Localist neural plasticity identified by mutual information

Gabriele Scheler[1*], Martin L Schumann[2] and Johann Schumann[3]

[1,3*]Carl Correns Foundation for Mathematical Biology, 1030 Judson Dr, Mountain View, 94040, CA, USA.

[2]Dept. of Computer Science, Ludwig Maximilians University, Munich, Germany.

*Corresponding author(s). E-mail(s): gscheler@gmail.com;
Contributing authors: martinlschumann.com; johann.schumann@gmail.com;

Abstract

We present a model of pattern memory and retrieval with novel, technically useful and biologically realistic properties. Specifically, we enter n variations of k pattern classes (n*k patterns) onto a cortex-like balanced inhibitory-excitatory network with heterogeneous neurons, and let the pattern spread within the recurrent network. We show that we can identify high mutual-information (MI) neurons as major information-bearing elements within each pattern representation. We employ a simple one-shot adaptive (learning) process focusing on high MI neurons and inhibition. Such 'localist plasticity' has high efficiency, because it requires only few adaptations for each pattern. Specifically, we store k=10 patterns of size s=400 in a 1000/1200 neuron network. We stimulate high MI neurons and in this way recall patterns, such that the whole network represents this pattern. We assess the quality of the representation (a) before learning, when entering the pattern into a naive network and (b) after learning, on the adapted network, during recall. The recalled patterns could be easily recognized by a trained classifier. The recalled pattern 'unfolds' over the recurrent network with high similarity, albeit compressed, with respect to the original input pattern. We discuss the distribution of neuron properties in the network, and find that an initial Gaussian or uniform distribution changes into a more heavy-tailed, lognormal distribution during the adaptation process. The remarkable result is that we are able to achieve reliable pattern recall by stimulating only high information neurons. This work provides a biologically-inspired model of cortical memory and may have interesting technical applications.

Keywords: pattern memory, cortical model, information theory, symbolic abstraction

1 Introduction

Storing patterns and achieving sequence-independent recall is a problem for neural network models. We are interested in exploring the concept of localist ensemble memory in cortical networks (Carpenter, 2000; Fuster, 1998). From biology we know that there are both synaptic and intrinsic memory (Langille & Brown, 2018), which have a 'hidden' component in cell-internal memory (Scheler, 2023). Therefore we want to explore the concept of storing memory by local plasticity in neurons within ensemble-like neuronal groups. For this purpose we define a realistic excitatory-inhibitory network with recurrent interactions and use simple, visually-defined patterns (LeCun & Cortes, 2005) as inputs to the network. We then examine the representations that develop on a naive network, where 'naive' means without previously stored patterns, and in the absence

of plasticity. Representations are classified by a downstream machine learning mechanism. The machine learning mechanism stands for other, for instance, contralateral, subcortical, interpretations, which are not explicitly modelled (cf. Fig. 1).

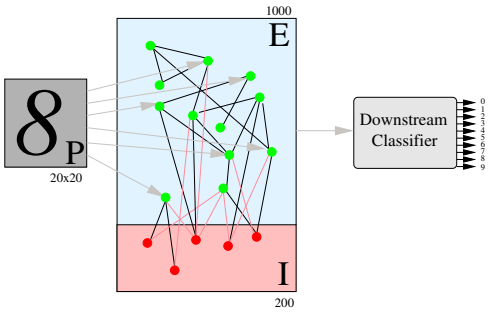

Fig. 1 Architecture of the system: patterns from a field of 20x20 excitatory P neurons are 'loaded' into a network of E neurons. The network of E (N=1000) and I neurons (N=200) generates a representation. The representation is identified by a downstream (e.g., subcortical) classifier

The construction chosen is a suitable model for cortical networks (Kim, Leahy, & Shlizerman, 2019; Scheler, 2018), it is also reminiscent of LSN (Maass, Natschläger, & Markram, 2002) and ESN (Jaeger, 2002) models. To implement plasticity, we perform a information-theoretic analysis over neurons and patterns (Schneidman, Bialek, & Berry, 2003; Sheintuch, Rubin, & Ziv, 2022) to find the neurons with the highest mutual information (MI) for each class of patterns. Such high MI response neurons form spontaneously in the representations. The existence of high MI neurons is an emergent property of the network setup. It turns out that while absolute MI values for neurons differ across classes, and the number of high MI neurons above a threshold is different for each class, we can rank the highest MI neurons for each pattern class. We select only the high MI neurons to use for plasticity. The idea is to 'compress' the pattern information into the high MI neurons' intrinsic properties and adjacency network, i.e. perform localist plasticity. Then by stimulating only these neurons, we can recall neural representations and we achieve high precision pattern recall (Scheler, 2016).

We also look at inhibitory neurons which play an important role in defining ensembles in real cortical networks. When an ensemble is activated, inhibitory plasticity guarantees that the activated ensemble is less inhibited, while inhibition remains strong for all other neurons. When there is low or no overlap in the high MI neurons for each pattern class, localist learning means that there is a disjunctive set of adaptations for each pattern set. Dissimilar pattern classes are stored free of interference. The goal of our experiments is to ensure that pattern information can be retrieved by stimulation of a few, high MI neurons alone - even though the patterns were learned from a full input representation.

2 Methods

2.1 Neurons and Network Structure

The network model is initialized as a fully connected model with E (=1000) excitatory neurons and I (=200) inhibitory neurons, plus P (=400) pattern input neurons, which are directly linked to P of E excitatory neurons, in order to 'load' patterns of length P into the network. The connections between the P neurons and corresponding E neurons is strong enough to ensure solid transfer of activity. Synaptic connections within the network are of type NMDA, AMPA and GABA-A and are modeled as in (Scheler, 2018).

