

Constraining the Connectivity of Sparse Neural Associative Memories

Philippe Tigréat, Vincent Gripon, Pierre-Henri Horrein

Electronics Department, Telecom Bretagne

Brest, France

Email: {philippe.tigreat, vincent.gripon, ph.horrein}@telecom-bretagne.eu

Abstract—Willshaw networks are a type of associative memories with a storing mechanism characterized by a strong redundancy. Namely, all the subparts of a message get connected to one another. We introduce an additional specificity, by imposing the constraint of a minimal space separating every two elements of a message. This approach results from biological observations, knowing that in some brain regions, a neuron receiving a stronger stimulation can inhibit its neighbors within a given radius. Theoretical arguments are derived to quantify the benefits of this method in terms of memory usage as well as pattern completion ability. We experiment with different values of the inhibition radius introduced, and we study its impact on the error rate in the retrieval of stored messages. We show that this added constraint can result in significantly better performance of the Willshaw network, either when reducing its set of connections, or when extending its set of neurons while maintaining the memory resource.

Keywords—Willshaw Networks; Clique-Based Neural Networks; Content-Addressable Memory; Lateral Inhibition; Sparse neural networks.

I. INTRODUCTION

Associative memories are a type of computer memories that are part of the broader category of content-addressable memories. A new model of associative memory based on neural networks was formerly introduced in [1]. Where addressable memories associate an address with a piece of data, associative memories have the characteristic of associating patterns to one another. Among this group, we distinguish between hetero-associative memories, and auto-associative memories. An hetero-associative memory will associate together patterns in pairs. For instance, if the pattern p_1 was associated with pattern p_2 , the request p_1 will bring the response p_2 . Auto-associative memories follow a different principle, as they will associate a pattern with itself. The main use case of these memories is pattern completion, where a request made of a subpart of a stored message will get as response the completed pattern. Associative memories can be found in several types of real-world applications, such as database engines [2], network routers [3], data compression devices [4], and computer vision systems [5].

Today, it is widely accepted that the working principle of the brain can often be likened to the operation of an associative memory. The focus is put here on the phenomenon observed in biological neural networks, called lateral inhibition [6]. It can also be referred to as surround suppression [7]. This translates in the inhibition exerted by some neurons on their close neighbors when these have an activity inferior to their own. Starting from the Willshaw model [8], we propose a neural associative memory that is improved in terms of plausibility, by the introduction of local inhibition that results in the prohibition of short-range connections. We show that

this modification brings a performance improvement in the retrieval of stored messages.

Section II introduces three associative memory models with relevant relationship to this work. Section III gives a formal presentation of Willshaw networks, including the usual message retrieval algorithm, and biological considerations motivating the modifications we introduce. Section IV details modifications in our implementation as compared to the classic Willshaw model, including the constraint applied on the space between connected neurons. Section V provides theoretical arguments showing the better usage of memory brought by this constraint. Section VI presents the results we obtain in pattern completion, and gives some theoretical explanations.

II. RELATED WORK

A. Hopfield Networks

The prominent model for associative memories was introduced by John Hopfield [9], [10]. Hopfield networks are made of a set of N neurons that are fully interconnected. The training of these networks, given n binary vectors x^μ of length N , consists in modifying the weight matrix W according to the formula:

$$w_{ij} = \frac{1}{n} \sum_{\mu=1}^n x_i^\mu x_j^\mu, \quad (1)$$

where element w_{ij} at the crossing between line i and column j of W is the real-valued connection weight from neuron i to neuron j .

As connections are reciprocal and not oriented, we have:

$$w_{ij} = w_{ji} \quad \forall i, j \in \llbracket 1, N \rrbracket \quad (2)$$

for any indices i and j in the list of neurons, which makes W symmetrical.

The binary values considered for the stored messages are usually -1 and 1, but can be adapted to work with other binary alphabets. The Hopfield model has a limited efficiency, in particular it does not allow a storage of more than $0.14N$ messages [11]. The limits of the model can be explained by the facts that each entry of the matrix is modified at every time step of the storing procedure, and that the changes are made in both directions and can, therefore, cancel each other out. This overfitted characteristics of associative memories is very different from that observed in learning applications. Indeed, an overfitted learning system recognizes only the training samples and fails at generalizing to novel inputs. To the contrary, an overfitted storing system recognizes everything and does not discriminate anymore between stored and nonstored data.

