

Article

Self-organized structuring of recurrent neuronal networks for reliable information transmission

Daniel Miner¹ , Florentin Wörgötter¹, Christian Tetzlaff^{1,†}  and Michael Fauth^{1,†} 

¹ Bernstein Center for Computational Neuroscience, Third Institute of Physics, Georg-August University, Friedrich Hund Platz 1, 37077 Göttingen, Germany

* Correspondence: mfauth[at]gwdg.de

† These authors contributed equally to this work.

Abstract: Our brains process information using a layered hierarchical network architecture, with abundant connections within each layer and sparse long-range connections between layers. As these long-range connections are mostly unchanged after development, each layer has to locally self-organize in response to new inputs to enable information routing between the sparse in- and output connections. Here we demonstrate that this can be achieved by a well-established model of cortical self-organization based on a well-orchestrated interplay between several plasticity processes. After this self-organization, stimuli conveyed by sparse inputs can be rapidly read out from a layer using only very few long-range connections. To achieve this information routing, the neurons that are stimulated form feed-forward projections into the unstimulated parts of the same layer and get more neurons to represent the stimulus. Hereby, the plasticity processes ensure that each neuron only receives projections from and responds to only one stimulus such that the network is partitioned into parts with different preferred stimuli. Along this line, we show that the relation between the network activity and connectivity self-organizes to a biologically plausible regime. Finally, we argue how the emerging connectivity may minimize the metabolic cost for maintaining a network structure under the above described constraints.

Keywords: self-organization; synaptic plasticity; information transfer

1. Introduction

The brain's visual system continuously transmits perceived information along a hierarchical topology. At each layer of this topology, predominantly short-range connections form a recurrent neuronal network while long-range connections convey inputs from the previous layer and send outputs to the subsequent layer. Hereby, the long-range connections are sparse, meaning that a neuron in one layer has a rather low probability to be connected to a neuron in the subsequent (target) layer. Sparsity across layers leads directly to the problem of routing the information from an input neuron via the recurrent network of the layer to a specific output neuron.

The foundation of the brain's topology is determined during development and the coarse of the morphology of axons and dendrites remains rather rigid afterwards [1]. The chemoaffinity hypothesis formulates that during development gradients of chemical signals [2] being sensed by axonal surface molecules [3,4] guide the axons towards their destination area. Hereby the molecular composition defines a "molecular code" to specify the target neuron. However, the low number of guidance molecules [5] and the stochastic of the involved processes [6] impede the clearness of the code. By contrast, anatomical studies imply a rather unspecific assignment of neurons [7–9]. Hence, in the worst case, which we consider in this study, the sparse in- and output connections of a layer could be unstructured and assigned randomly to different neurons.

After the initial developmental phase, the brain also has the ability to adapt to sensory experiences. For this, neuronal networks modify the properties of their neurons and synapses by means of local activity-dependent plasticity processes. Most prominently, synaptic plasticity adapts the transmission efficacy –or weights– of the established connections or synapses [10–17]. On the other hand, homeostatic processes such as synaptic



Citation: Miner, D.; Wörgötter, F., Tetzlaff, C. and Fauth, M. Self-organization of short-range connections enables communication through sparse long-range connectivity. *Preprints* 2021, 1, 0. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

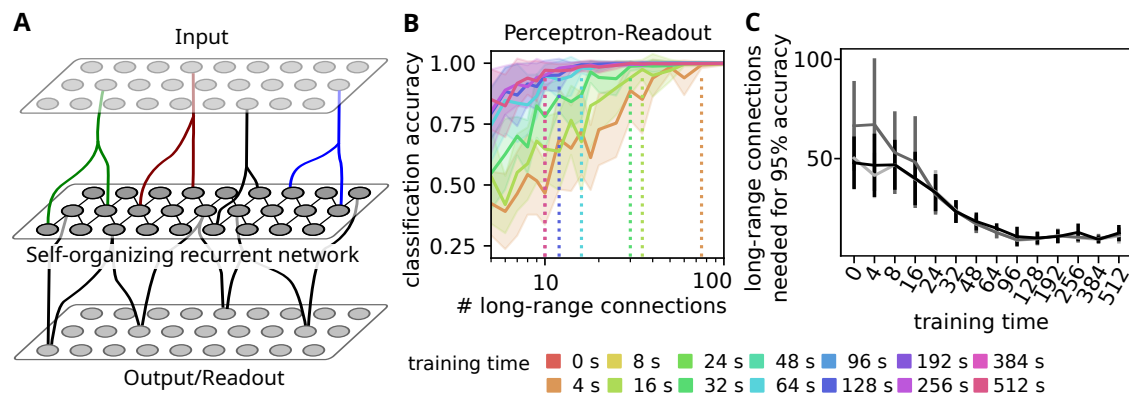


Figure 1. Self-organization of recurrent networks allows for efficient information transmission by sparse fixed in- and output-projections (A) Situation considered in this study: (Sensory) information is processed in a hierarchy of recurrent networks. Connections between layer of the hierarchy are costly, such that each layer receives sparse input and transmits outputs only from a subset of its cells via long-range output projections. As those require long-range axons, they are fixed and cannot be reorganized in response to stimulation. (B) Accuracy of a perceptron classifying the presented stimulus from the activity of a random sub-set of neurons in the self-organizing network. Different colors indicate training time (0 s for network without self-organization). Dashed lines indicate the minimum number of samples needed to achieve 95% classification accuracy. Initially around 100 neurons are needed but after around 100 s of self-organization, 10-15 neurons are sufficient to decode the stimulus identity. Transparent regions mark standard deviations of classification accuracy for a stratified 5-fold cross-validation and 6 different choices of readout connections. (C) Number of long-range connections needed to achieve at least 95% classification accuracy at different training times. Perceptron classifier (black curve) and non-neuronal classifiers such as 3-nearest-neighbor (light grey) or a (linear) support-vector machine (dark grey) arrive at similar results. Data points are mean and standard deviation of the last number of connections with more than 95% accuracy from 40 decreasing sequences.

scaling [16,18,19] and intrinsic plasticity [20,21] regulate synaptic weights and the excitability of neurons, respectively, such that the neuronal activity remains within a desired regime. A series of theoretical studies shows that the interplay of these processes leads to the self-organized formation of memory representations of observed input stimuli [19,22] and their correlations [23,24], while matching and explaining the experimentally observed connectivity on the microcircuit level [25,26]. These studies focused predominantly on investigating the neuronal and synaptic dynamics within the recurrent network forming one hierarchical layer neglecting its embedding in the hierarchical topology discussed before.

In this study, we investigate whether the interplay of activity-dependent plasticity mechanisms self-organizes a recurrent neuronal network to support the decoding of input stimuli by subsequent network layers on the condition that all connections are sparse, random and their source and target neurons are predetermined. Utilizing a mathematical neuronal network model of the visual cortex [23] matching a multitude of experimental data [25,26], in the first step, we show that the self-organizing principles lead to an optimal decoding performance despite sparse, random read-out connections. In the second step, we identify the effect of the self-organization on the neuronal activity and the information storage of each neuron in the recurrent network resulting in the decoding performance. Moreover, we investigate the underlying adaptation of the synaptic weight structure and compare it to experimental data. Finally, we provide an analysis suggesting that the resulting network architecture could be designed for low metabolic maintenance costs.