Neurons are modeled as spiking neural models as in (Izhikevich, 2004) with an equation for the membrane model v and an equation for the gating variable u (Eq 1), such that v is set back to a low membrane potential $v := c$; and the gating variable u is increased by an amount d ($u := u+d$), when a neuron spikes (at $v(t) = \theta$ and $\theta = 30mV$).

$$\begin{aligned} \dot{v} &= 0.04v^2 + 5v + 140 - u - I_{syn} \\ \dot{u} &= a(bv - u) \end{aligned} \tag{1}$$

For excitatory neurons, parameters are variable in order to capture different types of neurons and generate different distributions of intrinsic excitability. For excitatory neurons, parameters a and b are varied in our model (cd. (Scheler, 2018)), which results in different intrinsic excitability of a neuron, cf. Table 1. The gain is captured by $g = \gamma \cdot a$. We measure the baseline rates as spontaneous neural activity with a background input of around $0.15nA$. This variation of parameters also allows for intrinsic plasticity.

Table 1 Parameter ranges for excitatory and inhibitory neurons

Parameter	E	I	P
a	$0.006 \ldots 0.034$	0.3	0.1
b	$0.06 \ldots 0.34$	0.1	0.1
c	-70	-70	-70
d	3	2	3
γ	$2.5 \cdot 10^4$	$5.0 \cdot 10^4$	$2.0 \cdot 10^4$
baseline rate (Hz *at* $0.15nA$)	3-7	20	6

For inhibitory neurons, the parameters are

$$a = 0.3, b = 0.1, c = -70mV, d = 2, \theta = 30mV.$$

This results in inhibitory neurons firing around 20Hz. For the input pattern neurons P, the values are similar with a lower gain and longer delay ($a = 0.1, b = 0.1, c = -70mV, d = 3$).

We initialize excitatory neurons with a Gaussian distribution over its gain and baseline rate, such that we use a mean of $\mu = 5$ and distribution $\sigma^2 = 0.96$ for the rate. Attested values for cortical neurons in mice are rate distributions with $\mu = 4.96$ and $\sigma^2 = 0.31$ (Scheler, 2017).

Initially, the network has full connectivity for E-E, E-I and I-E synapses. There are no I-I connections. Synaptic strength is set to achieve baseline activity in the network with a background input ($\sim 0.15nA$). Accordingly, the network is initialized with a Gaussian distribution of synaptic strength for both AMPA and NMDA connections between E-E ($\mu = 0.0015, \sigma = 0.00027$) and E-I ($\mu = 0.014, \sigma = 0.00024$). For GABA-A connections, we use a Gaussian distribution with $\mu = -0.5$ and $\sigma^2 = 0.1$.

There is a signal delay between neurons, which is randomly distributed:

- AMPA: $10 \ldots 30ms$
- NMDA[1]: $10 \ldots 15ms$
- GABA-A: $5 \ldots 15ms$

We do not use this parameter during plasticity.

2.2 Patterns

In order to use a simple set of variations of patterns which can be easily classified into classes, we used the MNIST database (LeCun & Cortes, 2005) for 10 handwritten digits, with 50 variations for each digit. The format was an integer vector (for the grayscale values) of length 400 (10x50=500 patterns).

We load patterns into the E+I (=1200) processing network via P (=400) pattern neurons which have a single excitatory connection (AMPA) to 400 E neurons in the processing network. We apply one input pattern of length 400 at a time for 300ms. We measure the resulting neural representation for the pattern after end of application. Spike rates recorded from excitatory neurons (E=1000 neurons) for a length of about 300 ms are regarded as the neural representation of the pattern. After that time, the neural representation fades away in our set-up.

2.3 Classification of Representations

In our work we wanted to give an overall impression of whether the adaptation method used could demonstrate pattern learning (i.e. correct interpretation of new, never-seen patterns). We used an ML system, AutoGluon (Erickson et al., 2020a; Gijsbers et al., 2023; Schumann, 2025), which is a supervised training method using a mixture-of-experts approach. It employs several mechanisms for pattern classification in parallel - all of them have access to the pattern representation, and they are tested against a validation set. It achieves its performance by combining models from a variety of learning algorithms such as Random Forests, K-Neighbors, Categorical Boosting, Gradient Boosting Machines, Neural Networks, and combined ensemble models (Erickson et al., 2020b). The results are accumulated and those with the better prediction are weighted stronger for the goal of pattern classification of new patterns.

After training the basic models, the predictions of these models on a validation data set are used to train a meta-model, which will yield a

[1] NMDA has intrinsic delays of about 120ms

weighted combination of predictions of the basic models. Figure 7 shows how three basic learning algorithms are combined to achieve a combined classification result, which has a score value higher than each of the basic components.

We simply wanted to use this to demonstrate that acceptable pattern learning occurs. We made no attempt of improving our results, or of tackling complex datasets.

The Autogluon classifier is presented with the neural representations for the $50 * 10 = 500$ patterns. The classifier's task is to recognize the correct digit for each pattern. The original neural representations for patterns are labeled, and the algorithm is trained to classify them ('supervised learning').

Combining different pattern classification methods is very likely also a technique that is used in biology to arrive at best results, results which may be best suited to specific tasks. In this respect it is remarkable that cortical activation patterns are accessible to (one could say "observed by") different networks: contralateral cortical networks, striatal networks, hippocampus, midbrain areas like substantia nigra and VTA, the cerebellum etc. Quite obviously the ability to classify patterns is different in all those brain areas, and this diversity serves many purposes, such as task-dependent classification results, and will have different levels of precision. There is no need to assume that there is a single learning algorithm used by the brain.

The test phase consists in recognizing representations from stimulating of a few high MI neurons *alone* after neural plasticity. These 'recalled' neural representations were derived from m high MI neurons where $m = 5, 10, 20$. These representations are recognized by the classifier without any labels, based on their similarity with the original representations in an 'unsupervised' manner. This is a remarkable result.

2.4 Information Theoretical Analysis

Each pattern class s is presented with equal probability $p(s) = 1/N_{digits}$, where $N_{digits} = 10$.

The response for each neuron is the rate response r. To simplify, we distinguish 3 rate responses (A,B,C), with f^s_{rel} as the neuron's firing rate for digit pattern s relative to its baseline rate (Eq. 2).

$$r = \begin{cases} A & 80\% < f^s_{rel} < 120\% \quad \text{unchanged} \\ B & f^s_{rel} \leq 80\% \quad \text{low} \\ C & f^s_{rel} \geq 120\% \quad \text{high} \end{cases} \tag{2}$$

Now for each set of 50 variations of a digit, each neuron has a probability for each of the three responses. For instance, for digit 5, neuron j may have $A = 0.8, B = 0.1, C = 0.1$, or 40 times A, 5 times B and 5 times C. Assume $p(A) = 0.5, p(B) = 0.25, p(C) = 0.25$.