B. Willshaw Networks

Willshaw networks are another model of associative memories in which information is carried by the existence or absence of connections [8], [12]. Its material is made of a set of N neurons and N^2 potential connections between them. A message is then a fixed size subset of the N neurons, and can be represented by a sparse vector of length N with ones at these neurons' positions and zeros everywhere else. The connection weights are binary, and the active units in a message get fully interconnected as soon as it is memorized, thus forming a clique. Figure 1 gives an example of such a network. The performances of Willshaw networks are way superior to those of Hopfield memories, given that stored messages are sparse (i.e., they contain a small proportion of nonzero elements). Further theoretical and numerical comparison between Hopfield and Willshaw networks can be found in [13]–[16].

C. Clustered Cliques Networks

Recently, a novel type of associative memories was proposed by Gripon et al., called Gripon-Berrou Neural Networks (GBNNs) or clustered cliques networks (CCNs) [17], [18]. These associative memories make use of powerful yet simple error correcting codes. These networks consider input messages to be nonbinary, and more precisely to be words in a finite alphabet of size l . This specific structure allows the separation of nodes into different clusters, each being constituted of the same number l of nodes. Connections between nodes inside a given cluster are forbidden, only the connections between nodes in two different clusters are allowed. There again, this model brings a significantly improved performance as compared to the former state-of-the-art of associative memories, namely Willshaw networks [19]–[21]. For instance, it can be found experimentally that with 2,048 nodes and 10,000 stored messages of order 4 and 2-erasures queries, a Willshaw network will have an error rate close to 80%, while a clique-based neural network will only make 20% of wrong retrievals.

In both Hopfield and Willshaw models, the number of messages the network can store and retrieve successfully is linearly proportional to the number of nodes, with a greater proportionality constant for Willshaw networks [14]. In clique-based neural networks however, storage capacity grows quadratically as a function of the number of units.

One of the objectives of the present work is to explain the performance improvement brought by the separation of the network into clusters. Therefore, we study a network that can be considered as an intermediate between the Willshaw and Gripon-Berrou models. More precisely, our proposed model adds a locally exclusive rule for nodes to be active in the network.

III. WILLSHAW NETWORKS AND BIOLOGICAL CONSIDERATIONS

Willshaw networks are models of associative memories constituted of a given number of neurons. A stored message, or memory, is a combination of nodes taken in this set.

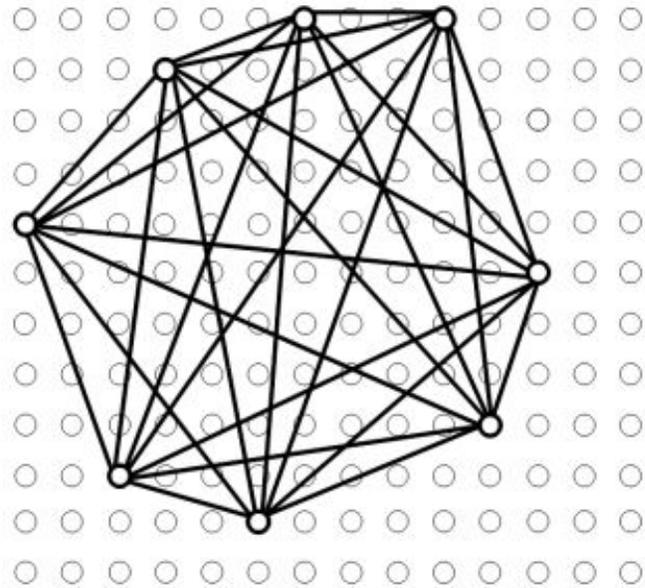


Figure 1. Willshaw network. A message composed of 8 nodes is displayed, the inter-connections being the means of its storage in the network.