2. Methods and Materials

2.1. Model network

Simulations are implemented in the Brian 2 simulator platform [27] and follow implementations of self-organizing spiking networks from earlier work [25,28].

The self-organizing recurrent network model (Fig. 1A) consists of a population of $n_E = 1000$ excitatory neurons with absolute refractory periods of $\tau_E^{\text{refrac}} = 10$ ms and $n_I = 200$ inhibitory neurons with absolute refractory periods of $\tau_I^{\text{refrac}} = 2$ ms.

We recurrently connect the excitatory neurons (without self-connections) and interconnect the inhibitory and excitatory pool (in both directions) with random sparse connections with probability $p_{\text{connect}} = 0.04$. Recurrent inhibitory connections are neglected. Recurrent excitatory connections are given an initial strength of $w = 0.5$ nS, and all other connections are given an initial strength of $w = 1.0$ nS.

2.1.1. Neuron model

We use conductance-based leaky integrate-and-fire neurons, with membrane voltages evolving according to:

$$\frac{dv_j}{dt} = \frac{g_{\text{leak}}(v_{\text{rest}} - v_j) + g_{\text{ampa},j}(v_{\text{ampa}} - v_j) + g_{\text{gaba},j}(v_{\text{gaba}} - v_j)}{c_{\text{membrane}}} + \frac{\sigma_{\text{noise}}\xi}{\sqrt{\tau_{\text{membrane}}}}, \quad (1)$$

and synaptic conductances evolving according to

$$\frac{dg_{\text{ampa},j}}{dt} = -\frac{g_{\text{ampa},j}}{\tau_{\text{ampa}}} \quad \text{and} \quad \frac{dg_{\text{gaba},j}}{dt} = -\frac{g_{\text{gaba},j}}{\tau_{\text{gaba}}}. \quad (2)$$

Here, v_j is the membrane voltage of neuron j , g_{leak} is the leak conductance, v_{rest} is the resting potential, c_{membrane} is the membrane capacitance, τ_x is the time constant for feature x , $g_{[\text{ampa}, \text{gaba}],j}$ is the conductance for each neurotransmitter type, $v_{[\text{ampa}, \text{gaba}]}$ is the reversal potential for each neurotransmitter type. ξ is an Ornstein-Uhlenbeck noise generator, and σ_{noise} is the noise variance. Please refer to Table 1 for the chosen parameter values.

2.1.2. Intrinsic plasticity

The firing threshold of neuron j , $v_{\text{threshold}}^j$, is adaptive and follows:

$$\frac{dv_{\text{threshold}}^j}{dt} = -\eta_{\text{decay}}^{\text{ip}} \quad (3)$$

with adaptation rate $\eta_{\text{decay}}^{\text{ip}}$. When $v_j > v_{\text{threshold}}^j$, we reset the potential $v_j \leftarrow v_{\text{rest}}$ and set $v_{\text{threshold}}^j \leftarrow v_{\text{threshold}}^j + \eta_{\text{spike}}^{\text{ip}}$, $g_{\text{ampa},i} \leftarrow g_{\text{ampa},i} + w_{ij}$ and $g_{\text{gaba},i} \leftarrow g_{\text{gaba},i} - w_{ij}$. Here, $\eta_{\text{spike}}^{\text{ip}}$ is the adaptation increment for the intrinsic firing threshold plasticity, and w_{ij} is the strength of the synaptic connection from presynaptic neuron i to postsynaptic neuron j .

2.1.3. Synaptic plasticity and normalisation

The synaptic weights w_{ij} between two excitatory neurons are modified by spike timing-dependent plasticity (STDP) [14,15,29,30], which changes them according to the temporal difference between adjacent pre- and postsynaptic spikes Δt :

$$\Delta w_{ij} = \begin{cases} A_+ \exp(-\Delta t/\tau_+) & \Delta t > 0, \\ A_- \exp(\Delta t/\tau_-) & \Delta t < 0, \\ 0 & \Delta t = 0. \end{cases} \quad (4)$$

Here, $A_{+/-}$ is the plasticity amplitude, $\tau_{+/-}$ is the decay time constant, and $+$ and $-$ signify potentiation and depression, respectively.

After each STDP-induced weight change, we also implement a biologically inspired synaptic normalization mechanism [18,31,32] modeled by

$$\mathbf{W}_i \leftarrow \mathbf{W}_i \frac{W_{\text{total}}}{\|\mathbf{W}_i\|_1} \quad (5)$$

where \mathbf{W}_i is the vector of incoming weights to neuron i , $\|\mathbf{W}_i\|_1$ its L1-norm, and W_{total} the target value for the total incoming weight.

Table 1. Simulation parameter values

parameter	value	parameter	value
g_{leak}	30 nS	v_{rest}	-70 mV
c_{membrane}	300 pF	τ_{membrane}	20 ms
τ_{ampa}	2 ms	τ_{gaba}	5 ms
e_{ampa}	0 mV	e_{gaba}	-85 mV
$\eta_{\text{decay}}^{\text{ip}}$	0.2 mV/second	$\eta_{\text{spike}}^{\text{ip}}$	0.066 mV
σ_{noise}	1 mV	τ_{noise}	20 ms
A_+	0.05 nS	A_-	0.05 nS
τ_+	20 ms	τ_-	20 ms
		W_{total}	50 nS

2.2. Training and testing paradigm

Our goal is to study how network self-organization shapes the transmission of information conveyed by stimuli. To this end we simulate the self-organisation of a recurrent network which is exposed to five different stimuli for varying durations and then assess the network properties. All simulations are conducted in four phases:

1. **Warm-up:** The network is simulated without input for a period of 50 seconds to allow all dynamical variables to converge to an equilibrium distribution.
2. **Training:** In the following, five non-overlapping stimulation groups of 40 excitatory neurons are driven through strong connections (20 nS) from a group-specific Poisson spike source firing at 50Hz when activated. Every 200 ms another source is activated for 100 ms.
3. **Relaxation:** Afterwards, plasticity and inputs are turned off and the homeostatic mechanisms are allowed to re-equilibrate for 50 seconds.
4. **Testing:** The network is presented with recall cues which consist of one precisely timed input spike to all neurons in one stimulation-group conveyed through a strong (20 nS) connection. To allow for sufficient network relaxation, there is only one recall stimulus every 500 ms for 100 s.

To assess the influence of the training phase duration, we conduct new simulations with the respective training phase duration every time.

2.3. Evaluation measures

2.3.1. Classification accuracy

To assess the decodability of stimulus information from the activities of the self-organizing recurrent network, we use standard classification algorithms which are not necessarily neuronally implemented. Hence, our focus lies on the question if information can be decoded by a subsequent network rather than how this is achieved by a neuronal system. The latter question is subject to intensive research and out of the scope for this study.