We calculate the mutual information (MI) (Zeng, 2015) between each digit s (10 with 50 variations) and rate response r for each neuron n as in (Eq. 3):

$$\begin{aligned} MI^s_n &= \textstyle\sum_{r \in \{A,B,C\}} p(s,r) * \log_2 \frac{p(s|r)}{p(s)} \\ &= p(s,A) * \log_2 \frac{p(s|A)}{p(s)} \\ &\quad + p(s,B) * \log_2 \frac{p(s|B)}{p(s)} \\ &\quad + p(s,C) * \log_2 \frac{p(s|C)}{p(s)} \end{aligned} \tag{3}$$

Assume $p(digit5) = 0.1$ overall. Since A is the 'neutral' option of a fluctuating value in an intermediate range, neuron j would not be very specific in recognizing pattern 5. Its mutual information (MI^5_j) would be $(0.054 + (-0.01322) + (-0.01322) = 0.0278\ bit)$.

Figure 2 shows the overall MI for each excitatory neuron of the network after processing 500 patterns without any plasticity. The neurons are arranged according to their intrinsic baseline firing rate (right-to-left, and top-to-bottom).

We can also calculate the mutual information for each neuron with respect to each digit pattern. For each neuron, we have 10 values for MI^s_n, since there are $s = 1..10$ digit patterns. We then rank for each digit pattern the m highest MI^s_n values. This is shown in Fig. 4**a**–**c**. If a neuron appears as a high MI neuron for more than one pattern, it is marked by a black dot.

2.5 Plasticity

Adaptation is based on a neuron's information content (MI). The system has access to the mutual information of a neuron for patterns. A neuron must have high MI (for some pattern) in order to undergo plasticity. For each pattern, we rank

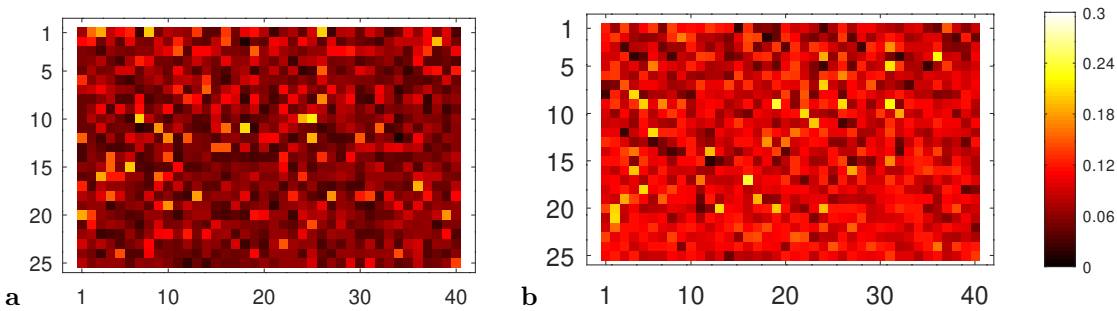

Fig. 2 No correlation between baseline firing rate and MI. 1000 E neurons are shown on a 25x40 grid. The numbering is ordered by the baseline firing rate of each neuron (left to right and top to bottom). MI is calculated for all 500 patterns. (**a**) before plasticity is applied, (**b**) after plasticity is applied

neurons by mutual information, and select the m highest neurons which are active for this pattern, for $m = 5, 10, 20$ (Fig. 4). Surprisingly, the specific size of m shows little difference in the behavior of the system.

We use two types of single-shot update rules, for intrinsic plasticity, and for synaptic plasticity.

- Intrinsic plasticity: For the adaptation of high MI neuronal parameters a, b, we use an exponential update rule:

$$a := a \cdot e^{(a-a_{min})/(\lambda(a_{max}-a_{min}))}$$

$$b := b \cdot e^{(b-b_{min})/(\kappa(b_{max}-b_{min}))}$$

  With this rule, we strengthen the excitability in a non-linear fashion with stretch factors $\lambda = 3$ and $\kappa = 3$. It turned out that a proportional update rule $(a \cdot (1 + \lambda))$ did not lead to desired results (s. Results).
- Synaptic connections:

  – For high MI neurons only, their postsynaptic connections are updated according to the activation of the input neurons i: if the response activation is high (type C, Eq 2), the connection will be strengthened

$$w_{ni} := w_{ni}(1 + \alpha_1)\,;\ \alpha_1 = 0.4$$

  The parameter was set manually such that recall was optimized. A single update of synaptic value is sufficient, and each neuron receives many unstrengthened synapses. Interestingly, a simple proportional update was sufficient. For responses A (unchanged) and B (low), w_{ni} is left unchanged.

  – Presynaptic connections are updated according to the activation response of the output neurons o: if the response is high (type C, Eq 2), the input synaptic connection is strengthened:

$$w_{ni} := w_{ni}(1 + \beta_1)\,;\ \beta_1 = 400$$

  Here $\beta_1 \sim 400$ shifts the weight to a much higher value.
  Synaptic connections for output neurons with unchanged or low responses are additionally weakened:

$$w_{no} = w_{no}(1 - \beta_2)\,;\ \beta_2 = 0.9$$

  . The updates for weakened synapses were not crucial in our experiments, they mainly serve to better calibrate the overall distributions.

- Tuning inhibition:

  - For each pattern, a few inhibitory neurons i are selected to receive strengthened connections from pattern-specific high MI neurons n:

$$w_{ni} := w_{ni} * (1 + \gamma_1)\,;\ \gamma_1 = 200$$

  - The GABA-A connections from inhibitory neurons to excitatory neurons with unchanged or negative response are also strengthened in order to suppress neurons which do not participate in coding for the pattern:

$$w_{io} := w_{io} * (1 + \gamma_2)\,;\ \gamma_2 = 8000$$

For inhibition, parameters selectively strengthen and supress synapses and in this

way a trace of the neural representation is 'carved' into the network.