The storage of this information element corresponds to the creation of connections with unitary weights between every two neurons in this message. The graphical pattern thus formed is termed "clique". The storing process of n binary vectors x^μ of length N , is equivalent to the modification of elements of the network's connection matrix W , according to the formula:

$$w_{ij} = \max_{\mu} x_i^{\mu} x_j^{\mu} \quad (3)$$

Note that here, the max operator is performed coefficient-wise. Equivalently, the connection weight between nodes i and j is equal to 1 if, and only if, those two nodes are both part of one of the n stored messages.

The network's density d is defined as the expected ratio of the number of ones in the matrix W to the number of ones it would contain if every possible message was stored. For cliques of order c , the number of connections they contain is $\binom{c}{2}$. Despite the correlation between these edges, the probability of a given connection to be picked when forming a message can be estimated as $\frac{\binom{c}{2}}{\binom{N}{2}}$. Provided that the n stored messages are uniformly distributed and independent, the density of the network equates to the probability for any given connection to belong to at least one of these messages. This leads to the formula:

$$d = 1 - \left(1 - \frac{\binom{c}{2}}{\binom{N}{2}} \right)^n \quad (4)$$

The efficiency of a connectionist associative memory is defined as the ratio of the maximal amount of information carried by the messages it is capable of storing then retrieving with high probability, over the total information quantity represented by its set of connection weights. For a Willshaw

network with N nodes, the number of potential connections, or binary resource, is

$$Q = \frac{N(N-1)}{2} [\text{bits}]. \quad (5)$$

After M messages have been stored in the network, the amount of information it contains is

$$B = M \left(\log_2 \left(\binom{N}{c} \right) \right) [\text{bits}]. \quad (6)$$

Hence the efficiency of a Willshaw network is

$$\eta = \frac{2M \left(\log_2 \left(\binom{N}{c} \right) \right)}{N(N-1)}. \quad (7)$$

The maximal attainable efficiency is $\ln(2)$ [22].

The stimulation of a Willshaw network with an input request can be performed as the product of the sparse input vector by the network's connection matrix. The resulting vector then contains the output scores of the network's neurons. The score of a neuron is thus the sum of unitary stimulations it receives from the request elements it is connected to. Neurons must then be selected based on their score. Algorithm 1 defines a procedure that can be used for the recovery of a complete message from a subpart of its content. The Global Winner-Takes-All step consists in discarding all active neurons with a score below the maximum.

Algorithm 1: Message retrieval procedure in a classic Willshaw network.

Data: Subpart x of a stored message

Result: Set of nodes z active after treatment

$z = x$

Repeat

$y = Wz$

$z = \text{GlobalWinnerTakesAll}(y)$

while (convergence not reached

and max. nb. of iterations not reached)

Return z

The probability of error in the retrieval of a message from the Willshaw network can be calculated when the process uses a single iteration. If only one vertex of the clique to complete has been erased, the probability of completing the message accurately after one iteration is the probability that no unit is connected to all elements of the query, other than the missing one:

$$P_{\text{retrieve}} = (1 - d^{c-1})^{N-c}. \quad (8)$$

The probability of error is then:

$$P_e = 1 - P_{\text{retrieve}} = 1 - (1 - d^{c-1})^{N-c}. \quad (9)$$

Knowing (4), this gives:

$$P_e = 1 - \left(1 - \left[1 - \left(1 - \frac{\binom{c}{2}}{\binom{N}{2}} \right)^n \right]^{c-1} \right)^{N-c}. \quad (10)$$

This holds for numbers of erasures c_e superior to 1, bringing:

$$P_e = 1 - \left(1 - \left[1 - \left(1 - \frac{\binom{c}{2}}{\binom{N}{2}} \right)^n \right]^{c-c_e} \right)^{N-c}. \quad (11)$$

We aim to modify classic Willshaw networks in a way that is relevant in regard to biological observations. Emphasis is put on lateral inhibition, a phenomenon that has been observed in several areas of the brain. It is notably present in sensory channels. For vision, it operates at the level of retinal cells and allows an increase in contrast and sharpness of signals relayed to the upper parts of the visual cortex [7] [23]. In the primary somatosensory area of the parietal cortex, neurons receive influx coming from overlapping receptive fields. The Winner-Takes-All operation resulting from the action of inhibitory lateral connections allows localizing precisely tactile stimuli, despite the redundancy present in the received information [24]. The same scheme of redundancy among sensory channels, and filtering via lateral inhibition, is present in the auditory system [6]. WTA is observed in the inferior colliculus and in upper levels of the auditory processing channel.