As the most biologically inspired algorithm, we employ a perceptron classifier using the implementation from the scikitlearn- framework [33]. This algorithm uses a single

readout neuron for each stimulus-class, which is trained to display high activity when the respective stimulus has been presented and low activity for all others. The prediction of the classifier amounts to the stimulus corresponding to the readout neuron with maximal activity (one-vs-all multi-class classification).

As a comparison we also use well-established machine learning algorithms: a support vector machine with linear kernels and a k-nearest neighbor classifier (with $k=3$), both also using the standardized implementation from the scikitlearn-framework.

To determine, if a certain number of long-range projections allows sufficient decoding, we repeatedly randomly select the given number of neurons, and train the above classifiers on their immediate spiking response to different stimuli (2.5ms in 0.5ms bins). To estimate the classification accuracy, we use only 80% of the responses for training and test the prediction on the remaining 20%. This process is repeated 5 times using a stratified cross-validation strategy.

2.3.2. Mutual information

As a model free measure of the information that the activity of each single neuron carries about the stimulus we use mutual information. Specifically, we calculate the mutual information between single neuron responses (spike or no spike) and the stimulus by subtracting the stimulus conditioned entropy of the responses from the unconditioned entropy of the responses

$$MI(response, stimulus) = H(response) - H(response|stimulus).$$

To determine these entropies, we use the probability p_i that a neuron fires within the first 2.5ms after presentation of stimulus $i \in \{A, B, C, D, E\}$:

$$MI(response, stimulus) = H_2\left(\frac{1}{5} \sum_{i \in \{A, B, C, D, E\}} p_i\right) - \sum_{i \in \{A, B, C, D, E\}} H_2(p_i)$$

where $H_2(p) = -p \cdot \log_2(p) - (1-p) \cdot \log_2(1-p)$. Note that, as the neurons response is binary (as time is not taken into account), this mutual information is maximally 1 bit. However, if the cell only responds to one stimulus, but with 100% probability, the value would be $-0.2 \log_2(0.2) - 0 = 0.46$ bit, assuming the stimuli occur equally often.

2.3.3. Decoding accuracy depending on response probabilities

We consider the situation where neurons respond to one preferred stimulus with probability p_{on} and to all other stimuli with p_{off} . We assume that our decoder receives input from an equal number n_{con} of neurons tuned to each of the n_{stim} stimuli and the firing of these neurons is independent. In that case, the probability that the neurons tuned to the presented stimulus fire more than any other group of neurons is:

$$P_{decode}(p_{on}, p_{off}) = \sum_{k=1}^{n_{con}} \binom{n_{con}}{k} p_{on}^k (1-p_{on})^{n_{con}-k} \left(\sum_{\tilde{k}=0}^{k-1} \binom{n_{con}}{\tilde{k}} p_{off}^{\tilde{k}} (1-p_{off})^{n_{con}-\tilde{k}} \right)^{n_{stim}-1}$$

For a given number of stimuli, we evaluated these probabilities for all combinations of p_{on} and p_{off} and obtained the 95% isolines for a varying number of connections per stimulus n_{con} . The number of necessary long-range connections is determined as $n_{stim} \cdot n_{con}$. In Fig. 2F the resulting isolines are shown for five stimuli corresponding to the situation in the simulations.

2.3.4. Correlation dependent densities

To assess whether network self-organization entails strong connectivity between correlated neurons as observed in [34], we perform a similar analysis as in these experiments: For each existing synapse, we determined the correlation of a pre- and postsynaptic neuron

pair using the Pearson correlation coefficient of the stimulus responses of the respective neurons. Hereby, we restricted the responses for each trial to the first 2.5ms after the stimulus presentation. We then sorted the synapses according to these correlation coefficients and evaluated the cumulative sum of synaptic weights as well as the number of synapses for increasing correlations, similar to the analysis conducted in [34].

2.4. Number of long-range connections needed to decode partly tuned networks

We consider a network where a fraction f of the neurons are tuned to one of n_{stim} stimuli and respond to it with $p_{on} = 1$ while not responding to other stimuli. The remaining fraction $1 - f$ of neurons is untuned. We further assume that the preferred stimuli are randomly chosen from the n_{stim} stimuli and every stimulus occurs with equal probability f/n_{stim} . The stimulus identity can be reliably read out from the network, if at least $n_{stim} - 1$ differently tuned neurons can be identified by the read outs. We determine the probability to observe at least $n_{stim} - 1$ differently tuned neurons after k draws following [35]:

$$p(k) = \sum_{l=0}^k \binom{k}{l} f^{k-l} (1-f)^l \sum_{m=1}^{n_{stim}} \sum_{s=0}^{\min(n_{stim}-m, 1)} \binom{n_{stim}}{m} (-1)^{n_{stim}-m-s} (n_{stim}-m)^s \left(\frac{m}{n_{stim}}\right)^{k-l}$$

97 .

To determine the minimal number of needed long-range connections to decode all stimuli with 95% certainty, we steadily increase the number of draws k until $p(k) > 0.95$. An analytic approximation of the number of needed long-range connections can be found in the appendix.

3. Results

3.1. Self-organization improves stimulus decoding through sparse readouts

We investigate the stimulation, self-organization and readout of a self-organizing recurrent network (SORN) comprising 1200 conductance-based leaky integrate-and-fire neurons equipped with intrinsic plasticity, spike-timing-dependent plasticity and synaptic normalisation [23,25].

The network is presented with five stimuli, which are modelled by increased rates in one of five input populations, whose activity is conveyed to the network by sparse input connections. The network is then allowed to self-organize in response to these stimuli for a specified training time.

We then test, whether the stimulus can be identified from a sparse readout that only sees the activity of a small subset of neurons in the recurrent network (Fig. 1A). For this, we track the spiking activity relative to the stimulus presentation time and try to decode the stimulus from a sub-set of the recorded neuron activities using multiple classifiers. To this end, we repeatedly select a subset of neurons and divide their responses into a training set, on which a decoder is trained, and a test set on which its classification accuracy is evaluated. At the beginning of the training the decoding accuracy rises only slowly with the number of neurons and a good accuracy above 95% is only reached with around 100 neurons (Fig. 1B and C). After around 100s of training the accuracy rises much faster such that only 10-15 neurons are needed to decode the stimulus with high accuracy (Fig. 1B). This result is preserved when other classifiers are used (Fig. 1C).

Thus, we conclude that self-organization improves the readout of stimulus information by sparse randomly distributed long-range connections to other layers or brain areas.