In this case, meta-parameters were hand-tuned to achieve appropriate results for localist recall. Automatic setting and re-calibration of parameters could be added and might uncover more parameter combinations with acceptable solutions ('regimes'). It is most likely that calibration by homeostatic plasticity, which occurs from time to time, would also stabilize the parameter regimes, and make some weakening rules unnecessary.

The intrinsic plasticity rule means that a neuron which carries much information for a pattern will increase its intrinsic frequency. It will also adjust its input and output connections. The pattern will leave a trace at that neuron. We have seen that intrinsic frequency and high MI are initially not correlated (Fig. 2). For small sets of high MI neurons ($m/P < 0.05$), there is limited overlap in the identity of the high MI 'concept' neurons for each pattern (Fig. 4), and therefore the update rules are mostly disjoint (separate). A high MI neuron's input synapses are updated only when they receive positive pattern input from input neurons i. However, all of its output synapses are updated. As a result, we have few, local updates which furthermore are applied only once, which makes plasticity highly efficient.

For the neural parameters a and b which we use to characterize the intrinsic activation (firing rate) of the excitatory neurons, the proportional update caused problems. When parameters a and b were already high, and the neuron had a high intrinsic activation, further adaptation of the parameters could cause undesirable behavior, such as intense bursting-like activity. Such behavior is an artefact of using a parametrized neuron model which is non-optimal for extreme values. Lowering the learning rate overall would lead to a lack of intrinsic learning for the bulk of low-firing neurons. Therefore we chose an exponential form (s. supplemental Figure 1), which ensured that learning intensity was matched to the existing activity level: strong learning for low activity neurons, weaker learning for high activity neurons. It is easy to imagine a biological basis for this.

3 Results

3.1 Neural Representations and Classification

We modeled the network similar to a cortical network, which receives pattern information from thalamic input neurons (Sampathkumar, Miller-Hansen, Sherman, & Kasthuri, 2021). Accordingly, we used a vector representation of visual patterns for perceptual input and added weak fluctuating background noise (of $1.5nA$). We initialized the network with random (Gaussian) distributions of neural parameters such that the network had heterogeneous intrinsic strengths and synaptic weights. The naive network, without any plasticity, developed neural representations for, in this case, 500 pictures from 10 classes. The neural representations, which lasted $\sim 300ms$, could be analyzed for information content of individual neurons using the tool of mutual information. It turned out, like we noticed before (Scheler, 2017), that neurons with high mutual information develop for each input pattern class. These neurons could represent pattern classes in the manner of 'symbolic' or 'concept' neurons. They stand for the whole pattern. We developed a localist form of plasticity to make pattern representation permanent. Then, by stimulating only very few ($m = 5$) high-MI neurons, we recall pattern representations sufficiently to allow for correct pattern classification by a machine learning classifier (Section 3.4).

For the naive representations, without any plasticity applied, we show that the projection of the 400 pixel input pattern onto the 1200 recurrent cortical network model resulted in recognizable neural representations. Figure 3 shows neural representations for digit "6".

1. before application of plasticity, when the pattern is presented to the naive network through input neurons P (Figure 3**a**),
2. after plasticity, when pattern is presented to the learned network through input neurons P (Figure 3**b**), and
3. after plasticity, when the representation is 'recalled' from stimulation of the 'concept' high-MI neurons alone (Figure 3**c**).

A reduction of recall precision after adaptation (Figure 3**b**) which allows to generalize to unseen patterns is in accord with many results

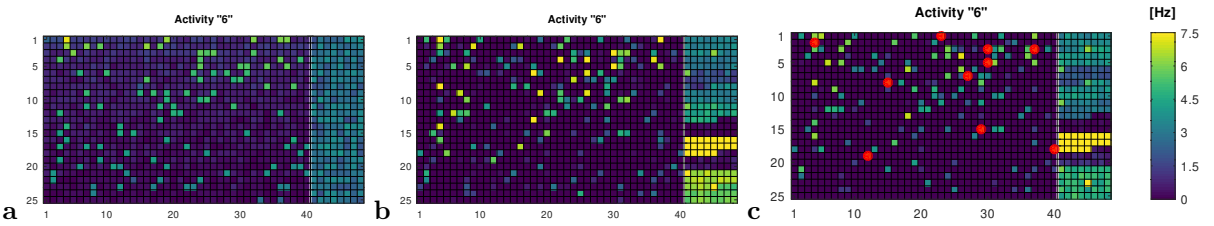

Fig. 3 Neural rate representation for digit patterns. Excitatory neurons on the left side of each panel, inhibitory neurons on the right. (**a**) Activity after presentation with a pattern "6" before plasticity, (**b**) The same after application of plasticity rules. (**c**) Activity after stimulation of 10 high MI neurons (red). Patterns (**a**–**c**) were all classified correctly as representing digit "6". The high inhibitory activity contributes to specificity of representation

from the computational ML literature. Something about compression

3.2 Mutual Information Analysis of Representations

When we analyzed the neural representations for information content, by measuring mutual information (MI) between each neuron and the 10 digit input patterns, we could see that there is no or weak correlation between high MI and high baseline rate overall in the network (Figure 2). After plasticity, the mutual information (for 500 patterns) in the network is higher for each pattern (Figure 5). This shows that the network has stored pattern-specific information. It has increased its information content specifically for each of the learned patterns.

However, the correlation between MI and baseline rate remains low ($r = -0.09$ (Pearson), and -0.10 (Spearman) before plasticity, $r = 0.048$, 0.06 after plasticity).

Our aim was to identify high MI neurons for each digit which could be regarded as 'symbolic' abstractions of the whole pattern. I.e. we were looking for the m neurons for each digit pattern, which are highest ranked by mutual information. Our results show that these neurons have a low amount of overlap (Figure 4). While with larger m more shared high MI neurons appear, there are mostly unique high MI neurons which are spread over the whole network. No calculations are offered here on the relations between data sets, network size, and number of MI neurons, which would generalize this form of plasticity beyond the chosen example.