IV. PREVENTING CONNECTIONS BETWEEN NEIGHBOR NEURONS

Classic Willshaw networks have no topology. Their material is constituted with a list of neurons each having an index as sole referent. There is neither a notion of spatial position in these networks, nor, a fortiori, of spatial distance. We get closer here to a real neural network, by arranging them on a two-dimensional map. In the model we propose, the respective positions of two neurons impact the possibility for them to get connected together. The considered network is composed of a number N of nodes evenly distributed along a square grid, of side $S = \sqrt{N}$. Stored messages are of constant order, meaning they are all constituted of the same number of neurons. We forbid connections between nearby neurons. To this end, we apply a threshold σ on the spatial length of a connection. Stored messages must necessarily be conform to this constraint. Each message is formed in a random manner, units are chosen iteratively. Each new element of the message is picked from the positions left available after the removal of the neighbors of the formerly selected nodes, as indicated in Figure 2. One can consider the introduced constraint as applied on the network's material, as the weights of a predetermined set of short-range connections will be enforced to stay null all along the network's life. During the formation of a message, it is practical to pick neurons to satisfy this constraint in a sequential manner, with a local inhibition applied on a neuron's neighborhood from the moment it is selected until the message generation is complete.

A link can be drawn between this approach and Kohonen Self-Organizing Maps, where close-by neurons encode more similar information [25]. Therefore, long-range distance separates information elements that are different in nature, whereas

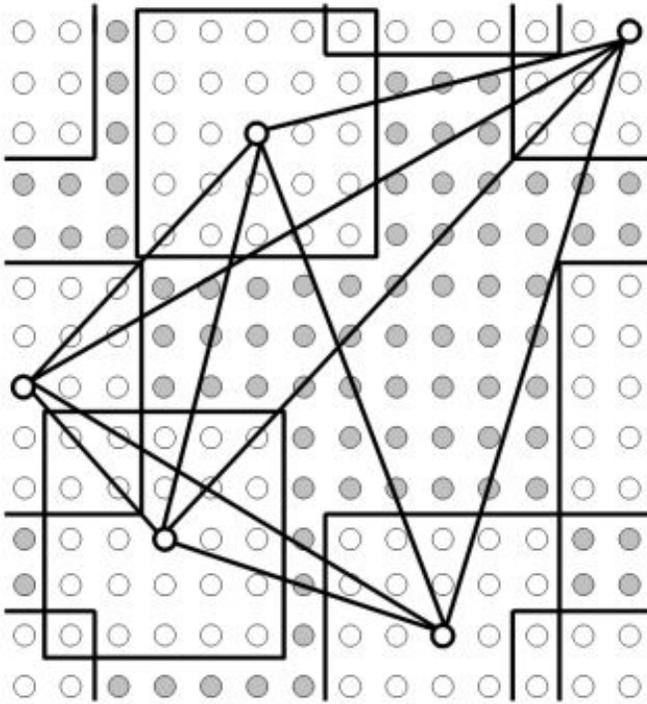


Figure 2. Cyclic Willshaw network with a constraint on local connections.

shorter-range distance depicts a difference in degree. Local competition is particularly relevant in this scheme.

During retrieval, the network is stimulated iteratively with a request that will most often change from one iteration to the next. Each node of the request will first stimulate every other element it is connected to. Scores are initialized with zero at the start of every iteration, and each stimulation is a unitary increment to the score of the receiver unit. For the first iteration, after the stimulation we apply a global Winner-Takes-All rule, which consists in excluding from the research scope all units that do not achieve the maximal score observed in the network. We know indeed that the neurons from the searched message will all have the maximum possible score, equal to the number of elements in the request. Once non-maximum elements are put to zero, we only pay interest in the remaining neurons during the rest of the retrieval process. Moreover, for every iteration after the first one, neurons in the new request are the only ones that can receive stimulation as the algorithm proceeds to only discard neurons from then on.

