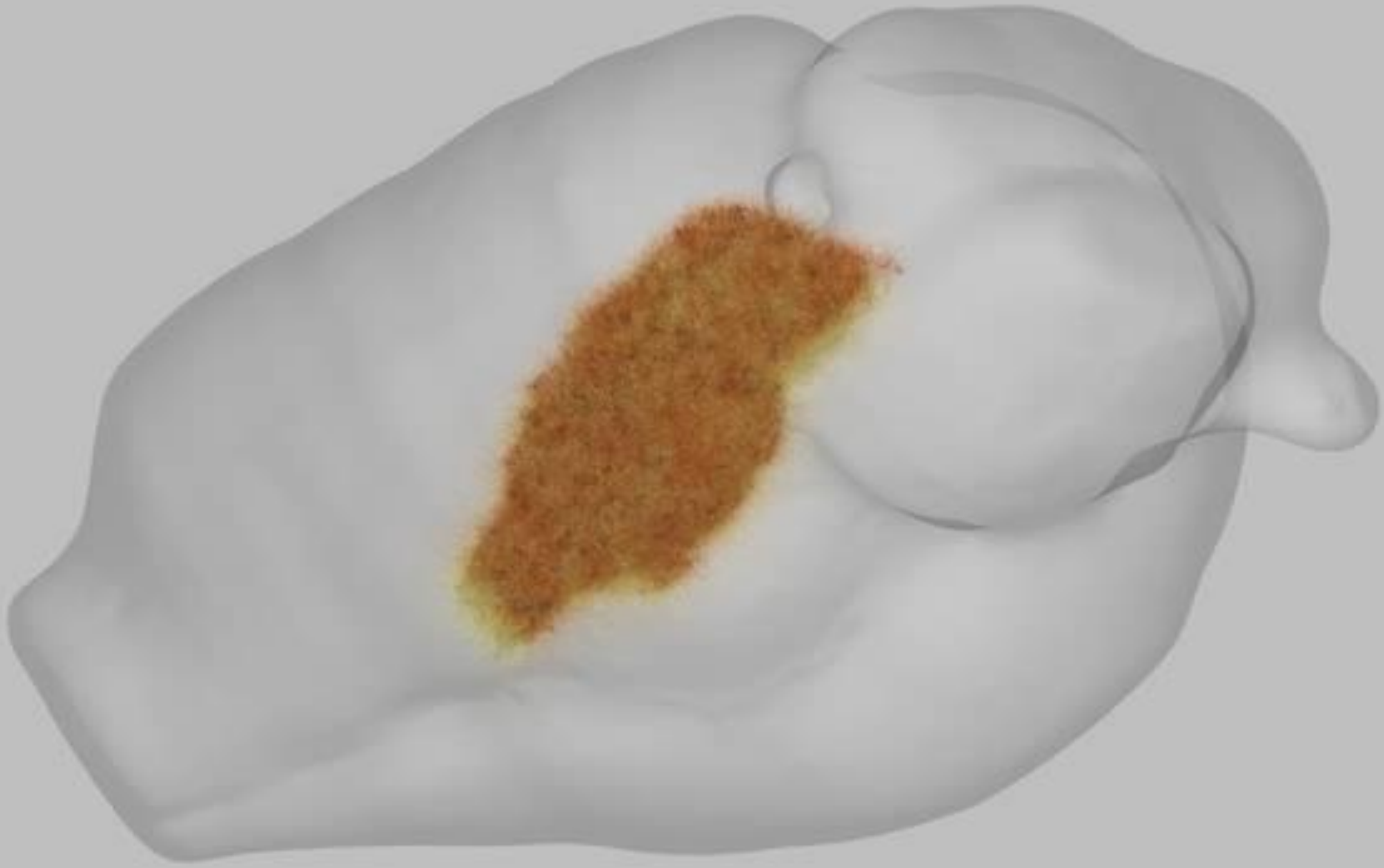
https://github.com/BlueBrain

https://bbp.epfl.ch/nmc-portal

Single Cell Tools
- NeuroM
- eFEL
- BluePyOpt
- BluePyMM

Viz Tools
- Brion (SONATA supp.)
- RTNeuron (SONATA supp.)
- Brayns
- Fivox

Blue Brain Nexus



HBP Brain Simulation Platform - Value Proposition: Collaborative Integrative Data-Driven Neuroscience

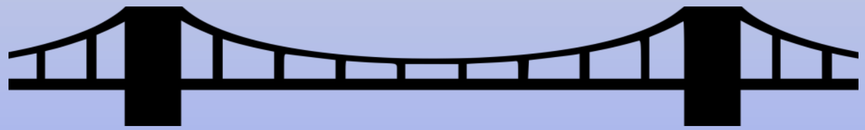
Data ⟷ **Neuro-informatics** ⟷ **Data-driven Models** ⟷ **Validation** ⟷ **Simpli-fication** ⟷ **Neuromorphic Neurorobotics Cognition**

Experiment

Strategic Mouse Brain Data

Strategic Human Brain Data

Neuroinformatics Platform



Brain Simulation Platform

Collaborative Modeling
Community “reference” models
Automation
Reproducibility
Transparency