3.2. Self-organization distributes information by tuning all neurons to a single stimulus.

In the next step, we aim to identify the features of self-organized neuronal code in the recurrent network that allows for the above described improved decoding. Most prominently, we find that, after a short training interval, also the cells that receive no stimulation exhibit a rapid response, mostly to one of the stimuli (Fig. 2A). However, as the actual time of the spike is slightly jittered from trial to trial (Fig. 2A), we also tracked the cumulative probability that a cell spikes within the first 2.5ms after the stimulus

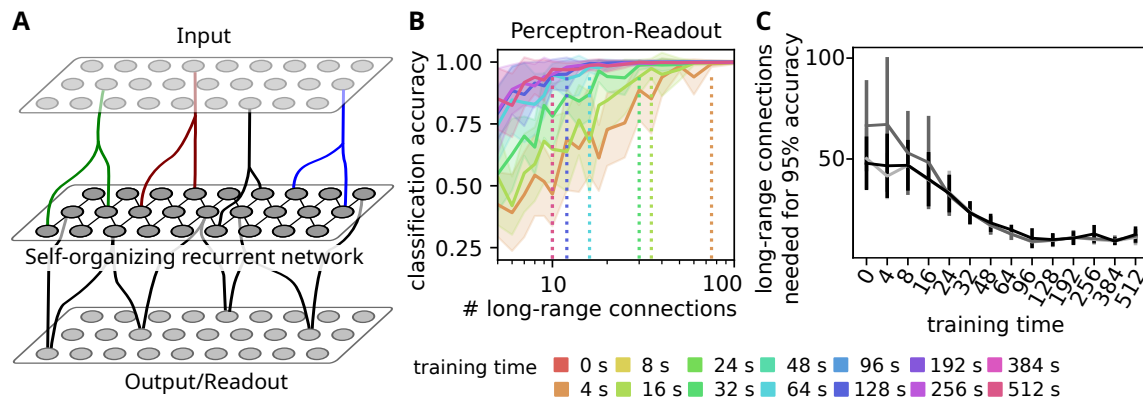


Figure 2. Self-organization tunes all neurons reliably to a single stimulus allowing for readout with less connections. (A) Stimulus triggered responses (spike rasters for 10 trials and mean response) to the five possible stimuli for three (unstimulated) cells (colors) after 48 s of training. At this stage, clear preferences for single stimuli are visible. (B) Spiking probability withing the fist 2.5 ms after the stimulus for the three example cells. If the probability for a stimulus exceeds the threshold (dashed line), the cell is considered to be tuned to that stimulus. (C) Distributions of mutual information between stimulus and single neuron activity (up to 2.5 ms after the stimulus) at different training times. (color coded as in Fig. 1B). (D) Distribution of the mutual information for neurons that respond to the same number of stimuli (with more than 20% probability) indicated on the x-axis. Neurons responding to one or two stimuli convey maximal information. Color code for number of tunings as in panel E. (E) Evolution of the the fraction of neurons that responds to $n=0,1,2,3$ or 4 stimuli. (F) Response probabilities for preferred stimulus (p_{on} , x-axis) and all other stimuli (p_{off} , y-axis) determine the number of readout connections needed to decode the stimulus. Blue curves mark the boundary between combinations of response probabilities for which the stimulus can be decoded from the indicated number of neurons per stimulus with a 95% accuracy. Colored dots mark the mean response probabilities observed in simulations. Color code as in Fig. 1B. (G) Fraction of neurons responsive to n -stimuli depending on the response threshold at different times during network training (indicated in title). Color code for number of tunings as in panel E. After training most neurons respond to a single stimulus.

132 presentation. As expected from the spike-rasters, these distribution exhibit clear peaks for
 133 one of the stimuli and only small probabilities to fire for one of the other stimuli (Fig. 2B).

134 We then set a threshold for the stimulus-dependent firing probability (20%, dashed
 135 line in Fig. 2B) above which a cell is considered as tuned to the respective stimulus and
 136 evaluated how many cells are tuned to one, two or more stimuli at different training times.
 137 While initially most cells are tuned to no stimulus (Fig. 2E, blue curve) and only the 200
 138 cells that are stimulated are tuned to one stimulus (orange curve), the number of cells that
 139 are tuned to no stimulus decreases continuously over the first 100 s of training, such that at
 140 the end, every cell responds to external stimuli. Interestingly, nearly all of the cells become
 141 tuned to only one of the stimuli (orange curve), while double or triple tunings remain
 142 sparse (green/ red curves).

143 We wondered how this strong preference for one of the stimuli influences the infor-
 144 mation that the activity of a single cell carries about the stimulus. Thus, we evaluated
 145 the mutual information between the stimulus and the activity of individual cells and
 146 tracked the distribution of this mutual information over time (Fig. 2C; thickness of the
 147 shape signifies the relative frequency of mutual information value indicated at y-axis at the
 148 training time indicated at the x-axis). Initially the activity of only a few cells – presumably
 149 the stimulated ones – carries information about the stimulus, while the activity of all other
 150 cells is uninformative. As the number of cells that are tuned to at least one stimulus rises,
 151 the probability mass shifts towards higher mutual information values until ultimately the
 152 activity of each individual cell carries around 0.5 bit of information about the stimulus
 153 (training times larger than 96s). Note, while in principle the spiking probability could
 154 carry up to 1 bit of information, a perfectly tuned cell responding to one stimulus with
 155 100% probability and remaining silent for the rest of the stimuli has a mutual information
 156 value of 0.46 bit. This indicates that self-organization does not necessarily lead to a optimal
 157 representation of information in the spiking activity, but to one that can be easily read

out (see below). Yet, when the mutual information distributions are calculated separately for cells that are tuned to different numbers of stimuli (Fig. 2D), we find that cells with a tuning for one stimulus actually yield slightly higher information content than cells with multiple tunings. Hence, for the network we consider here, the final state with a single tuning for each cell might be the best choice.

The number of tunings, however, could strongly depend on the threshold for the firing probability that is used. Therefore, we varied the threshold for tuning and evaluated the fraction of tuned cells at different training durations (indicated in panel title, Fig. 2G). The tendency to acquire a single tuning (orange) is observed over a large range of thresholds. Tunings to two or more stimuli are mostly observed for very small thresholds during the beginning of training (Fig. 2G). At very high thresholds, the number of untuned neurons (blue) first shrinks and then rises again, which can be attributed to intrinsic plasticity which prevents the neurons from being overly active and responding to each stimulus presentation with 100% probability.

3.3. Decodability improves through increasing the response to preferred stimulus

In summary, the above results indicate that the majority of neurons in the network responds to a single stimulus, but with a probability below 1. We use this simplified view to gain an analytical insight, how many long-range connections would be needed to read out from such a trained network and decode the stimulus with a given accuracy. For this we consider the situation, where each neuron is tuned to one preferred stimulus and responds to it with probability p_{on} , whereas it responds to all other stimuli with probability p_{off} . For a given number of long-range connections per stimulus, we can then calculate which combinations of p_{on} and p_{off} would allow us to discern the correct stimulus in 95% of the cases (by choosing the population with the maximal activity). In general, the higher the response probability to the preferred stimulus, the higher also the response probability for the other stimuli can be (Fig. 2F, blue curves with number of long-range connections indicated). Using less connections however, will require higher response probabilities to the preferred stimulus or smaller ones to the not preferred stimulus. We then determined the average of these response probabilities for our simulated network in different training stages. The self-organization initially increases the average response probability to the preferred, but also to the not preferred stimuli (Fig. 2F, square markers, color indicates training time). Later on, the response to the non-preferred stimuli is decreased again, most likely due to the homeostatic mechanisms that prevent excessive firing. Comparing the time-course of the response probabilities to the curves indicating the number of connections necessary for decoding reveals that self-organization transforms the network from a state where around 20 long-range connections would be needed per stimulus, to a point where only one or two connections are necessary for a reliable readout, mainly by increasing the response probability to the preferred stimulus.