Thereafter, we can keep using the global WTA principle iteratively, but other algorithms such as Global Winners-Take-All (GWsTA) or Global Losers-Kicked-Out (GLsKO) [26] are more efficient in discriminating the right nodes from the spurious ones that can appear during retrieval.

GWsTA relies on the calculation of a threshold score to select winner neurons. This threshold is chosen such that neurons with an activity above it are in number at least as large as the order of stored messages.

GLsKO consists in putting off, at each iteration, all the units that do not have the highest score, or a subgroup sampled

randomly in this ensemble.

These two algorithmic techniques allow to getting rid of an important proportion of false-positives. In the clique-based CCN, clusters play a similar role.

The iterative nature of the process means that a message retrieved as output from the network is typically reinjected in it until input and output no longer differ. A limited number of iterations is applied in the case where the network would not converge to a stable solution, an observable case in which it can oscillate between two states.

In addition to these two stopping criteria that are the maximum number of iterations and convergence, comes a third one which is the identification of a clique. Indeed, if we observe that the units still active after an iteration are in number equal to the order of stored messages, and that they all have the same score, this means it is a stored message. This ensemble is then retained as the response given by the network for the current request.

Algorithm 2 shows the message retrieval procedure used in the results we present. Phase II uses GLsKO.

Algorithm 2: Message retrieval procedure in the modified Willshaw network with spacing constraint.

Data: Subpart x of a stored message

Result: Set of nodes z active after treatment

Phase I

$$y = Wx$$

$$z = \text{GlobalWinnerTakesAll}(y)$$

Phase II

Repeat

$$y = Wz$$

$$a = \text{active nodes in } y$$

$$m = \text{nodes in } a \text{ with minimal score}$$

$$z = a - m$$

while (convergence not reached

and max. nb. of iterations not reached)

Return z

We experiment the storage of messages of order c in the connection matrix of the network. Messages are formed with the constraint of a minimal space between connected nodes. Two units in a message must be spaced apart at a distance superior to a minimum σ . In order to ease computations and avoid edge effects, we choose to use the L_1 distance, even though we believe this method should work using any distance. This way, when picking a node x for a message, all nodes located in a square grid centered on x , of side $2\sigma+1$, are excluded from the possible choices for the elements of the message remaining to be filled. Moreover, this distance is applied in a cyclic way, meaning a node located on the right edge of the grid will be considered a direct neighbor of the element located at the crossing between the same line and the left edge of the grid. All four corners of the grid will also be neighbors to one another. We call the network so described a torus.

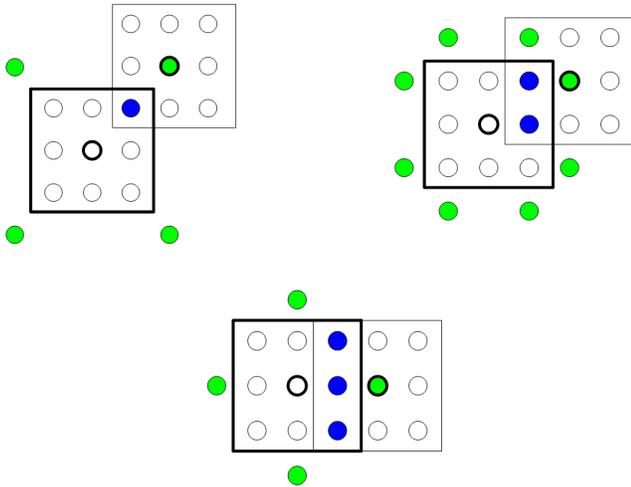


Figure 3. Different overlapping configurations between the inhibition areas around two neurons. Here with $\sigma = 1$, the union of the two overlapping areas can contain 17, 16, or 15 neurons.

