AFRL-AFOSR-VA-TR-2020-0064

Principles of Robust Learning Derived from the Structure and Function of the Cortical Column

Armen Stepanyants
NORTHEASTERN UNIVERSITY

12/29/2019
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

DISTRIBUTION A: Distribution approved for public release.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 17-06-2020	**2. REPORT TYPE** Final Performance	**3. DATES COVERED** *(From - To)* 30 Sep 2015 to 29 Sep 2018

4. TITLE AND SUBTITLE Principles of Robust Learning Derived from the Structure and Function of the Cortical Column	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER FA9550-15-1-0398
	5c. PROGRAM ELEMENT NUMBER 61102F
6. AUTHOR(S) Armen Stepanyants	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NORTHEASTERN UNIVERSITY 360 HUNTINGTON AVE BOSTON, MA 02115 US	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AF Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/AFOSR RTA2
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-VA-TR-2020-0064

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The goal of this project is to uncover the effects of learning and long-term memory storage on synaptic connectivity, thus, creating the basis for quantitative analyses of these fundamental brain functions. To that end, we developed a model of a biologically-inspired recurrent neural network, capable of storing and reliably retrieving temporal sequences of network states. The model incorporates many basic elements of local connectivity in the mammalian neocortex, including excitatory and inhibitory neuron classes, neuron morphologies, the homeostatic constraint on connection weights, and several types of errors and noise in signal transmission. The model was solved analytically and numerically with the replica and convex optimization methods. In addition, a perceptron-type learning rule was developed to load associative memory sequences into the network in a biologically-plausible online manner. Our results revealed that when individual neurons are robustly loaded with a near-maximum amount of memories they can support, the network develops many structural and dynamical properties that are consistent with experimental observations.

15. SUBJECT TERMS
Computational Cognition, Machine Learning

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			LAWTON, JAMES
Unclassified	Unclassified	Unclassified	UU		

					19b. TELEPHONE NUMBER *(Include area code)* 703-696-5999

Final Report
Principles of Robust Learning Derived from the Structure and Function of the Cortical Column

Grant Number:	FA9550-15-1-0398
Principal Investigator:	Armen Stepanyants
Period of Performance:	September 30, 2015 – September 29, 2019
Program:	Computational Cognition and Machine Intelligence
Program Manager:	Dr. James Lawton

Project objectives

The goal of this project is to uncover the effects of robust associative learning and long-term memory storage on synaptic connectivity, thus, creating the basis for quantitative analyses of these fundamental brain functions. The following objectives were proposed. (1) Develop a biologically-realistic model of robust associative learning by cortical circuits. The model must be derived from the hypothesis that synaptic connectivity in a given circuit of the adult cortex is functioning in a steady-state in which the associative memory storage capacity of the circuit is maximal and learning of new associations is accompanied by forgetting some of the old ones. The model must integrate the current knowledge of excitatory and inhibitory neuron classes, with structural connectivity constraints imposed by the morphologies of axonal and dendritic arbors of cortical neurons, with homeostatic constraints on numbers and strengths of synaptic connections. (2) Simulate steady-state learning based on one of the best-studied networks in the mammalian neocortex – the barrel-centered column of rodent somatosensory cortex. The simulations must be embedded in the structural connectivity of the column built from the morphologies of neurons reconstructed in three-dimensions from various cortical depths. (3) Validate the structural and dynamical properties of the steady-state circuits with a large dataset of experimental studies reporting probabilities of connections between neurons, probabilities of specific higher-order connectivity motifs, distributions of unitary postsynaptic potentials, and strengths of laminar and inter-laminar projections in rodent barrel cortex. The dataset, created as part of the project, must encompass connectivity of major excitatory and inhibitory neuron classes from different cortical layers.

Summary of major accomplishments and results

1. The associative memory storage model (Objective 1)

We developed a model of associative learning and memory storage based on the knowledge of connectivity in the cerebral cortex. The model includes multiple classes of inhibitory and excitatory neurons and incorporates various types of homeostatic and sign-constraints on connection weights. The model neural networks can learn temporal sequences and entire basins of network states, which is believed to be the foundation of all cognitive functions of the brain. One of the main advantages of the model is that the memory storage capacity of the network and the properties of its connectivity can be determined analytically. This makes it possible to explore the effects of model parameters and constraints on the network functions.