3.4. Tuning spreads due to strong feed-forward connections

The phenomenon that most cells tune to a single stimulus can be explained by the self-organisation mechanisms in our network: The intrinsic plasticity mechanism adapts the firing threshold of the unstimulated cells to prevent them from being inactive. Hence, they eventually become very sensitive and can be triggered via random connections by the stimulated cells. The resulting correlated firing leads to a strengthening of those connections by STDP. As stronger weights also induce more correlated firing, this constitutes a positive feedback loop. However, synaptic normalisation also introduces competition between input weights to a cell and, hence, a winner-take-all mechanism that ultimately only allows strong weights from the stimulated cells of a single stimulus.

Evidence for this mechanism can be found in the time evolution of histograms of different weight types in our network (Fig. 3A,C). First of all, neurons that receive the same stimulus exhibit correlated firing, such that the weights between them (intra-stimulus, red) are continuously growing. However, also the weights that project from the stimulated

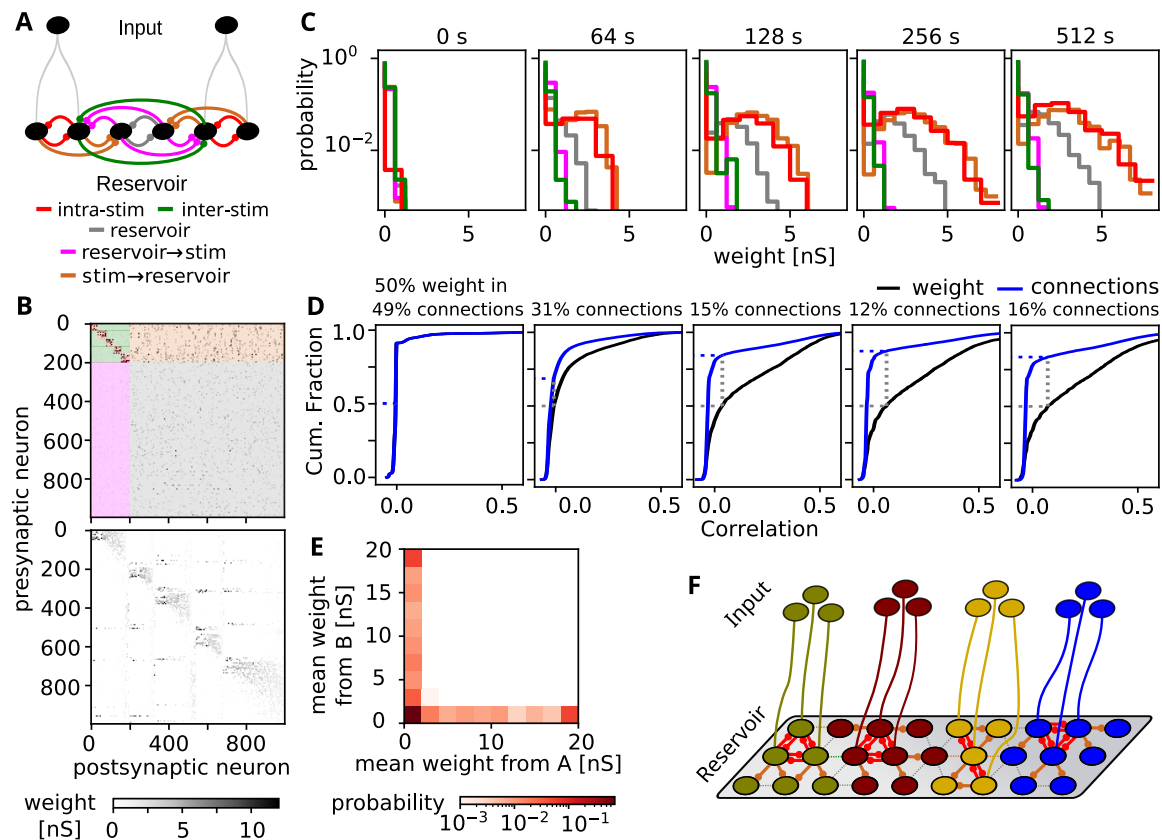


Figure 3. Self-organization of connectivity during training (A) Connection-types emerging from the stimulation paradigm: We differentiate between intra-stimulus connections (red) between neurons stimulated by the the same input-group and inter-stimulus connections (green) between cells stimulated by different inputs. Furthermore there are connections from stimulated cells to the unstimulated cells (brown), between unstimulated cells (grey) and from the unstimulated to stimulated cells (magenta). (B, top) Excitatory synaptic weight matrix after 512 s training. Note the recurrent connection blocks between the stimulated neurons (0-199) and the feed-forward projections to the unstimulated neurons. Color code as defined in panel A. (B, bottom) Weight matrix with neurons resorted according to the stimulated neuron group from which they receive the maximum input and the size of that weight. (C) Histogram of the synaptic weights of the different connection types. Connections within and from stimulated groups form stronger weights. Also weights between unstimulated neurons increase. (D) Cumulative distribution of synaptic connections (blue) and synaptic weight (black) with connections sorted according to the correlation between pre- and postsynaptic signals. Similar as in experimental results from visual cortex (compare [36]), large fractions of the total synaptic weight stem from synapses between neuron pairs with high signal correlation. (E) Pairwise histogram of the mean incoming weights from the stimulated neuron groups receiving stimuli A and B for the same neuron. Neurons with large weights from one stimulated group do not have large weights from the other stimulated group. (F) Schematic drawing of the connectivity resulting by training. The network is partitioned in strongly connected sub-networks for each of the presented stimuli.

neurons to the unstimulated neurons (brown) grow at the same pace. The weights that project back to the stimulated neurons, however, remain weak that could arise from the asymmetric nature of the STDP rule, which senses that the stimulated neurons trigger the firing in the unstimulated ones, but not vice versa. Weights between unstimulated neurons also grow, although a bit less than those from stimulated neurons. Finally, firing between neurons that receive different stimuli are uncorrelated and the weights between them remain low (inter-stimulus, green).

As a consequence, after training, we observe strong recurrent connections between the neurons receiving the same stimulus (Fig. 3B top, indices 0-199, five groups with 40 neurons). Moreover, we observe strong connections from these neurons to the rest of the network (upper right block), but not back (lower left block) or between the unstimulated neurons (lower right block). When the neuron indices are rearranged according to the

stimulated group from which they receive the strongest cumulative input weight and the size of that weight (Fig. 3B bottom), the connectivity matrix exhibits a block diagonal structure, which is in line with the observation that each neuron tunes to one stimulus.

Furthermore, neurons that receive strong weights from one stimulated group, do not receive strong weights from another (Fig. 3E), which likely emerges by the winner-take-all structure induced by synaptic normalization.