V. EFFICIENT USE OF THE MEMORY RESOURCE

When applying a constraint σ on the minimal spacing between connected neurons, the number of potential connections in the modified Willshaw network becomes

$$Q = \frac{N(N - (2\sigma + 1)^2)}{2} [bits]. \tag{12}$$

Let the total number of messages one can form in it under the spatial constraint, be denoted \bar{M} . The entropy per message b is given by

$$b = (\log_2(\bar{M})) [bits]. \tag{13}$$

The amount of information contained in the network after the storage of M messages is then

$$B = bM = M (\log_2(\bar{M})) [bits]. \tag{14}$$

Hence the efficiency of the network with lateral inhibition is

$$\eta = \frac{2M (\log_2(\bar{M}))}{N(N - (2\sigma + 1)^2)}. \tag{15}$$

Predicting the diversity of the spatially constrained Willshaw network given N , c and σ , is not trivial for most values of these parameters. Indeed, once one neuron has been picked among N , the remaining choice for the second element of a message is naturally $N - (2\sigma + 1)^2$. However, there are then several possibilities for the number of remaining allowed components, as the inhibition areas around the first two elements can overlap. Figure 3 shows this behavior with $\sigma = 1$. In this case, two overlapping inhibition areas can either share one, two or three neurons. Given a fixed position for the first neuron, there are four positions for the second one that will give an intersection of one, eight that will give an intersection of two, and four positions will give intersections of three neurons.

For $N - (4\sigma + 1)^2$ neurons, the two inhibition areas do not overlap, and there are $N - 2(2\sigma + 1)^2$ remaining possible

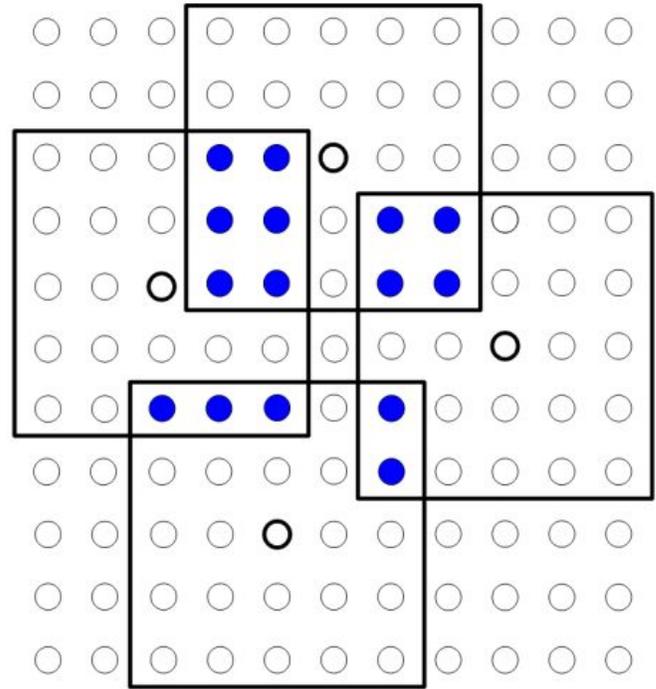


Figure 4. Overlapping inhibition areas for a combination of 4 selected neurons satisfying the spatial constraint $\sigma = 2$. Intersections are made of 2, 3, 4 and 6 neurons.

choices for the third node.

The possible number of ordered arrangements of three nodes respecting the spacing constraint $\sigma = 1$ is thus

$$|\mathcal{A}_{\sigma=1}(N, 3)| = N((N - 25)(N - 18) + 4(N - 17) + 8(N - 16) + 4(N - 15)). \tag{16}$$

Finally, as the considered messages are unordered, we divide this formula by $c!$, here six, to get the number of allowed messages:

$$\bar{M} = \frac{N^3 - 27N^2 + 194N}{6}. \tag{17}$$

Similarly, for other sets of parameters, the diversity of the spatially constrained network will be obtained by sum-product compositions [27] of polynomials with integer roots. As c grows, the number of overlapping configurations between inhibition areas for neurons in a message is multiplied, and finding a formula to predict the diversity becomes increasingly complex.