Theory

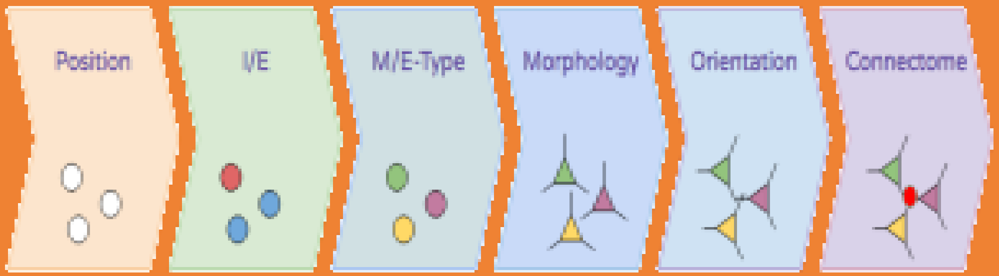
Neuromorphic Computing Platform

Neurorobotics Platform

Cognitive Architectures

High Performance Computing Platform

HBP Brain Simulation Platform – Architecture Overview



Workflows

Entities

Community

SP7 UNICORE

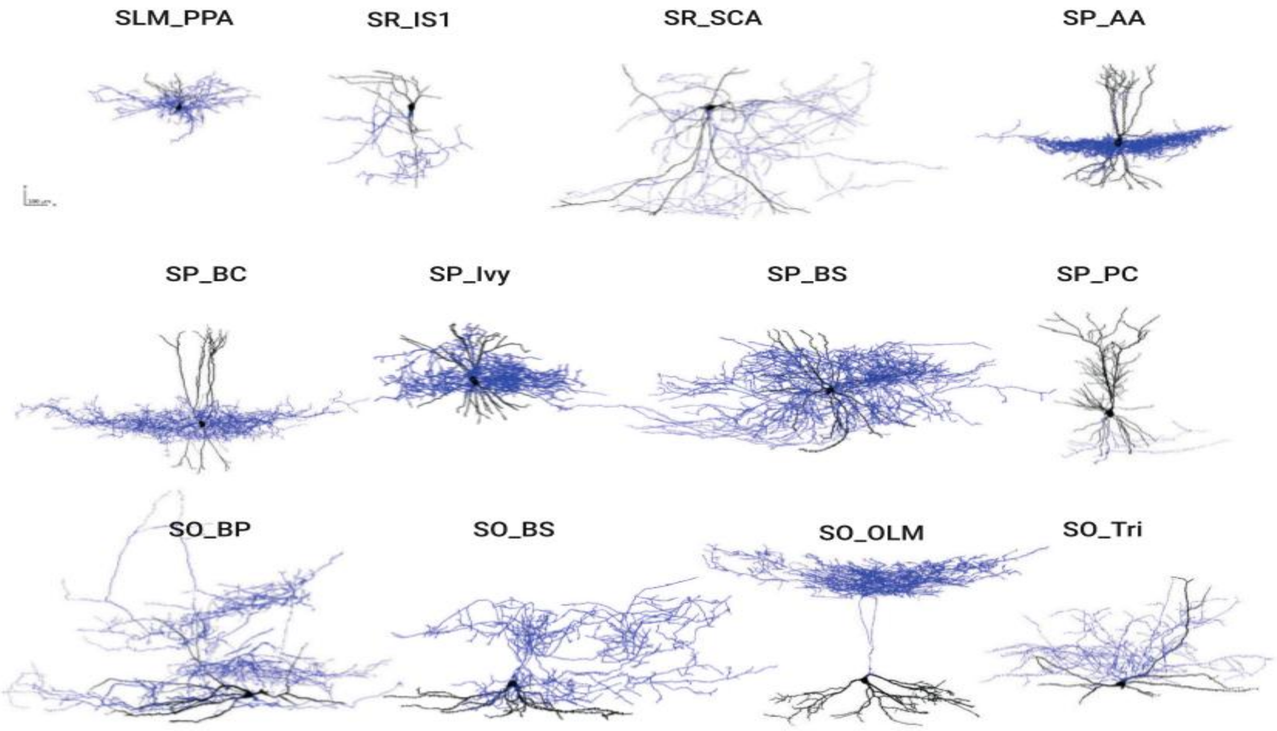
SP5 Blue Brain Nexus

Brain Simulation Platform foundation

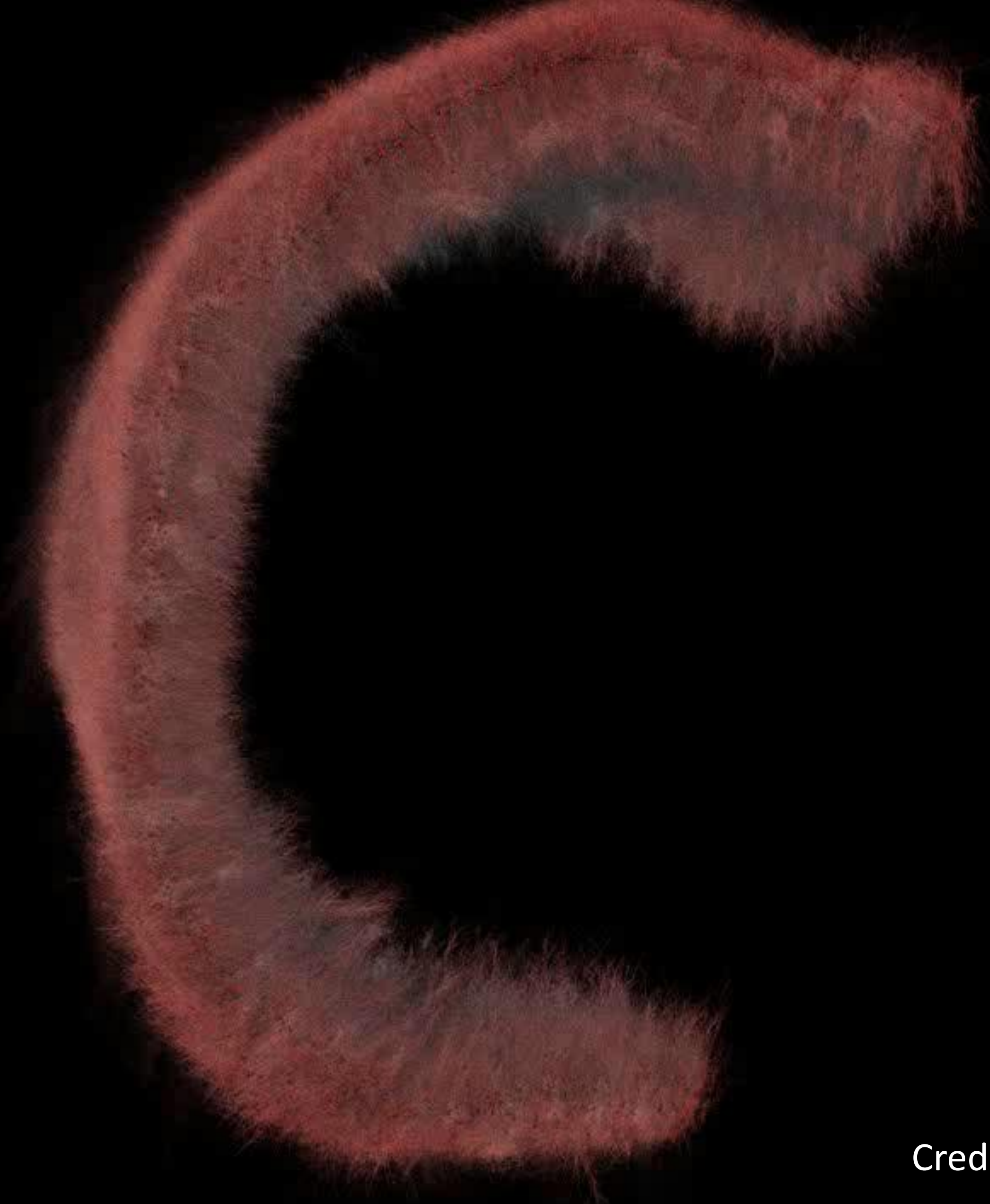
Reconstruction and Simulation of a full-scale model of Rat Hippocampus CA1

Romani A[1], Antille N[1], Budd J[3], Courcol JD[1], Devresse A[1], Ecker A[1], Falck J[2,6], Favreau C[1], Gevaert M[1], Gulyas AI[3], Hagens O[7], Hernando J[1], Jimenez S[1], Kali S[3], Kanari L[1], King JG[1], Lange S[2,5], Mercer A[2], Migliore M[4], Pezzoli M[7], Ramaswamy S[1], Reimann M[1], Rössert CA[1], Sáray S[3,8], Shi Y[1], Thomson AM[2], Van Geit WAH[1], Vanherpe L[1], Markram H[1], Muller EB[1]

[1] Blue Brain Project (BBP), Brain Mind Institute, EPFL, Lausanne, Switzerland; [2] University College London, United Kingdom; [3] Institute of Experimental Medicine, Hungarian Academy of Sciences, Hungary; [4] Institute of Biophysics, National Research Council, Italy;