In summary, we see a spread out of tuned stimulus responses from the stimulated neurons to the whole unstimulated network (Fig. 3F). Thus, the self-organizing dynamics determined by the interplay between STDP, synaptic normalization, and intrinsic plasticity allocates the maximum possible neuronal resources to represent the stimuli. Intuitively, this spread-out of the stimulus representation due to self-organization is beneficial for a sparse readout by a subsequent network layer or brain area. This is because all long-range readout connections – no matter how they are distributed – can be expected to sample from tuned neurons whose activity carries information about the stimulus (see above).

3.5. Comparison to experimental findings

We wondered whether the above-described interplay between neural activity and connectivity is biological plausible. To this end, we repeated an analysis from visual cortex [34,36,37], which evaluates how the correlation between the activities of neurons influences the synaptic weight of the connection between them. We find that after the self-organization of the network, a large fraction of the total synaptic weight in the network is distributed to a small fraction of connections between neurons with the strongest correlation (Fig. 3D, fraction of connections for 50% of the total synaptic weight indicated in title). This finding is in line with experimental results from visual cortex [34,36,37] indicating that the here discussed self-organization of activity and connectivity is biologically plausible.

3.6. Is the spread-out of stimulus representation energy-efficient?

One reason why the long-range connectivity between layers is sparse may be to save energy. A cost-efficient implementation of neural computation provides an evolutionary advantage. Hence, it is assumed that networks are arranged such that the total wiring cost stays small [38,39]. However, to allow for a proper readout with less long-range connections (LRC), our network uses strong recurrent or short-range connections (SRC). Thus, it is unclear whether and under which conditions the self-organization principles discovered here decrease the metabolic cost of the network structure.

To investigate this, we assume that the metabolic cost E to maintain the connectivity of the network is proportional to a weighted sum of number of short-range connections n_{SRC} and the number of long-range connections n_{LRC} : $E \propto E_{LRC}n_{LRC} + E_{SRC}n_{SRC}$, where E_{LRC} and E_{SRC} signify the metabolic cost per long-range and short-range connection, respectively. Using the ratio $\gamma = E_{LRC}/E_{SRC}$, which indicates how much more expensive a long-range connection is compared to a short-range connections, this can be simplified to

$$E \propto \gamma n_{LRC} + n_{SRC}.$$

Hereby, both n_{SRC} and n_{LRC} are determined by the architecture of the network. In the following we will analyze the metabolic cost for the class of network architectures which emerge during the above studied self-organization process. In these networks, a fraction f of the neurons are assumed to acquire a tuning and to respond to one stimulus (with $p_{on} = 1$), whereas the remaining fraction $1 - f$ of neurons shows no stimulus specific response.

We further assume that, as in our self-organized network, each tuned neuron receives strong short-range connections from all input neurons that are stimulated by its preferred stimulus (if these connections physically exist). Thus, for a network with N neurons, the number of short range connections is approximately $n_{SRC} = fNp_{con}\beta$, where p_{con} is the connection probability within the network and β the number of stimulated cells for each individual stimulus.

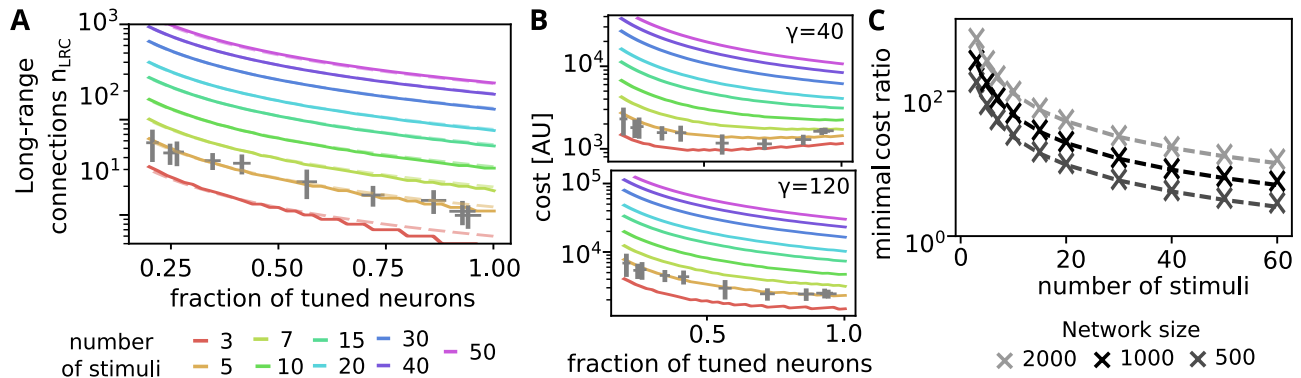


Figure 4. Energy-efficiency of information transmission depending on the fraction of tuned neurons. (A) Number of long-range connections needed to sample at least $n_{stim} - 1$ different tunings, assuming that a fraction f of the neurons are tuned to a single stimulus and all tunings are equally probable. Color indicates the number of stimuli n_{stim} . For each number of stimuli, the dependency on f is well fit by the function L/f (dashed lines). Grey error-bars mark the simulation corresponding to $n_{stim} = 5$ (fraction of single tuned neurons vs long-range connections needed for classification). (B) Metabolic cost for connection maintenance assuming that long-range connections are 40x (top) or 120x (bottom) more expensive than the short-range connections in the recurrent network. Grey bars mark simulation results (as above, number of short-range connections from weights above 5 nS). (C) Maximal cost-ratio between long- and short-range connections for which a number of stimuli (x-axis) is transmitted most cost-efficient through the here presented self-organizing recurrent network. Gray scale indicates the size of the recurrent network, which determines the number of short range connections.

266 The number of long-range connections that are needed to correctly read out the stimulus
 267 identity has been determined numerically for different fractions of tuned neurons and
 268 different numbers of stimuli through simulation (Fig. 4A, solid curves). These numbers
 269 of long-range connections match well with those needed to correctly read out from our
 270 self-organizing network at different training stages (grey crosses, n_{LRC} and f determined
 271 as in Fig 1C and Fig 2E, respectively). Using an analytical approximation (see appendix),
 272 we found that it scales with f like $n_{LRC} = L(n_{stim})/f$ (best fits shown in Fig. 4A, dashed
 273 lines).

274 Using the above relations between the fraction of tuned neurons and the numbers of
 275 long- and short-range connections, we can determine the metabolic costs for different cost
 276 ratios γ (Fig. 4B top/bottom).

For small numbers of stimuli and small cost ratios, there is an optimal fraction of tuned neurons below 1 (Fig. 4B top for $n_{stim} \in \{3, 5\}$). This optimal fraction of tuned neurons at which the total metabolic cost is minimized can be expressed as

$$f_{min} = \sqrt{\gamma L(n_{stim}) / (N p_{con} \beta)}.$$

277 For all other cases (where $f_{min} > 1$) the metabolic cost is monotonically decreasing with f
 278 (Fig. 4B), such that networks with $f = 1$ are the most cost efficient architecture. Thus, in
 279 these cases the here investigated self-organization, which drives the network towards $f = 1$
 280 (see above), is constructing a network architecture that transmits the stimulus information
 281 with minimal metabolic cost.