3.3 Localist Plasticity and Recall

In order to store the representations of pattern input that appear on the naive, randomly initialised network, we apply plasticity rules. Our goal is to compress the information into selected neurons and their surrounding ensembles, and recall patterns by stimulating those high-level, 'concept' neurons. There are experimental indications for such constructions (Josselyn & Tonegawa, 2020; Park, Ko, Frankland, & Josselyn, 2024), and they also have enormous advantages in intelligence applications.

We have shown that MI is distributed in such a way that a few neurons have high MI. We want to use these neurons like symbolic abstractions. In a number of experimental publications (Agetsuma et al., 2023; Gal et al., 2021), a hub-spoke representation for each pattern has been suggested (Gu, Qi, & Gong, 2019). We believe that such structures imply great advantages for recall. It should be sufficient to target only the high-order neurons for activation, which then activate their feature structures, and in this way reconstruct the whole pattern (cf. also (Choucry, Nomoto, & Inokuchi, 2024; Uytiepo et al., 2024)).

Accordingly, we use a simple, one-shot plasticity rule, focused on high MI neurons. We may label this type of plasticity rule "localist", because it only affects few neurons and their synaptic environment. The basic idea is to form a trace of the input pattern activation as recorded in the intrinsic excitability and synaptic connections of selected neurons (cf. Section 2.5). The localist restriction of parameter updates is the decisive difference to a distributed network update rule, and ensures its specific recall properties. It is also computationally very efficient.

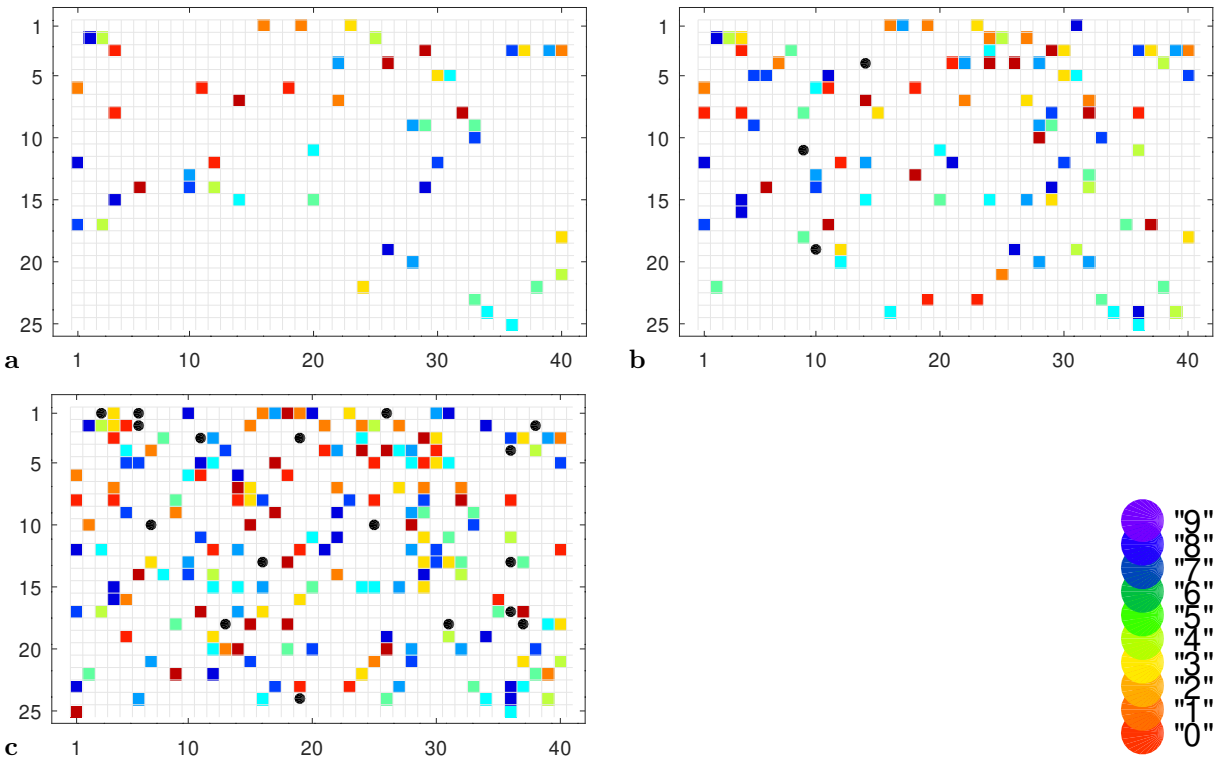

Fig. 4 Ranked high MI neurons (m) for a naive network representation, unique for each digit (0-9), marked by color. (**a**) $m = 5$ neurons per pattern, (**b**) $m = 10$ neurons, and (**c**) $m = 20$ neurons. Shared high MI neurons are marked as round and black. They are sparse even at higher m

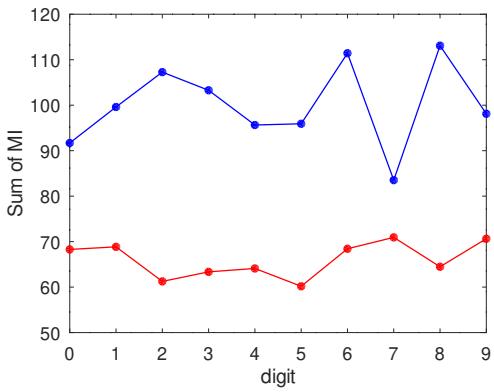

Fig. 5 Sum of MI per digit before applying plasticity (red) and after application (blue)

3.4 Classification of recalled patterns

The AutoGluon classifier had been trained with the neural representations recorded over 300ms of the 50 variants for each digit, presented to the network before plasticity is applied.

The trained classifier was subjected to the recalled representations on the network, after plasticity has been applied (Figure 8).The representations unfold from selected stimulation ($m = 5, 10, 20$) of high-MI neurons. We see that the results are uneven and not directly dependent on the size m of stimulated neurons (Figures 8). The results show, without further analysis, that recall of a complete representation from a few 'concept' neurons is possible and that the recalled representations are similar enough to the original representations to be recognized.