[5] University of Westminster, London, United Kingdom; [6] Deutsches Zentrum für Neurodegenerative Erkrankungen, Germany; [7] Laboratory of Neural Microcircuitry (LNMC), Brain Mind Institute, EPFL, Lausanne, Switzerland;
[8] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary



JÜLICH FORSCHUNGSZENTRUM · UCL · Blue Brain Project · EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE · Human Brain Project · European Commission



Credit: Cyrille Favreau, Armando Romani

Blue Brain Project

HBP Hippocamp 2017: Collaborative and Integrative Modeling of Hippocampus

23rd-24th May 2017, Paris, France

eITN

European Institute for Theoretical Neuroscience

http://neuralensemble.org/meetings/Hippocamp2017/

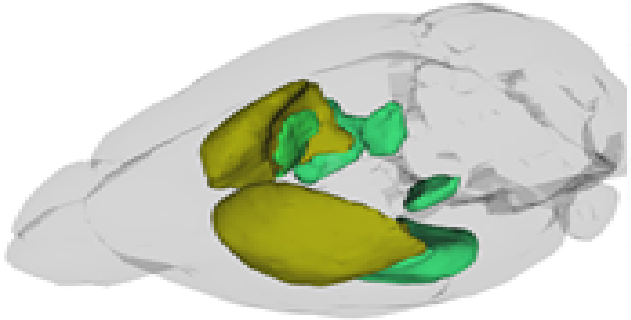
HBP Brain Simulation Platform Co-Design Drivers

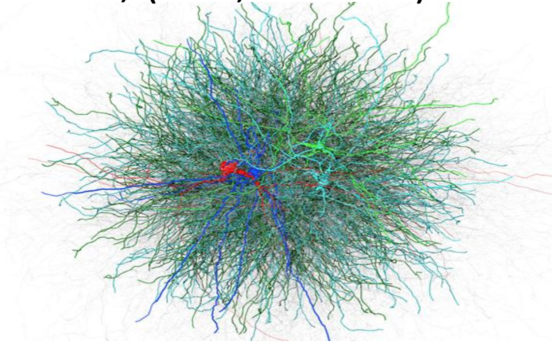
Hippocampus

Lead: Szabolcs Kali (IEM-HAS, Budapest)