We used a McCulloch and Pitts neural network to model a local cortical circuit in which N_{inh} inhibitory neurons and $(N - N_{inh})$ excitatory neurons are all-to-all potentially connected (Figure 1A). Associative memory in the model is a connected graph of successive network states (directed edges termed associations), $\left\{X^{\mu} \to X'^{\mu}\right\}$, in which every node has no more than one daughter node (Figure 1B). Vectors X^{μ} and X'^{μ} contain binary activities (0 or 1) of individual neurons within an association μ. In general, an associative memory can be in the form of a point attractor, an associative sequence, a limit cycle, and an entire basin of attraction.

Learning in the network is mediated by changes in connection weights of individual neurons, J_{ij}, in the presence of several biologically inspired constraints. (1) The input connection weights of each neuron are sign-constrained to be non-negative if the presynaptic neuron is excitatory and non-positive if it is inhibitory (Dale's principle). (2) The input weights of each neuron are homeostatically constrained to have a predefined l_1-norm. (3) Each neuron attempts to learn its associations robustly so that they can be recalled correctly in the presence of postsynaptic noise.

Each neuron in the network (e.g. neuron i), independently from other neurons, attempts to learn a set of input-output associations $\left\{X^{\mu} \to X_i'^{\mu}\right\}$, in which a vector X^{μ} represents the neuron's input for an association μ, and a scalar $X_i'^{\mu}$ is the desired output of the neuron derived from the subsequent network state X'^{μ} (orange boxes in Figure 1B). To simplify the notation, in the following expressions we drop index i, replace $X_i'^{\mu}$ with y^{μ}, and summarize the learning problem like so:

$$
\begin{aligned}
&\theta\left(\sum_{j=1}^{N} J_j X_j^{\mu} - h + \eta\right) = y^{\mu}; \quad \mu = 1,\ldots,m \\
&J_j g_j \geq 0; \quad j = 1,\ldots,N \\
&\frac{1}{N}\sum_{j=1}^{N}\left|J_j\right| = w \\
&\left|\eta\right| \leq \kappa \\
&\text{Prob}\left(X_j^{\mu}\right) = \begin{cases} 1-f, & X_j^{\mu} = 0 \\ f, & X_j^{\mu} = 1 \end{cases}; \quad \text{Prob}\left(y^{\mu}\right) = \begin{cases} 1-f, & y^{\mu} = 0 \\ f, & y^{\mu} = 1 \end{cases}
\end{aligned} \tag{1}
$$

In these expressions, θ denotes the Heaviside step-function, h is the neuron's firing threshold, and η denotes its postsynaptic noise, which is bounded by the robustness parameter κ, i.e. $\left|\eta\right| < \kappa$. To enforce sign-constraints on the neuron's presynaptic connection weights, we introduced parameters $\{g_j\}$ and set g_j to 1 if the connection j is excitatory and –1 if it is inhibitory. Parameter w, referred to as the average absolute connection weight, is introduced to impose the l_1-norm constraint on the weights of presynaptic connections. Binary input and output states, X_j^{μ} and y^{μ}, are randomly drawn from the Bernoulli probability distribution: 0 with probability $1-f$ and 1 with probability f.

The above network model is governed by the following parameters: number of neurons in the network, N, fraction of inhibitory neurons, N_{inh}/N, threshold of firing, h, firing probability of neurons in the associative states, f, average absolute connection weight, w, robustness parameter κ, and memory load, $\alpha = m/N$.