This shows that we can retrieve a version of the original input representation after plasticity learning. Thus there is local storage at a small set of neurons for large patterns of hundreds of neurons, such that stimulation of these neurons allows to 'unfold' and spread activations to retrieve large representations. For each pattern, only a small subset of neurons and synapses are affected by plasticity. As we can see from Fig. 9, there is reduction of activation on all neurons not affected by an input pattern as in Fig. 9**a** (pre-plasticity)

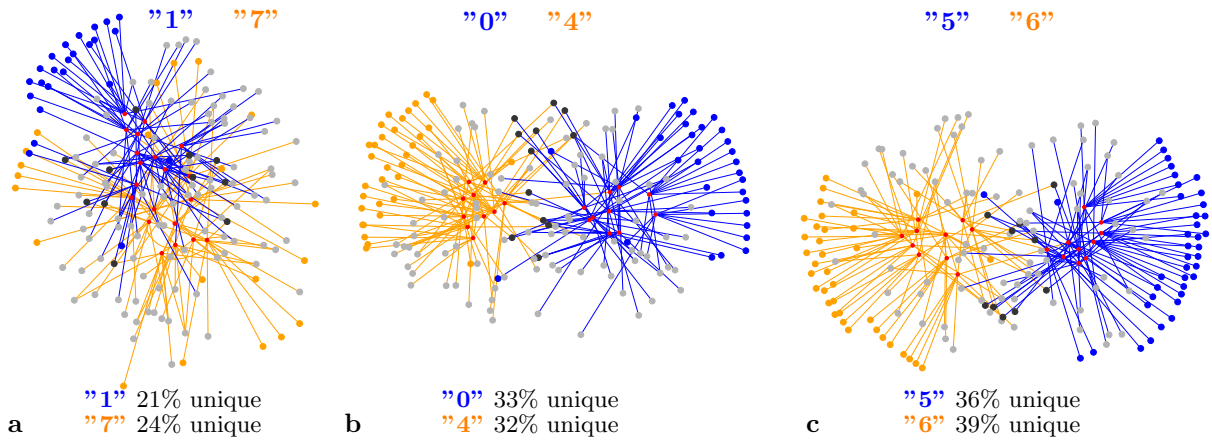

Fig. 6 Network connection graph after plasticity for three examples. Target neurons for the strongest 100 connections from high-MI neurons (red) and their connections are shown. Unique feature neurons are colored, black indicates a shared feature neuron, grey feature neurons overlap with other digits. **a**: fewer unique neurons indicating higher similarity, **b**: more unique, and **c**: most unique features.

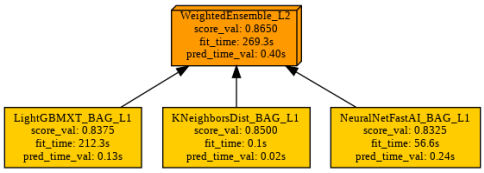

Fig. 7 AutoGluon: combination of classification models in a stack. The "score_val" is the score for the validation data set. The weighted combination (Model Weighted_Ensemble_L2) achieves a higher value by a learnt weighted combination of the predictions of the 3 individual models.

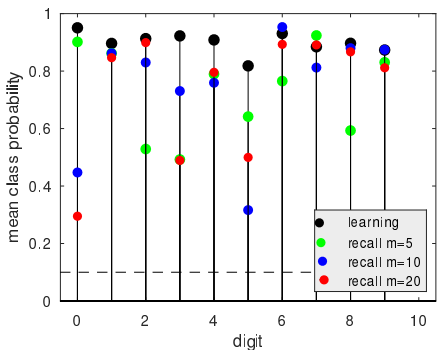

Fig. 8 Pattern recall: classification for each digit pattern (**a**) by supervised learning (black) and (**b**) from recalled representations with $m = 5$ (green), $m = 10$ (blue), and $m = 20$ (red) (testing)

vs. **b** (post-plasticity). These percentages are very stable across patterns.

What the percentages show is that in a naive network there is a large number of neurons with minimal response to patterns. In a trained network most neurons show suppressed activation in response to a pattern. about 20% have no response, and about 10% raise their activation level. The naive state could be said to have a 'reservoir' of unallocated neurons (Jaeger, 2002; Maass et al., 2002). One could also reduce the size of the network after plasticity and have a 'compressed' active representation

3.5 Synthesis

It is remarkable that all pattern memories are stored in parallel, and stimulation of pattern-specific neurons will reproduce the associated distributed pattern with considerable accuracy over the whole network. Learning a pattern with localist adaptation minimally affects the response of the network to other patterns. The network shows considerable independence in its response to input. Our simple learning rule could be further developed.

For instance, the questions of unique features for patterns vs. overlap of features (Fig. 6. colored vs. grey/black target neurons) is important for abstraction and generalization. It is an open questions, which patterms are suitable for a concept-feature 'symbolic' abstraction and whether a learning mechanism is successful which *imposes* a structure onto the material for ease of simple one-shot learning. It is entirely possible that cortical plasticity structures material and is

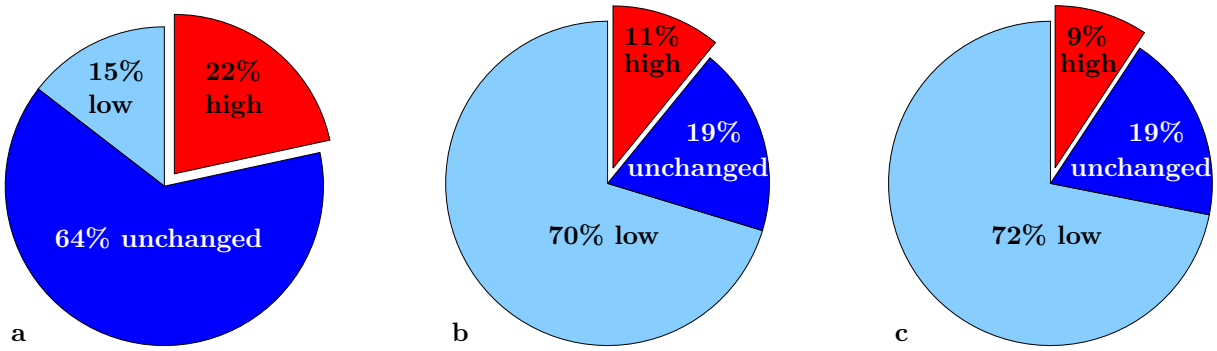

Fig. 9 Exemplification of localist learning (mean of 50 variations for "1"): (**a**) patterns prior to plasticity, (**b**) after plasticity, and (**c**) after targeted stimulation with ($m = 10$ per digit). The percentage of E neurons with no response (**a**) becomes low activity in (**b**, **c**). Number of high-activity neuron (**a**) becomes compressed (**b**, **c**)