Basal Ganglia

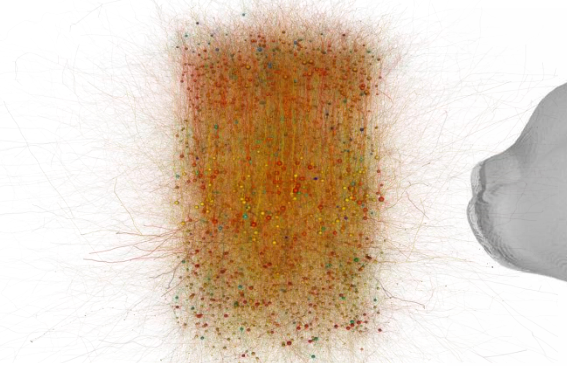
Lead: S. Grillner, J. Hellgren-Kotaleski, (KTH, Sweden)

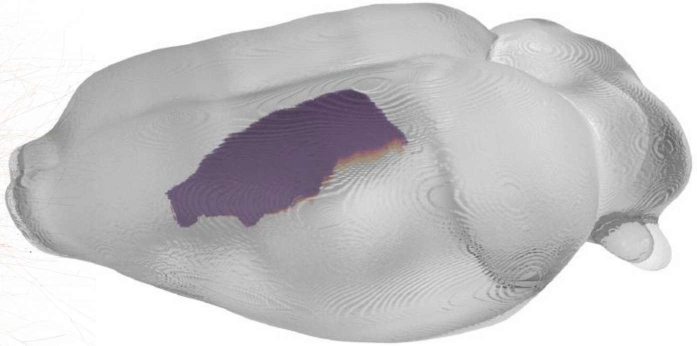




Somatosensory Cortex

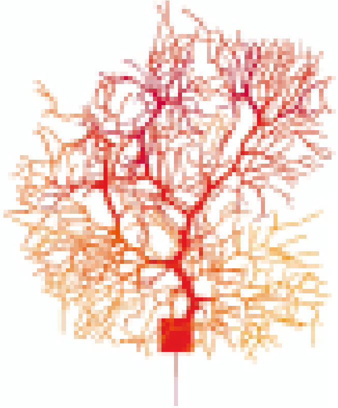
Lead: E. Muller (EPFL, Switzerland)