The minimal cost ratio, for which this ($f = 1$) is the case, can be determined from the above expression as

$$\gamma_{min} = \underbrace{p_{con} \cdot \beta}_{\text{in-degree of tuned cells}} \cdot \underbrace{N / L(n_{stim})}_{\approx \text{tuned cells per stimulus}}.$$

282 The cost ratio is higher the more cells are tuned to each stimulus and the more synapses
 283 they receive from the cells stimulated by the inputs βp_{con} . In Figure 4C we show how this

minimal cost ratio scales for different network sizes and number of stimuli (using $\beta = 40$ and $p_{con} = 0.04$ as in the simulations).

We would like to point out, that the metabolic cost to maintain a connection depends on the axonal membrane area [40–42], which is in turn proportional to the axonal length. The length-ratios between local projections (order of $100\mu m$) and long-range projections (order of millimeters) can be expected to be around 10-100. As this already exceeds most of the minimal cost ratios determined in Figure 4C, the here discussed self-organization, which entails that every neuron in the network responds to one stimulus, may indeed be a biologically plausible mechanism to ensure a metabolically cost efficient routing of stimulus information.

4. Discussion

In this study we used a well established model of the self-organization of cortical connectivity and demonstrated that the resulting intra-layer connectivity supports information processing and transmission using sparse long-range connections to other brain areas higher or deeper in the cortical hierarchy. Specifically, we demonstrated that the stimulated neurons form feedforward projections into the unstimulated parts of the network and thereby acquire more neurons to represent the stimulus by rapid stereotypic spiking responses. Hereby, each neuron only receives feedforward projections from one externally stimulated group such that the network is partitioned into parts with different preferred stimuli (Fig. 3F). Finally we showed that the self-organization is consistent with experimental findings and leads to metabolically cost efficient network architectures.

4.1. Model predictions

Several pre- and postdictions can be derived from our model: First of all, it predicts that each neuron is tuned to a stimulus. This is in line with neurons establishing new tunings when they are deafferented (e.g. [43]). Note, however, in a given experiment it is still possible to observe neurons which are not tuned to any experimental parameter as the animals still encounter other stimuli outside of the experiment.

Secondly, we see that an increasing training and adaptation time increases the response probabilities of the neurons. Hence, the neural responses to a stimulus will become more and more stereotypical when training continues. Similar effects have been observed in the olfactory domain [44]

Most importantly, however, we predict that an increased entrainment to a set of stimuli will enable the identification of the stimulus from smaller and smaller subsets of neurons. This might be verified by future experiments where learning is tracked in animals with implanted electrode arrays.

4.2. Limitations and possible extensions

In this study we focused on routing of uncorrelated stimuli which are moreover conveyed to disjunct or orthogonal groups of neurons. Interestingly, the sets of neurons responding to the stimuli are also disjunct / orthogonal such that this assumption is self-consistent with the transmission of stimulus information over many layers. However, often there is a significant correlation between presented stimuli and the corresponding responses. This may lead to cells with mixed responses that support complex computations [45]. Moreover, stimuli might not stem from distinct classes but rather from a continuum, as for example orientation and speed of bars and moving gratings in the visual system. In this case we would also expect more continuous tuning curves. An investigation of the self-organisation and information routing in response to correlated stimuli is therefore an interesting direction for future work.

Author Contributions: Investigation, D.M, Formal Analysis, D.M and M.F, Methodology, D.M, M.F, C.T, Visualization and Writing—original draft preparation, M.F, Writing—review and editing, C.T., F.W., M.F, Funding Acquisition and Project Administration, C.T., F.W..

All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the H2020-FETPROACT project Plan4Act (#732266).

Institutional Review Board Statement: -

Informed Consent Statement: -

Data Availability Statement: All relevant methods to reproduce this study are given in the Materials and Methods section.

Acknowledgments:

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI Multidisciplinary Digital Publishing Institute
STDP Spike-timing-dependent plasticity

Appendix A Estimation for the number of needed long-range connections

Here we approximate the number of long-range connections that is needed to discern n_{stim} different stimuli. We assume that they randomly emerge from cells in the network. Hereby a fraction f is tuned to one of the stimuli (equally distributed, whereas a fraction of $1 - f$ is untuned.

A reliable readout needs to sample at least from $n_{stim} - 1$ different stimuli. Thus, to determine the minimal number of long-range connection, we then determine the number of samples for which the probability to observe at least $n_{stim} - 1$ different tunings is above 95%. Using the probability that two stimuli are not chosen within the first n_{LRC} connections, we get

$$0.95 > 1 - \frac{n_{stim}(n_{stim} - 1)}{2} \left[1 - 2 \cdot f \cdot \frac{1}{n_{stim}} \right]^{n_{LRC}}.$$

Hence, using $\ln(1 - x) \approx -x$ we obtain

$$n_{LRC} > \frac{\ln(2 \cdot 0.05 / n_{stim} / (n_{stim} - 1))}{\ln\left(1 - 2 \cdot f \cdot \frac{1}{n_{stim}}\right)} \approx \frac{n_{stim}}{2f} \ln\left(\frac{n_{stim}(n_{stim} - 1)}{2 \cdot 0.05}\right) =: L(n_{stim}) / f.$$