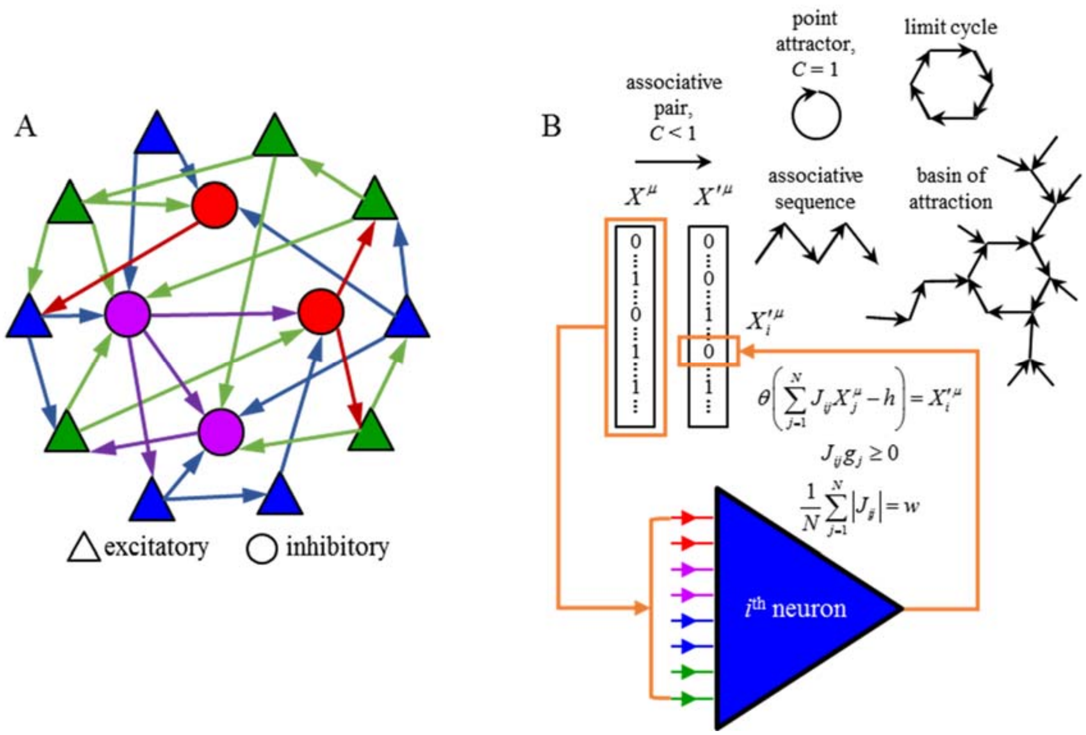

Figure 1: Associative memory storage in recurrent networks of excitatory and inhibitory neurons. **A.** A recurrent network of various classes (color) of all-to-all potentially connected excitatory and inhibitory neurons. Note that the arrows indicate actual (or functional) connections. **B.** Associative memory in the model is a connected graph of successive network states (directed edges termed associations), $\{X^{\mu} \rightarrow X'^{\mu}\}$, in which every node has no more than one daughter node. Each neuron in the network (e.g. neuron *i*) must learn a set of input-output associations derived from the memory (orange boxes) by modifying the strengths of its input connections, J_{ij}, under the constraints on connection signs and l_1-norm.

2. Theoretical solution of the model (Objective 1)

We used the replica theory from statistical physics to solve the above model analytically, in its most general formulation. This solution is described in detail in Zhang, et al., J Neuroscience, (2019). It yields the neuron's critical (maximum) capacity, α, the probabilities of excitatory and inhibitory connections, $P^{con}_{exc/inh}$, and the probability densities of non-zero excitatory and inhibitory connection weights, $p^{PSP}_{exc/inh}$:

$$
\begin{aligned}
&\alpha\left(\tilde{w},\frac{N_{inh}}{N},f,\rho\right)=\frac{2\rho^2}{\sigma^2\left(u_+ + u_-\right)^2}\frac{fD\left(u_-\right)+(1-f)D\left(u_+\right)}{\left(fE\left(u_-\right)+(1-f)E\left(u_+\right)\right)^2}\\
&P_{inh}^{con}\left(\tilde{w},\frac{N_{inh}}{N},f,\rho\right)=E\left(v_+\right)\\
&P_{exc}^{con}\left(\tilde{w},\frac{N_{inh}}{N},f,\rho\right)=E\left(v_-\right)\\
&p_{inh}^{PSP}\left(\tilde{J}\mid\tilde{w},\frac{N_{inh}}{N},f,\rho\right)=\frac{\theta\left(-\tilde{J}\right)}{\sqrt{2\pi}\sigma\tilde{w}E\left(v_+\right)}e^{-\left(\frac{\tilde{J}}{\sqrt{2}\sigma\tilde{w}}+v_+\right)^2}\\
&p_{exc}^{PSP}\left(\tilde{J}\mid\tilde{w},\frac{N_{inh}}{N},f,\rho\right)=\frac{\theta\left(\tilde{J}\right)}{\sqrt{2\pi}\sigma\tilde{w}E\left(v_-\right)}e^{-\left(\frac{\tilde{J}}{\sqrt{2}\sigma\tilde{w}}-v_-\right)^2}
\end{aligned}
\tag{2}
$$