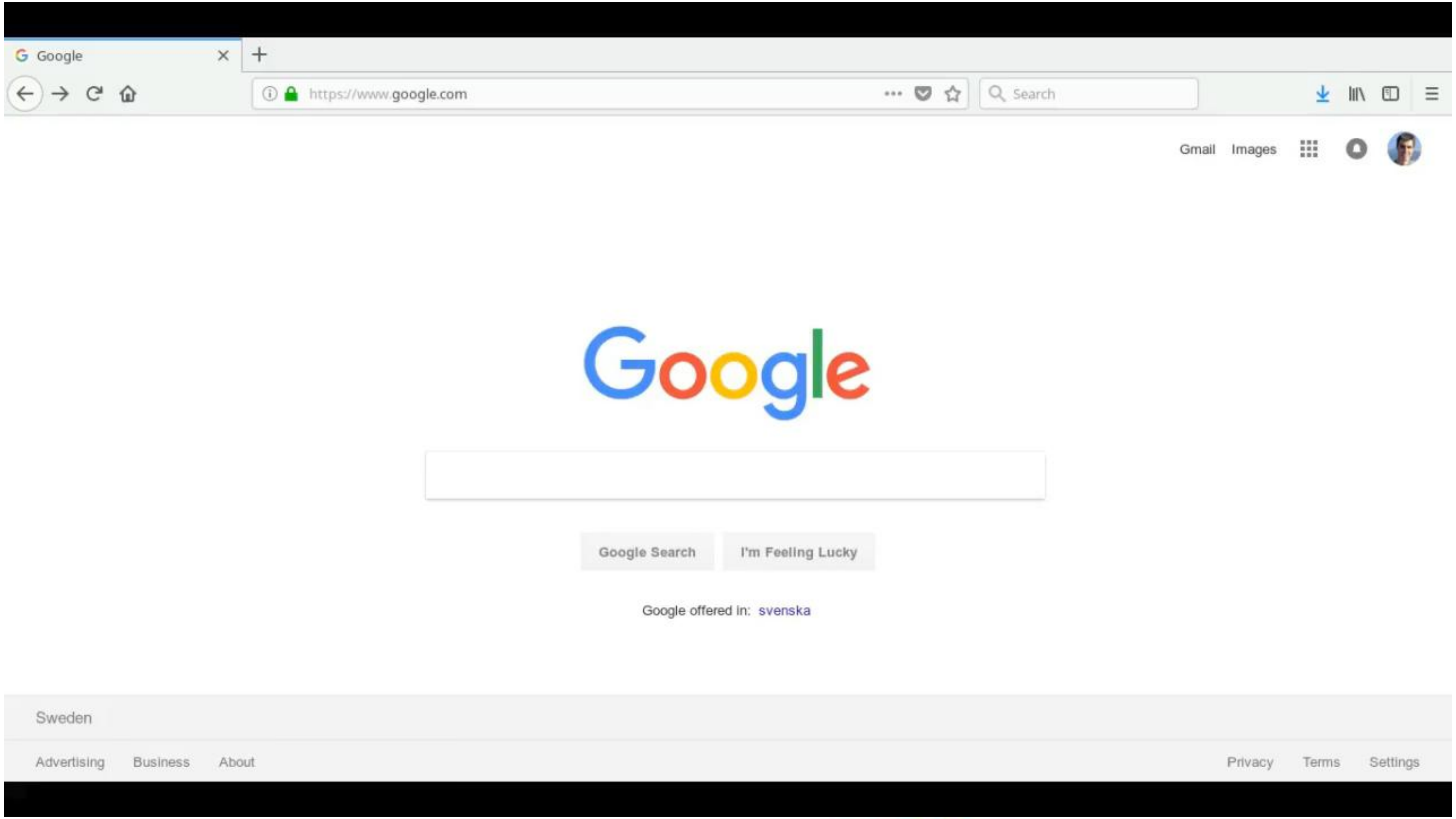
Cerebellum

Lead: E. D'Angelo, U. Pavia, Italy



Simulation Neuroscience MOOC #1 - Single Neurons

Blue Brain Project



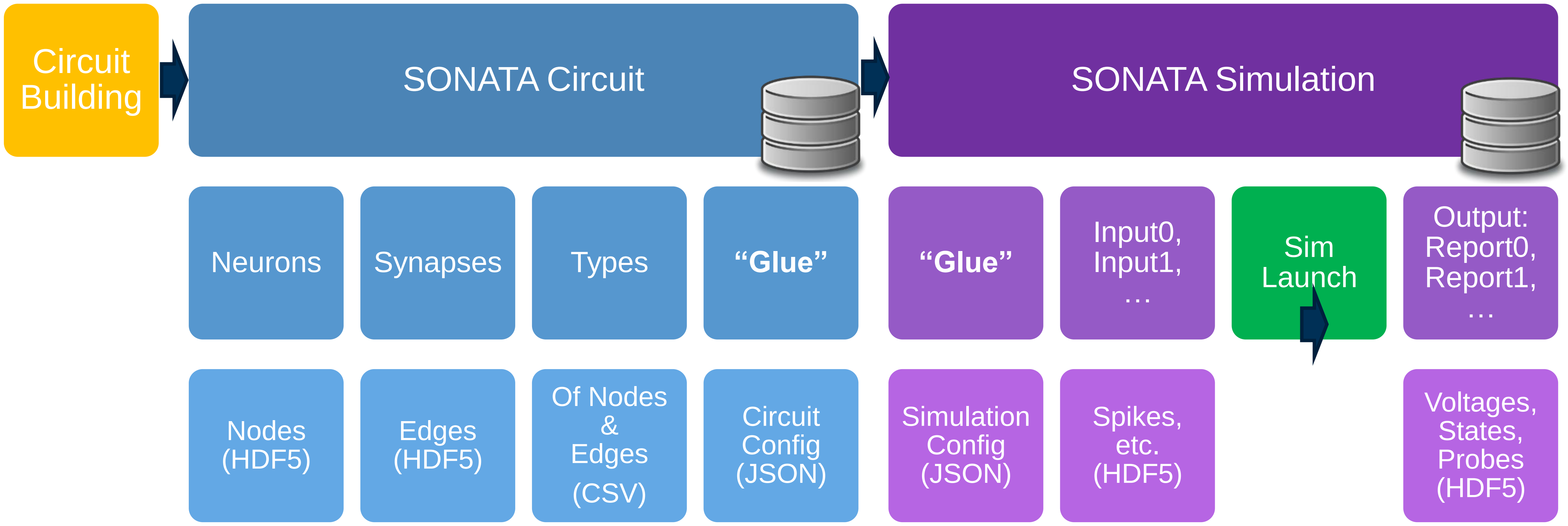
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

SONATA Data Format

Overview - SONATA Architecture

Overarching design goals:

- Convergence of AIBS and BBP file format approaches
- Enable large-scale, high performance: sim, viz, analysis