References

1. Fauth, M.; Tetzlaff, C. Opposing effects of neuronal activity on structural plasticity. *Front. Neuroanat.* **2016**, *10*, 75.
2. Sperry, R.W. Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl. Acad. Sci. U. S. A.* **1963**, *50*, 703.
3. De Wit, J.; Ghosh, A. Specification of synaptic connectivity by cell surface interactions. *Nat. Rev. Neurosci.* **2016**, *17*, 4.
4. Seiradake, E.; Jones, E.Y.; Klein, R. Structural perspectives on axon guidance. *Annu. Rev. Cell Dev. Biol.* **2016**, *32*, 577–608.
5. Stoeckli, E.T. Understanding axon guidance: are we nearly there yet? *Development* **2018**, 145.
6. Hassan, B.A.; Hiesinger, P.R. Beyond molecular codes: simple rules to wire complex brains. *Cell* **2015**, *163*, 285–291.
7. Fares, T.; Stepanyants, A. Cooperative synapse formation in the neocortex. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 16463–16468.
8. Hill, S.L.; Wang, Y.; Riachi, I.; Schürmann, F.; Markram, H. Statistical connectivity provides a sufficient foundation for specific functional connectivity in neocortical neural microcircuits. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E2885–E2894.
9. Reimann, M.W.; King, J.G.; Muller, E.B.; Ramaswamy, S.; Markram, H. An algorithm to predict the connectome of neural microcircuits. *Front. Comp. Neurosci.* **2015**, *9*, 28.
10. Hebb, D.O. *The organization of behavior: a neuropsychological theory*; Wiley book in clinical psychology, Wiley, 1949.
11. Bliss, T.V.; Lomo, T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* **1973**, *232*, 331–356.
12. Dudek, S.M.; Bear, M.F. Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 4363–4367.
13. Markram, H.; Lübke, J.; Frotscher, M.; Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **1997**, *275*, 213–215.
14. Bi, G.Q.; Poo, M.M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **1998**, *18*, 10464–10472.
15. Song, S.; Miller, K.D.; Abbott, L.F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* **2000**, *3*, 919–926. doi:10.1038/78829.
16. van Rossum, M.C.; Bi, G.Q.; Turrigiano, G.G. Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* **2000**, *20*, 8812–8821.
17. Markram, H.; Gerstner, W.; Sjöström, P.J. Spike-timing-dependent plasticity: a comprehensive overview. *Front. Synaptic Neurosci.* **2012**, *4*, 2.
18. Turrigiano, G.G.; Leslie, K.R.; Desai, N.S.; Rutherford, L.C.; Nelson, S.B. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **1998**, *391*, 892–896. doi:10.1038/36103.
19. Tetzlaff, C.; Kolodziejewski, C.; Timme, M.; Wörgötter, F. Analysis of Synaptic Scaling in Combination with Hebbian Plasticity in Several Simple Networks. *Front. Comput. Neurosci.* **2012**, *6*. doi:10.3389/fncom.2012.00036.
20. Desai, N.S.; Rutherford, L.C.; Turrigiano, G.G. Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nat. Neurosci.* **1999**, *2*, 515–520.
21. Triesch, J. A Gradient Rule for the Plasticity of a Neuron's Intrinsic Excitability. *Artificial Neural Networks: Biological Inspirations – ICANN 2005*; Duch, W.; Kacprzyk, J.; Oja, E.; Zadrozny, S., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp. 65–70.
22. Zenke, F.; Agnes, E.J.; Gerstner, W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nat Commun* **2015**, *6*, 6922. doi:10.1038/ncomms7922.
23. Klos, C.; Miner, D.; Triesch, J. Bridging structure and function: A model of sequence learning and prediction in primary visual cortex. *PLOS Comp. Biol.* **2018**, *14*, 1–22. doi:10.1371/journal.pcbi.1006187.
24. Hartmann, C.; Miner, D.C.; Triesch, J. Precise Synaptic Efficacy Alignment Suggests Potentiation Dominated Learning. *Front. Neural Circuits* **2015**, *9*, 90. doi:10.3389/fncir.2015.00090.
25. Miner, D.; Triesch, J. Plasticity-Driven Self-Organization under Topological Constraints Accounts for Non-random Features of Cortical Synaptic Wiring. *PLoS Comput. Biol.* **2016**, *12*, e1004759. doi:10.1371/journal.pcbi.1004759.
26. Zheng, P.; Dimitrakakis, C.; Triesch, J. Network self-organization explains the statistics and dynamics of synaptic connection strengths in cortex. *PLoS Comput. Biol.* **2013**, *9*, e1002848.
27. Stimberg, M.; Brette, R.; Goodman, D.F. Brian 2, an intuitive and efficient neural simulator. *eLife* **2019**, *8*, e47314. doi:10.7554/eLife.47314.
28. Miner, D.; Tetzlaff, C. Hey, look over there: Distraction effects on rapid sequence recall. *PLOS ONE* **2020**, *15*, 1–22. doi:10.1371/journal.pone.0223743.
29. Gerstner, W.; Kempter, R.; van Hemmen, J.L.; Wagner, H. A neuronal learning rule for sub-millisecond temporal coding. *Nature* **1996**, *383*, 76–78.
30. Kempter, R.; Gerstner, W.; van Hemmen, J.L. Hebbian learning and spiking neurons. *Phys. Rev. E* **1999**, *59*, 4498–4514.
31. Elliott, T. An Analysis of Synaptic Normalization in a General Class of Hebbian Models. *Neural Computation* **2003**, *15*, 937–963. doi:10.1162/08997660360581967.
32. Lazar, A.; Pipa, G.; Triesch, J. SORN: a self-organizing recurrent neural network. *Front. Comput. Neurosci.* **2009**, *3*, 23.
33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

34. Ko, H.; Cossell, L.; Baragli, C.; Antolik, J.; Clopath, C.; Hofer, S.B.; Mrsic-Flogel, T.D. The emergence of functional microcircuits in visual cortex. *Nature* **2013**, *496*, 96. doi:10.1038/nature12015.
35. Emigh, T.H. On the Number of Observed Classes from a Multinomial Distribution. *Biometrics* **1983**, *39*, 485–491.
36. Ko, H.; Hofer, S.B.; Pichler, B.; Buchanan, K.A.; Sjöström, P.J.; Mrsic-Flogel, T.D. Functional specificity of local synaptic connections in neocortical networks. *Nature* **2011**, *473*, 87–91. doi:10.1038/nature09880.
37. Ko, H.; Mrsic-Flogel, T.D.; Hofer, S.B. Emergence of Feature-Specific Connectivity in Cortical Microcircuits in the Absence of Visual Experience. *J. Neurosci.* **2014**, *34*, 9812–9816. doi:10.1523/JNEUROSCI.0875-14.2014.
38. Cuntz, H.; Forstner, F.; Borst, A.; Häusser, M. One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS Comput. Biol.* **2010**, *6*. doi:10.1371/journal.pcbi.1000877.
39. Betzel, R.F.; Bassett, D.S. Specificity and robustness of long-distance connections in weighted, interareal connectomes. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4880–E4889, doi:10.1073/pnas.1720186115.
40. Attwell, D.; Laughlin, S.B. An Energy Budget for Signaling in the Grey Matter of the Brain. *J. Cereb. Blood Flow Metab.* **2001**, *21*, 1133–1145, PMID: 11598490, doi:10.1097/00004647-200110000-00001.
41. Harris, J.J.; Jolivet, R.; Attwell, D. Synaptic energy use and supply. *Neuron*, *75*, 762–777. doi:10.1016/j.neuron.2012.08.019.
42. Harris, J.J.; Attwell, D. The energetics of CNS white matter. *J. Neurosci.*, *32*, 356–371. doi:10.1523/JNEUROSCI.3430-11.2012.
43. Keck, T.; Mrsic-Flogel, T.D.; Vaz Afonso, M.; Eysel, U.T.; Bonhoeffer, T.; Hubener, M. Massive restructuring of neuronal circuits during functional reorganization of adult visual cortex. *Nat. Neurosci.* **2008**, *11*, 1162–1167. doi:10.1038/nn.2181.
44. Bazhenov, M.; Stopfer, M.; Sejnowski, T.J.; Laurent, G. Fast Odor Learning Improves Reliability of Odor Responses in the Locust Antennal Lobe. *Neuron* **2005**, *46*, 483–492. doi:https://doi.org/10.1016/j.neuron.2005.03.022.
45. Rigotti, M.; Barak, O.; Warden, M.R.; Wang, X.J.; Daw, N.D.; Miller, E.K.; Fusi, S. The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*, 585–590. doi:10.1038/nature12160.