These quantities are expressed as functions of five latent variables, u_+, u_-, v_+, v_-, and σ, which can be obtained by solving the following system of equations and inequalities:

$$
\begin{cases}
fF\left(u_-\right)-(1-f)F\left(u_+\right)=0\\
\frac{N_{exc}}{N}F\left(v_-\right)+\frac{N_{inh}}{N}F\left(v_+\right)=\frac{\sqrt{2}}{\sigma}\\
\frac{N_{exc}}{N}F\left(v_-\right)-\frac{N_{inh}}{N}F\left(v_+\right)=\frac{\sqrt{2}}{\tilde{w}f\sigma}\\
\frac{N_{exc}}{N}D\left(v_-\right)+\frac{N_{inh}}{N}D\left(v_+\right)=\frac{2\rho^2}{\sigma^2\left(u_+ + u_-\right)^2}\\
\sigma=\frac{\sqrt{2}\rho^2}{\left(u_+ + u_-\right)\left(\frac{1}{\tilde{w}f}\left(v_+ - v_-\right)-\left(v_+ + v_-\right)\right)}\frac{fF\left(u_-\right)+(1-f)F\left(u_+\right)}{fE\left(u_-\right)+(1-f)E\left(u_+\right)}\\
u_+ + u_- > 0;\quad \sigma>0
\end{cases}
\tag{3}
$$

$$
E\left(x\right)=\frac{1}{2}\left(1+\operatorname{erf}\left(x\right)\right);\quad F\left(x\right)=\frac{1}{\sqrt{\pi}}e^{-x^2}+x\left(1+\operatorname{erf}\left(x\right)\right);\quad D\left(x\right)=xF\left(x\right)+E\left(x\right)
$$

We note that the distributions of inhibitory and excitatory connection weights are composed of Gaussian functions (standard deviation σ) truncated at zero and finite fractions of zero-weight connections. MATLAB code for generating the replica theory solution of the associative memory storage model is available at https://github.com/neurogeometry/AssociativeLearning.

3. Numerical solution of the associative memory model with a method based on convex optimization (Objective 1)

While the above theoretical solution can describe a network in the $N \to \infty$ limit at its maximum memory storage capacity, numerical methods are required for the analyses of finite networks and for the account of the learning process. For these reasons, we developed two numerical algorithms: first, based on the primal-dual, convex optimization formulation of the problem, and second, based on a biologically more plausible, perceptron-type learning rule.

For the first algorithm, we made the problem feasible by introducing a slack variable $s^{\mu} \geq 0$ for every association and chose the solution that minimizes the sum of these variables by solving the following linear optimization problem:

$$
\begin{aligned}
&\underset{\{J_j\}}{\arg\min}\left(\sum_{\mu=1}^{m} s^{\mu}\right) \\
&\left(2y^{\mu}-1\right)\left(\sum_{j=1}^{N} J_j X_j^{\mu} - h\right) + s^{\mu} \geq \kappa, \quad \mu = 1,\ldots,m \\
&s^{\mu} \geq 0 \\
&\frac{1}{N}\sum_{j=1}^{N}\left|J_j\right| = w \\
&J_j g_j \geq 0, \quad j = 1,\ldots,N
\end{aligned}
\tag{4}
$$

The code is available at https://github.com/neurogeometry/AssociativeLearning.

4. Numerical solutions of the model with a perceptron-type learning rule (Objective 1)

For the second algorithm, the sum of slack variables was minimized online yielding a perceptron-type learning rule, which is biologically more plausible. Here, a single not yet learned association is chosen at random and the synaptic weights of the neuron are updated in four consecutive steps:

$$
\begin{aligned}
&J_j \mapsto J_j + \beta\left(2y^{\mu} - 1\right)X_j^{\mu}, \quad j = 1,\ldots,N \\
&J_j \mapsto J_j\theta\left(J_j g_j\right) \\
&J_j \mapsto J_j + \left(w - \frac{1}{N}\sum_{j=1}^{N}\left|J_j\right|\right)g_j \\
&J_j \mapsto J_j\theta\left(J_j g_j\right)
\end{aligned} \tag{5}
$$

Unlike the standard perceptron learning rule, Eqs. (5) enforce the sign and homeostatic l_1-norm constraints during learning. The first update in Eqs. (5) is a standard perceptron learning step, in which parameter β is referred to as the learning rate. The second step was introduced to enforce the sign constraints, while the last two steps combined implement the homeostatic l_1-norm constraint and are equivalent to the soft thresholding used in LASSO regression. MATLAB implementation of the perceptron-type learning rule is available at https://github.com/neurogeometry/AssociativeLearning.