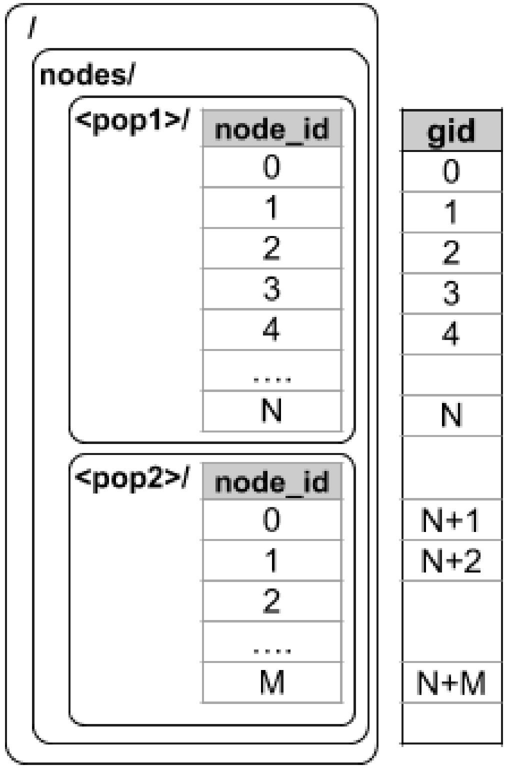


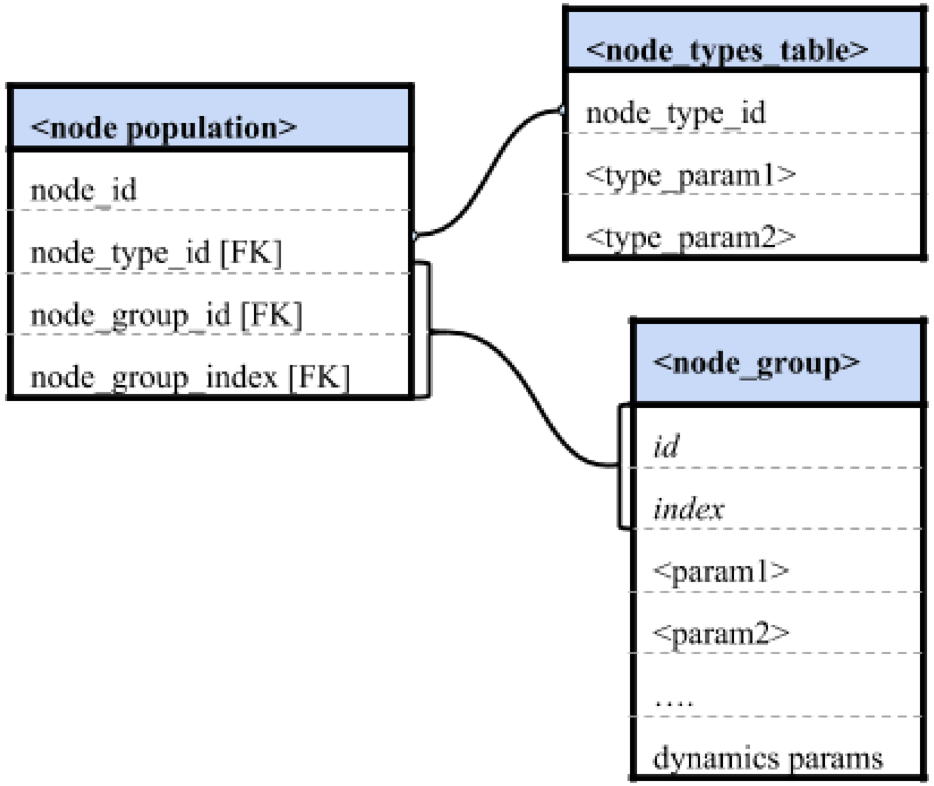
SONATA Circuit

Representation of Nodes

Highlights

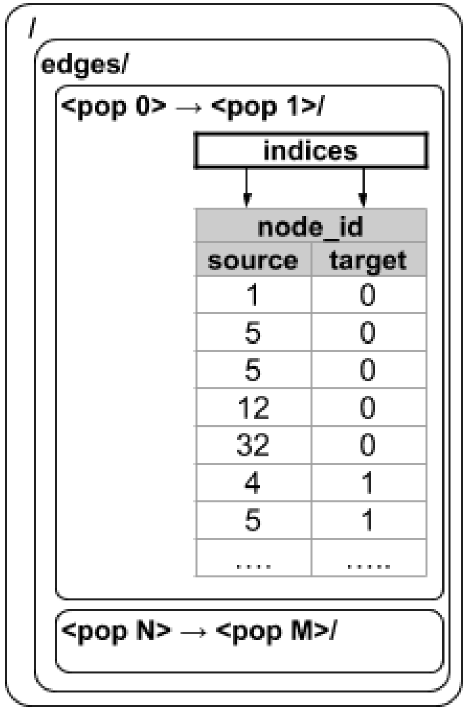
- Global features can be stored in “node type” table (CSV), and node-specific features in “nodes” table (HDF5).
- Users can add their own fields (in addition to reserved ones, such as “model_type”, “morphology”, “x”, “y”, “z”, …).
- Support for NeuroML description of cell models