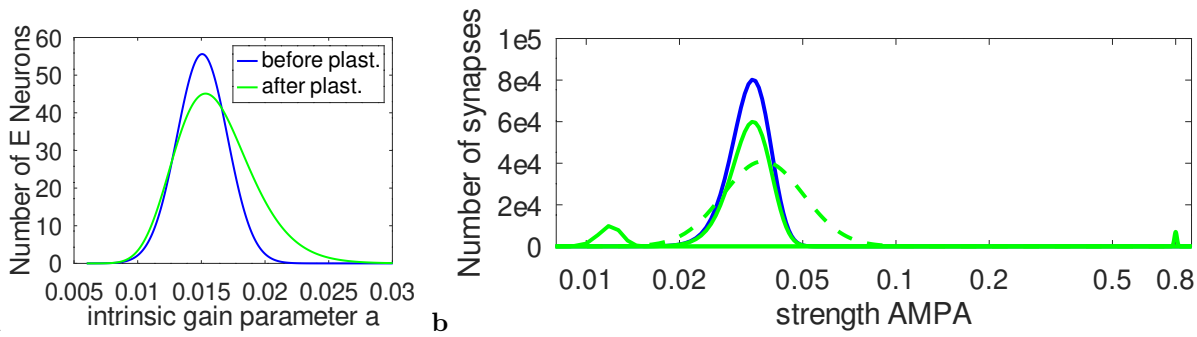

Fig. 10 (**a**) Distribution of the intrinsic gain parameters a and b before applying plasticity (blue), and after plasticity (green). The Gaussian distribution is streteched to a lognormal distribution. (**b**) Distribution of strength of E-E AMPA connections prior to applying plasticity (blue), fitted to a Gaussian. After plasticity, the distribution consists of three different groups (green): strong synapses, which have been strengthened during plasticity, a group of weakened synapses, and a majority of unchanged synapses. We assume that 'housekeeping' by homeostatic plasticity (sleep) would lead to an attested distribution.

complemented by other types of structural learning in subcortical areas. The enormous simplicity of the plasticity rule and the relative success of the given example problem makes sense on the backgroud of an imposed structure learning.

There are implications for biology and technical applications.

- biology: We show that a cortical model is able to form high-level 'concept' neurons where individual pattern information is stored in a localized manner. A version of the pattern information can be retrieved from these neurons, forming a feature ensemble as in Fig. 6. These results fit with recent experimental work (Josselyn & Tonegawa, 2020). It has been empirically shown that up-regulation of intrinsic excitability occurs for those neurons which are part of an ensemble (engram) (Pignatelli et al., 2018).
- technical: Input patterns are learned as high information 'concept' neurons and their connections in a single trial. Over time, a number of different patterns can be stored in this way. Questions of pattern storage and separation, overlap and similarity (generalization, abstraction) by feature neurons are an obvious next step.

3.6 Analysis

The goal of localist plasticity is to allow restoration of representations with high accuracy from activation of high MI neurons alone.

Foremost, we can show that the mutual information values for each pattern as the sum of the MI values over the high-ranking neurons for $m = 10$ increases as a result of plasticity (Fig. 5).

With our model, pattern recall by localist stimulation of such 'concept' neurons can be investigated. To understand the process we look at the neural representation after plasticity has been applied. The feature representations for each concept may be overlapping, and the same features can be re-used in multiple representations (cf. Fig.6).

Nonetheless, the digit patterns retain a significant proportion of features, which are unique to each digit.

Interestingly, our localist plasticity changes the initial Gaussian distribution to a wider lognormal distribution with heavy-tail characteristics (cf. Figure 10), (Scheler, 2017). The overall distribution of intrinsic gain parameters a and b after plasticity approaches a lognormal distribution, where the high-MI neurons are substantially strengthened (Figure 10**a**).

The picture is more complex for synaptic weights, because only few weights are changed by plasticity in our example. For $m = 10$ high MI (=concept) neurons, only ≈ 6400 AMPA connections (out of $\approx 10^6$, that is 0.5%) are adapted. This leads to an uneven distribution of synaptic weights (Figure 10**b**). Extrapolating for larger numbers of weight updates could lead to a biologically attested lognormal distribution.

4 Discussion

We developed a mechanism to store patterns in cortical-like networks using a neuronal ensemble, more precisely a concept-feature ensemble. The plasticity rule stores patterns on the network by localist adaptation of high MI neurons and their synaptic connections, creating a hub-spoke structure. By stimulating only single-pattern high MI 'concept' neurons, the resulting representation unfolds on the network and is similar enough to be recognized by a trained classifier. Activating select neurons by direct stimulation results in a whole set of related neurons to recreate a similar pattern to the original. In our network model individual neurons are recruited as pattern storage elements. We thus achieve a localist memory with a distributed component.

Such a concept-feature ensemble need not be just a passive storage of pattern information. The 'concept' neurons may also act as 'control' neurons, when they are interacting with each other, forming a set of high-level neurons with access to their feature neurons as needed.

It is noticeable that the learning method imposes a structure on the network that is characteristic, could be significant for biological cognition, and well-suited for symbolic computation.

We have not discussed the biological mechanisms which could underlie such selective plasticity, but there are possibilities in the cell-internal memory which filters the information by a number of indicators, such as small molecules and proteins (Scheler, 2023).

Our method of representing pattern information in a network allows for efficient storage of concept-feature ensembles. The possibilities for structure-imposed pattern abstraction learning by cortical networks have only been exemplified here for the first time. We have also not yet analysed the possibilities for blends and interference between patterns which may serve to create higher-order memory.