5. Dataset of cortical connectivity (Objective 3)

We created a dataset of features of synaptic connectivity in local cortical circuits. The dataset describes a total of 856 projections based on 152 articles published in peer-reviewed journals since 1990. It includes detailed information about probabilities of connection and connection strengths in various animals and brain regions, for various laminar projections, classes of pre- and postsynaptic neurons, ages, experimental conditions, and recording techniques. The dataset is available at the Neurogeometry group website, http://www.neurogeometry.net/resources/datasets.

6. Experimental analysis of learning and circuit plasticity (Objective 3)

We collaborated with the laboratories of Anthony Holtmaat from the University of Geneva and Graham Knott from EPFL, to detect and interpret circuit changes that accompany learning and long-term memory formation *in vivo*. To that end, we developed *BoutonAnalizer* software which makes it possible to measure structural changes in connectivity during perceptual learning based on time-lapse, 2-photon microscopy imaging of axonal boutons. The software was validated on a dataset of boutons imaged and reconstructed at both light and electron microscopy levels. *BoutonAnalizer* was used to create a large dataset of connectivity changes in the barrel cortex of mice involved in long-term, perceptual learning tasks. More than 5,000 synaptic boutons were tracked in five animals over a period of two months. This unprecedented dataset was used to constrain and guide our theoretical efforts. One article resulting from this collaboration was published, Gala, et al., eLife (2017), and another is in preparation. *BoutonAnalizer* software is available for download at https://github.com/neurogeometry/BoutonAnalyzer. The dataset used to validate the software is available for download at the Neurogeometry group website, http://www.neurogeometry.net/resources/datasets.

7. Properties of networks as functions of memory load and robustness (Objective 3)

We performed numerical simulations for networks of up to 1000 neurons and also solved the problem theoretically in the thermodynamic limit to study the effects of memory load and robustness on structural and dynamical properties of networks. Our results show that when a network is robustly loaded with associative memories up to capacity, it develops features like those observed in many cortical systems. These features include (1) probabilities of excitatory and inhibitory connections, (2) shapes of connection weight distributions, (3) overrepresentations of specific three-neuron motifs, (4) distributions of connection numbers in clusters of 3 to 8 neurons, (5) sustained, irregular, and correlated firing activity, and (6) tight balance of excitatory and inhibitory postsynaptic potentials. The confluence of these results serves as a validation of the proposed associative memory storage model.

8. Generalization of the theory on the case of unreliable signal transmission (extension of Objective 1)

An important aspect of associative learning in the brain that has not been fully explored is that memories are recalled in the presence of noise. It is well known that sources of noise are present at every step of signal transmission through the network. These sources include errors in the presynaptic input to the neurons, noise in synaptic transmission, and fluctuations in the neurons' postsynaptic potentials. Collectively they lead to errors in the neurons' outputs which are, in turn, injected into the network, completing the noise propagation cycle. In our original model, all sources of noise were described by way of a generic robustness parameter. Subsequently, we solved a more general formulation of the associative memory storage model, in which memories are retrieved in the presence of the noise propagation cycle. Interestingly, our results show that errors and noise during learning increase the probability of memory recall. There is a tradeoff between the capacity and reliability of stored memories, and the information contained in the retrieved memories is maximal when the level of noise during learning exceeds that during memory retrieval. As before, networks loaded with associative memories to capacity display many structural and dynamical features observed in local cortical circuits. The manuscript describing these results was submitted for publication. The numerical algorithms developed for the solution of the associative memory storage model in the presence of errors and noise are available at https://github.com/neurogeometry/Associative_Learning_with_Noise.

9. Associative learning in a structurally-constrained cortical column (Objective 2)

Inspired by the results of the associative memory storage model for local networks, we extended the model to a structurally constrained network of the cortical column. To that end, we used 55 morphological types from the Blue Brain Project and constructed an instance of structurally (potentially) connected cortical column. The column includes 31,346 neurons, totaling ~40,000,000 potential connections. We loaded associative memory sequences into the column and obtained its connectivity. We showed that as in the case of local networks, associative learning alone is enough to explain the bulk of experimental knowledge about connectivity between different neuron classes and network dynamics in the column. The manuscript related to these results is in preparation.