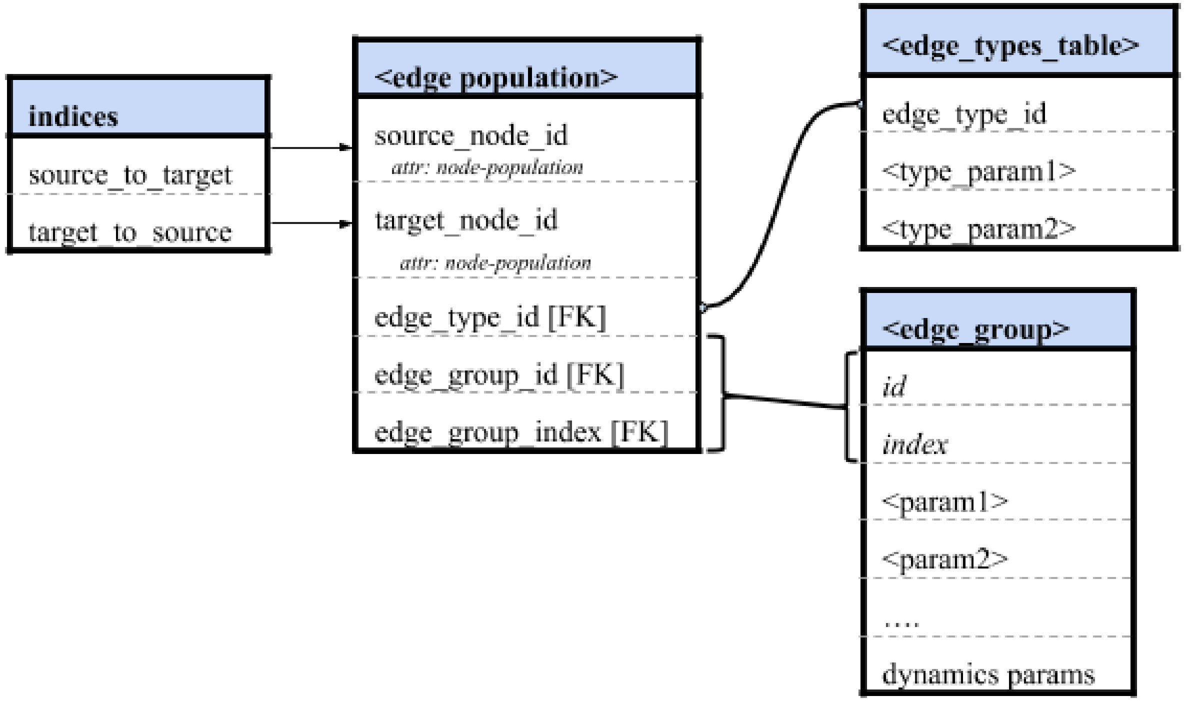




github.com/allenInstitute/sonata

Representation of Edges



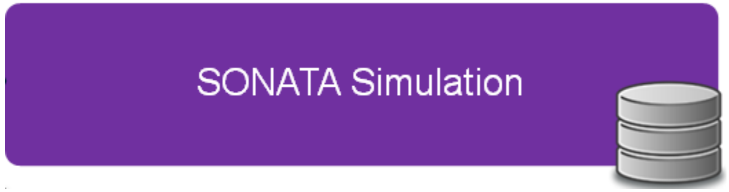


Highlights

- Like for nodes, there are global features (“edge type” CSV table) and specific features (“edges” HDF5 table).
- Users can add their own fields in addition to reserved ones (“syn_weight”, “delay”, “sec_id”, …) and provide exact specification of synaptic placement or rule-based specification.
- Convenient indexing, e.g. for efferent and afferent views
- Synapse models specified by strings, which map e.g. to NEURON MOD files.

github.com/allenInstitute/sonata

JSON-based Simulation Configuration Files



- Circuit config file – defines relative locations of circuit components (e.g., where the morphologies are stored).

- Simulation config file – defines simulation conditions and inputs for the circuit (e.g., simulation duration, directories where inputs spike data are stored, directories for writing output, …).

sonata_sim_config.json

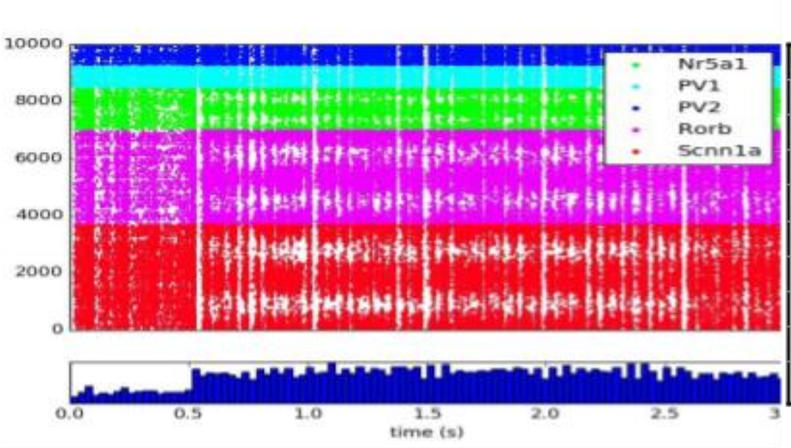

```
{
  "run": {
    "tstop": 3000.0,
    "dt": 0.1,
    "dL": 20.0,
    "spike_threshold": -15,
    "nsteps_block": 5000
  },
  "conditions": {
    "celsius": 34.0,
    "v_init": -80
  },

  "output":{"log_file": "$OUTPUT_DIR/log.txt"...},

  "inputs": {
    "spikes": {
      "input_type": "spikes",
      "module": "h5",
      "input_file": "inputs/exc_spike_trains.h5",
      "node_set": "biophysical"
    },

    "current_clamp": {
      "input_type": "current_clamp",
      "module": "IClamp",
      "node_set": "biophysical",
      "amp": 0.2200,
      "delay": 500.0,
      "duration": 1000.0
    }
```


github.com/allenInstitute/sonata

Simulation Output

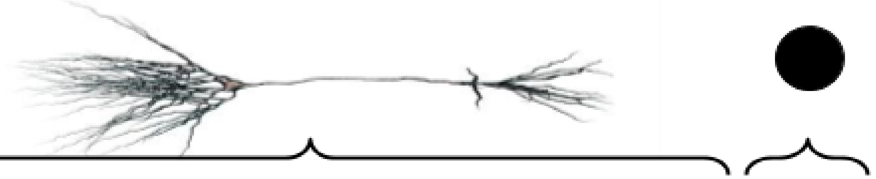
SONATA uses HDF5 to store and share the results of different simulation variables including:

Spike Trains



gid	spikes (ms)
0	1.02
0	11.10
0	22.98
0	28.01
0	….
1	9.12
1	15.22
1	19.98

Single- and multi-compartment variables

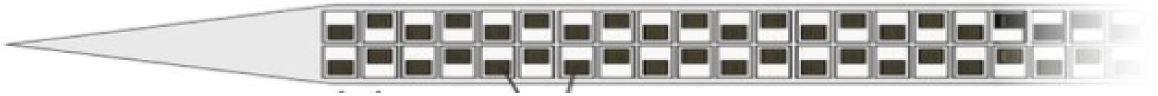


Time

	gid 0						gid 1	gid 2
0	-80.123	-80.123	-80.022	-80.001	-79.124	….	0.000	
1	-80.023	-80.123	-79.322	-80.000	-81.294	….	0.012	
2	-80.011	-80.110	-79.211	-79.052	-79.615	….	0.019	
3	-79.950	-80.070	-78.999	-78.799	-77.118	….	0.019	
4	-78.550	-80.001	-77.999	-77.999	-77.118	….	0.021	
5	-78.450	-79.070	-77.819	-77.559	-77.118	….	0.021	
6	-75.150	-80.070	-78.999	-79.991	-77.118	….	0.022	

voltage traces per cell/compartments

Extracellular time series



Time

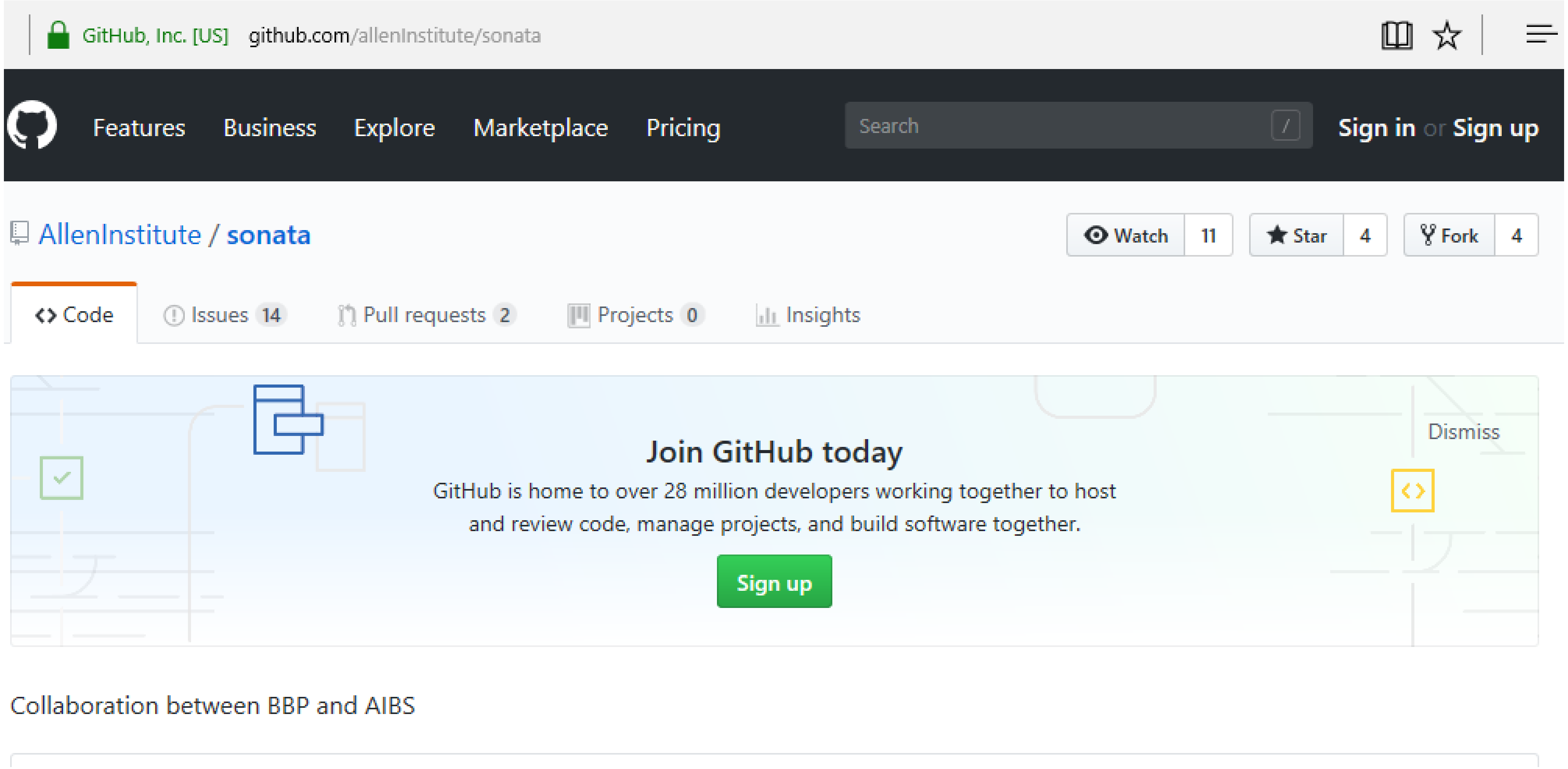
	0	1	2	4	5	6
0	-1.001	-3.140	-3.177	-3.511	-3.212	….
1	-1.022	-4.709	-4.976	-5.267	-4.677	
2	-2.550	-5.485	-5.796	-6.135	-7.125	
3	-3.001	-5.855	-6.188	-6.551	-7.252	
4	-5.548	-6.011	-6.353	-6.726	-7.300	
5	-5.730	-6.047	-6.399	-6.766	-9.598	
6	….					

extracellular voltage per channel

github.com/AllenInstitute/sonata

Get Involved!

github.com/AllenInstitute/sonata



Acknowledgments

Kael Dai

Jean-Denis Courcol

Yazan Billeh

Juan Hernando

Adrien Devresse

Sergey Gratiy

Michael Gevaert

James King

Werner Van Geit

Daniel Nachbauer

Arseny Povolotskiy