Declarations

The work for this paper did not receive any funding. The authors have no competing interests to declare that are relevant to the content of this article.

References

Agetsuma, M., Sato, I., Tanaka, Y., Carrillo-Reid, L., Kasai, A., Noritake, A., ... Nagai, T. (2023). Activity-dependent organization of prefrontal hub-networks for associative learning and signal transformation. *Nature Communications*, *14*, 1, https://doi.org/10.1038/s41467-023-41547-5

Carpenter, G.A. (2000). Combining distributed and localist computations in real-time neural networks. *Behavioral and Brain Sciences*, *23*(4), 473–474, https://doi.org/10.1017/S0140525X00283350

Choucry, A., Nomoto, M., Inokuchi, K. (2024, June). Engram mechanisms of memory linking and identity. *Nature Reviews Neuroscience*, *25*(6), 375–392, https://doi.org/10.1038/s41583-024-00814-0

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., Smola, A. (2020a). *AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data.* arXiv preprint arXiv:2003.06505.

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., Smola, A. (2020b). Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505*, ,

Fuster, J.M. (1998). Cellular dynamics of network memory. *Zeitschrift Für Naturforschung C*, *53*(7–8), 670–676, https://doi.org/10.1515/znc-1998-7-819

Gal, E., Amsalem, O., Schindel, A., London, M., Schürmann, F., Markram, H., Segev, I. (2021). The role of hub neurons in modulating cortical dynamics. *Frontiers in Neural Circuits*, *15*, 1, https://doi.org/10.3389/fncir.2021.718270

Gijsbers, P., Bueno, M.L.P., Coors, S., LeDell, E., Poirier, S., Thomas, J., ... Vanschoren, J. (2023). *Amlb: an automl benchmark.* Retrieved from https://arxiv.org/abs/2207.12560

Gu, Y., Qi, Y., Gong, P. (2019). Rich-club connectivity, diverse population coupling, and dynamical activity patterns emerging from local cortical circuits. *PLOS Computational Biology*, *15*(4), 1-34, https://doi.org/10.1371/journal.pcbi.1006902

Izhikevich, E. (2004, 10). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, *15*, 1063 - 1070, https://doi.org/10.1109/TNN.2004.832719

Jaeger, H. (2002). Adaptive nonlinear system identification with echo state networks. *Proc. NIPS* (Vol. 15). MIT Press. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2002/file/426f990b332ef8193a61cc90516c1245-Paper.pdf

Josselyn, S., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science*, *367*, eaaw4325, https://doi.org/10.1126/science.aaw4325

Kim, J., Leahy, W., Shlizerman, E. (2019). Neural interactome: Interactive simulation of a neuronal system. *Frontiers in Computational Neuroscience*, *13*, 1, https://doi.org/10.3389/fncom.2019.00008

Langille, J.J., & Brown, R.E. (2018). The synaptic theory of memory: A historical survey and reconciliation of recent opposition. *Frontiers in Systems Neuroscience*, 1, https://doi.org/10.3389/fnsys.2018.00052

LeCun, Y., & Cortes, C. (2005). *The MNIST database of handwritten digits.* Retrieved from https://api.semanticscholar.org/CorpusID:60282629

Maass, W., Natschläger, T., Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, *14*, 2531-60, https://doi.org/10.1162/089976602760407955

Park, S., Ko, S., Frankland, P., Josselyn, S. (2024). Comparing behaviours induced by natural memory retrieval and optogenetic reactivation of an engram ensemble in mice. *Philosophical Transactions B*, *379*, 20230227, https://doi.org/10.1098/rstb.2023.0227

Pignatelli, M., Ryan, T., Roy, D., Lovett, C., Smith, L., Muralidhar, S., Tonegawa, S. (2018). Engram cell excitability state determines the efficacy of memory retrieval. *Neuron*, *101*, 1, https://doi.org/10.1016/j.neuron.2018.11.029

Sampathkumar, V., Miller-Hansen, N.A., Sherman, S.M., Kasthuri, N. (2021). Integration of signals from different cortical areas in higher order thalamic neurons. *Proceedings of the National Academy of Sciences*, *118*(30), e2104137118, https://doi.org/10.1073/pnas.2104137118

Scheler, G. (2016). Extreme pattern compression in lognormal networks. *F1000Research* ***6****:2177(poster)*, 1, https://doi.org/10.7490/f1000research.1113011.1

Scheler, G. (2017). Logarithmic distributions prove that intrinsic learning is Hebbian [version 2; peer review: 2 approved]. *F1000Research*, *6*, 1222, https://doi.org/10.12688/f1000research.12130.2

Scheler, G. (2018). Neuromodulation influences synchronization and intrinsic read-out. *F1000Research*, 7:1277, https://doi.org/10.12688/f1000research.15804.1

Scheler, G. (2023). *Sketch of a novel approach to a neural model.* arXiv:q-bio.NC. Retrieved from https://arxiv.org/abs/2209.06865

Schneidman, E., Bialek, W., Berry, M.J. (2003). Synergy, redundancy, and independence in population codes. *Journal of Neuroscience*, *23*(37), 11539–11553, https://doi.org/10.1523/JNEUROSCI.23-37-11539.2003

Schumann, M.L. (2025). Enhancing semi-supervised learning with a meta-feature based safeguard system. *Proc. Symposium on Intelligent Data Analysis (IDA).*

Sheintuch, L., Rubin, A., Ziv, Y. (2022). Bias-free estimation of information content in temporally sparse neuronal activity. *Plos Computational Biology*, *18*(2), e1009832, https://doi.org/10.1371/journal.pcbi.1009832

Uytiepo, M., Zhu, Y., Bushong, E., Polli, F., Chou, K., Zhao, E., . . . Maximov, A. (2024). Synaptic architecture of a memory engram in the mouse hippocampus. *bioRxiv*, 1, https://doi.org/10.1101/2024.04.23.590812

Zeng, G. (2015). A unified definition of mutual information with applications in machine learning. *Mathematical Problems in Engineering*, 1, https://doi.org/10.1155/2015/201874