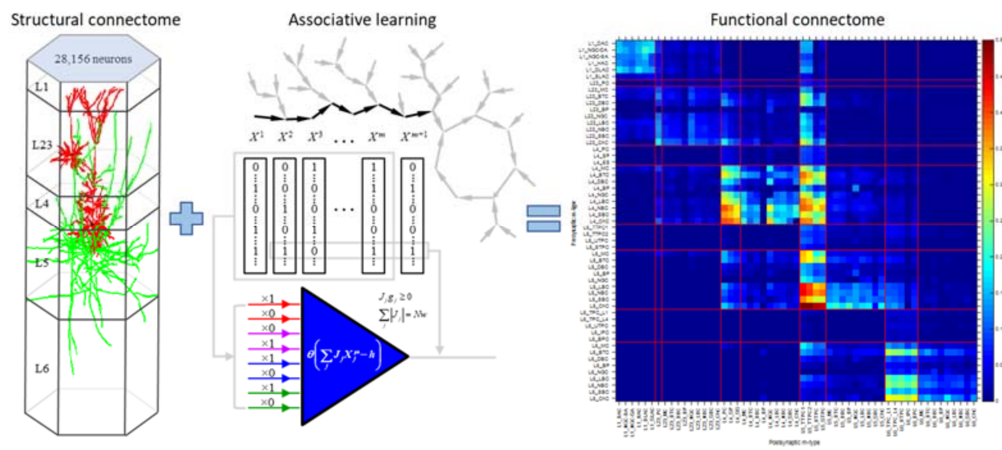

Figure 2: Functional connectivity results from biologically-constrained associative learning in a morphologically-constrained neural network.

10. Online learning without catastrophic forgetting (new objective)

Long-term memories in the brain are stored in patterns and strengths of synaptic connections, and it is believed that a single memory is written into a large population of synapses and that each synapse is involved in multiple memories. Due to this distributed and shared nature of memory representation, new learning can affect synapses involved in existing memories, leading to indiscriminate, or catastrophic, forgetting. How catastrophic forgetting is avoided in the brain is not known, and there are no adequate strategies for dealing with this problem in artificial neural networks. We use numerical simulations to show that, in the associative memory storage model, memory lifetime increases super-linearly with robustness. Therefore, it is possible to substantially extend the lifetimes of high-priority memories by loading them into the network with relatively high values of robustness. This will minimize the need for memory maintenance or retraining.

Personnel supported

- Armen Stepanyants, Faculty PI
- Danke Zhang, Postdoctoral Research Associate
- Chi Zhang, Graduate Student

Manuscripts in preparation

- Zhang, D., Zhang, C., and Stepanyants, A., Reconstructing the connectome of a cortical column through biologically constrained associative learning
- Gala, R. and Stepanyants, A., Anatomical correlates of perceptual learning in mouse barrel cortex

Articles published and submitted

- Zhang, C., Zhang, D., and Stepanyants, A., Noise in neurons and synapses enables reliable associative memory storage in local cortical circuits (submitted, an abridged version is available on **bioRxiv** 10.1101/583922)
- Zhang, D., Zhang, C., and Stepanyants, A., Robust associative learning is sufficient to explain the structural and dynamical properties of local cortical circuits, **J Neuroscience**, 3218-18 (2019)
- Gala, R., Lebrecht, D., Sahlender, D.A., Jorstad, A., Knott, G., Holtmaat, A., and Stepanyants, A., Computer assisted detection of axonal bouton structural plasticity in *in vivo* time-lapse images, **eLife**, 6:e29315 (2017)
- Mizusaki, B.E.P, Stepanyants, A., Chklovskii, D.B., and Sjöström, P.J., Neocortex: a lean mean memory storage machine, **Nature Neuroscience** (News & Views), 19(5) 643-644 (2016)

Peer-reviewed conference abstracts

- Zhang, D., Zhang, C., and Stepanyants, A., Reconstructing connectome of the cortical column with biologically-constrained associative learning, **Organization for Computational Neuroscience Annual Meeting**, BMC Neuroscience 20 (1), 56: P292 (2019)

- Zhang, C., Zhang, D., and Stepanyants, A., Fluctuations in neural activity are reflected in the structure of associative memory networks, **Computational and Systems Neuroscience meeting**, II-21 (2019)
- Zhang, D., Zhang, C., and Stepanyants, A., Structural and dynamical properties of local cortical networks result from robust associative learning, **Organization for Computational Neuroscience Annual Meeting**, BMC Neuroscience 19 (2), 64: O3 (2018)
- Zhang, D., Zhang, C., and Stepanyants, A., Properties of recurrent networks at maximum capacity for storing sequences of network states, **Organization for Computational Neuroscience Annual Meeting**, BMC Neuroscience 18 (1), 60: P254 (2017)
- Zhang, D. and Stepanyants, A., Structure and dynamics of robust associative networks operating at maximum capacity, **Computational and Systems Neuroscience meeting**, III-107 (2017)
- Rohan Gala, Daniel Lebrecht, Anthony Holtmaat, and Armen Stepanyants, A three-state model helps to find anatomical correlates of perceptual learning, **Computational and Systems Neuroscience meeting**, III-30 (2016)

Presentations

- Danke Zhang, Chi Zhang, and Armen Stepanyants, "Reconstructing connectome of the cortical column with biologically-constrained associative learning", Organization for Computational Neuroscience Annual Meeting, Barcelona, Spain, July 2019
- Chi Zhang, Danke Zhang, and Armen Stepanyants, "Fluctuations in neural activity are reflected in the structure of associative memory networks", Computational and Systems Neuroscience Meeting, Lisbon, Portugal, March 2019
- Danke Zhang, Chi Zhang, and Armen Stepanyants, "Structural and dynamical properties of local cortical networks result from robust associative learning", Organization for Computational Neuroscience Annual Meeting, Seattle, WA, July 2018
- Chi Zhang, Danke Zhang, and Armen Stepanyants, "Robustness to fluctuations in neural activity is reflected in the structure of critical associative memory networks", Society for Neuroscience Annual Meeting, Washington DC, November 2017
- Danke Zhang, Chi Zhang, and Armen Stepanyants, "Structural and dynamical properties of critical memory networks", Society for Neuroscience Annual Meeting, Washington DC, November 2017
- Danke Zhang, Chi Zhang, and Armen Stepanyants, "Properties of recurrent networks at maximum capacity for storing sequences of network states", Organization for Computational Neuroscience Annual Meeting, Antwerp, Belgium, July 2017
- Chi Zhang, Danke Zhang, and Armen Stepanyants, "Structural and dynamic properties of neural networks with robust associative memory recall in the presence of fluctuations in

neuron firing", International Conference on Mathematical Neuroscience, Boulder, CO, June 2017

- Danke Zhang, Chi Zhang, and Armen Stepanyants, "Order-to-chaos phase transition in recurrent networks operating at maximum capacity for storing sequences of network states", International Conference on Mathematical Neuroscience, Boulder, CO, June 2017
- Rohan Gala, Daniel Lebrecht, Daniela Sahlender, Anne Jorstad, Graham Knott, Anthony Holtmaat, and Armen Stepanyants, "Detection of structural changes in axonal boutons in time-lapse in vivo imaging experiments", Wiring the Brain, Cold Spring Harbor, NY, April 2017
- Danke Zhang, Chi Zhang, and Armen Stepanyants, "Structure and dynamics of robust associative networks operating at maximum capacity", Computational and Systems Neuroscience Meeting, Salt Lake City, UT, February 2017
- Rohan Gala, Daniel Lebrecht, Marta-Milena Dimanico, Stephane Pages, Daniela Sahlender, Karin Morandell, Graham Knott, Daniel Huber, Anthony Holtmaat, and Armen Stepanyants, "Tracking boutons and analyzing their structural plasticity to reveal correlates of perceptual learning", Neural Circuits, Cold Spring Harbor NY, April 2016
- Rohan Gala, Daniel Lebrecht, Anthony Holtmaat, and Armen Stepanyants, "A three-state model helps to find anatomical correlates of perceptual learning", Computational and Systems Neuroscience Meeting, Salt Lake City UT, February 2016
- Rohan Gala, Daniel Lebrecht, Anthony Holtmaat, and Armen Stepanyants, "Tracking structural changes in chronic in vivo images of neurites", Society for Neuroscience Annual Meeting, Chicago IL, October 2015
- Julio Chapeton, Rohan Gala, and Armen Stepanyants, "Associative memory storage and synaptic connectivity in homeostatically constrained networks of excitatory and inhibitory neurons", Society for Neuroscience Annual Meeting, Chicago IL, October 2015
- Daniel Lebrecht, Rohan Gala, Marta-Milena Dimanico, Stéphane Pagès, Karin Morandell, Daniel Huber, Armen Stepanyants, Anthony Holtmaat, "Anatomical correlates of perceptual learning in mouse barrel cortex", Society for Neuroscience Annual Meeting, Chicago IL, October 2015

Other products

- **Dataset of connection probabilities and strengths in local brain circuits in mammals**. This dataset is a compilation of 152 articles published in peer-reviewed journals from 1990 to 2016, describing a total of 856 projections. It includes detailed information about probabilities of connection and connection strengths in various animals and brain regions, for various laminar projections, classes of pre- and postsynaptic neurons, ages, experimental conditions, and recording techniques. The dataset is available at the Neurogeometry group website, http://www.neurogeometry.net/resources/datasets.

- **Dataset of correlative light and electron microscopy images of cortical axons**. This dataset includes (1) an image volume, acquired with two-photon laser scanning microscopy *in vivo* (2PLSM), showing axons of GFP expressing cortical neurons, (2) 3D electron microscopy reconstruction of these axons, their boutons, postsynaptic densities, axon centerlines and cross-sections, and (3) traces of the same axons reconstructed with NCTracer from the 2PLSM image. The dataset is available at the Neurogeometry group website, http://www.neurogeometry.net/resources/datasets.

- ***BoutonAnalyzer* software for detection and tracking of structural changes in boutons.** *BoutonAnalizer* was used to measure structural changes in *en passant* boutons during perceptual learning. The software can: (1) optimize traces of axons, (2) generate axon intensity profiles, (3) register traces across multiple imaging sessions, (4) annotate traces, (5) detect, edit, and match boutons in multiple imaging sessions, and (6) measure bouton weights. It is available for download on GitHub, https://github.com/neurogeometry/BoutonAnalyzer.

- **The associative memory storage model.** Two numerical algorithms were developed for the solution of the associative learning model. The algorithms are designed for loading large associative memory sequences of various topologies (Figure 1B) into recurrent networks of inhibitory and excitatory neurons. The first algorithm is based on the solution of Eqs. 4 with a method of convex optimization and the second on the perceptron-type learning rule described in Eqs. 5. The algorithms are available at https://github.com/neurogeometry/AssociativeLearning.

- **The associative memory storage model in the presence of errors and noise.** The numerical algorithms developed for the storage of associative memories into recurrent networks of inhibitory and excitatory neurons in the presence of errors and noise are available at https://github.com/neurogeometry/Associative_Learning_with_Noise.

Scientific impact and DoD benefits

We developed a model of a biologically-inspired recurrent neural network, capable of storing and robustly retrieving temporal sequences of network states. Based on this model, we showed that neural networks in the brain operate in a state of maximum memory storage capacity, thus also proving that there is a local learning rule capable of attaining optimal performance. We examined the effects of resource constraints and robustness to noise on network memory storage capacity, connectivity, and dynamics. We embedded the associative memory storage model in a morphologically constrained network of the cortical column, which includes 31,346 neurons and

40 million potential connections. These theoretical and computational results can help guide the development of neuromorphic hardware.

Signal transmission in the brain is accompanied by many sources of errors and noise, and yet, neural networks can reliably process sensory information, store and retrieve memories. We developed a way to incorporate various sources of noise into our model, without jeopardizing its theoretical tractability. Our results reveal that errors and noise should not be viewed as a nuisance, but that they are essential components of the reliable learning mechanism implemented by the brain. There is a tradeoff between the capacity and reliability of stored memories, and the optimal retrieval of stored information is achieved when the level of noise present during learning is larger than that during memory retrieval. This result shows how a biological system can harness noise, which is both free and inescapable, and use this power to enhance the reliability of its basic functions. The result parallels ideas from machine learning, where an augmentation of training examples with noise and dropping out neurons and connections during training have been shown to significantly reduce overfitting and training time.

Because associative learning of temporal sequences is the foundation of many cognitive functions of the brain, our results are also expected to have a profound impact on neuroscience. We have uncovered a direct link between learning and synaptic connectivity. Therefore, by measuring connection probabilities between neurons and recording connection strength in routine electrophysiological experiments, one can determine the memory load of the brain area and gauge the reliability of stored memories. We think that such knowledge can be used to understand the hierarchy of brain areas and to assess the effects of neurological disorders and aging on learning and memory.

The proposed associative memory storage model incorporates basic elements of local connectivity in the mammalian neocortex. It accounts for excitatory and inhibitory neurons, neuron morphology and structural connectivity, the homeostatic constraint on connection weights, and four types of errors and noise which accompany signal transmission in the brain. The model has been extensively validated against experimental measurements. The developed perceptron-type learning rule makes it possible to load associative memory sequences into the network in a biologically-plausible online manner. It serves as a proof of principle that a similar mechanism may be implemented in the brain. Therefore, we believe that the derived learning rule has profound implications for neuroscience, machine learning, neuromorphic computing, and, consequently, the capabilities of the DOD.