Figure 4 shows a possible configuration for the intersections of the inhibition areas of four neurons that can be part of the same message, when $\sigma = 2$. This illustrates the variety of intersection configurations that arise as c and σ grow larger.

For some values of σ and c , the total number of legitimate messages can be predicted by polynomial formulas that can be easy to find experimentally. Table I shows a list of such formulas, for the diversity of the constrained network as well as its number of prohibited messages, for different values of c and σ . $\sigma = 0$ corresponds to the unconstrained Willshaw

Table I:

POLYNOMIAL FORMULAS FOR THE NUMBERS OF ALLOWED AND FORBIDDEN MESSAGES UNDER THE SPATIALITY CONSTRAINT ON CONNECTIONS, AS A FUNCTION OF THE NUMBER N OF NEURONS, FOR DIFFERENT CLIQUE ORDERS AND INHIBITION RADII. CASES WITH $\sigma = 0$ ARE EQUIVALENT TO WILLSHAW NETWORKS. THE VALIDITY OF THESE FORMULAS HOLDS FOR NETWORK SIZES SUFFICIENTLY LARGE BEFORE σ .

c	σ	Allowed messages	Forbidden messages
2	0	$\frac{N^2 - N}{2}$	0
	1	$\frac{N^2 - 9N}{2}$	$4N$
	2	$\frac{N^2 - 25N}{2}$	$12N$
	3	$\frac{N^2 - 49N}{2}$	$24N$
3	0	$\frac{N^3 - 3N^2 + 2N}{6}$	0
	1	$\frac{N^3 - 27N^2 + 194N}{6}$	$4N^2 - 32N$
	2	$\frac{N^3 - 75N^2 + 1514N}{6}$	$12N^2 - 252N$
	3	$\frac{N^3 - 147N^2 + 5834N}{6}$	$24N^2 - 972N$
4	0	$\frac{N^4 - 6N^3 + 11N^2 - 6N}{24}$	0
	1	$\frac{N^4 - 54N^3 + 1019N^2 - 6798N}{24}$	$2N^3 - 42N^2 + 283N$
	2	$\frac{N^4 - 150N^3 + 7931N^2 - 149550N}{24}$	$6N^3 - 330N^2 + 6231N$
	3	$\frac{N^4 - 294N^3 + 30539N^2 - 1133958N}{24}$	$12N^3 - 1272N^2 + 47248N$

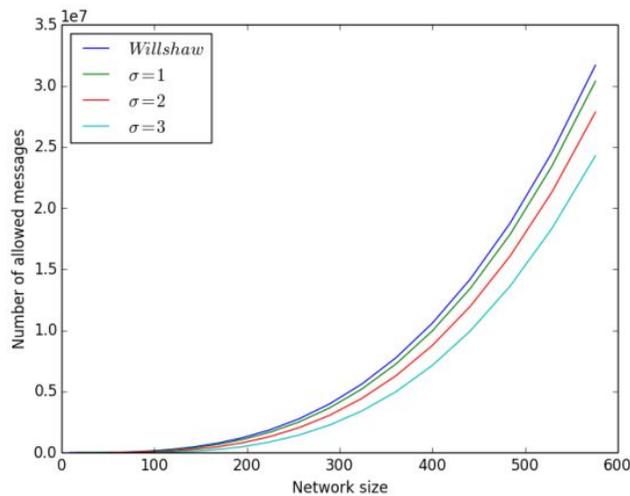


Figure 5. Number of allowed messages as a function of the total number of neurons, under constraints $\sigma = 1, 2,$ and 3 compared with a Willshaw network, for $c = 3$.

network. Figure 5 shows the evolution of the number of legitimate messages for different constraints on connection length, for $c = 3$.

With $c = 3$ and $\sigma = 1$, the network's efficiency is given by

$$\eta = \frac{2M \left(\log_2 \left(\frac{N^3 - 27N^2 + 194N}{6} \right) \right)}{N(N - 9)}, \quad (18)$$

while the efficiency of the corresponding Willshaw network is

$$\eta = \frac{2M \left(\log_2 \left(\frac{N^3 - 3N^2 + 2N}{6} \right) \right)}{N(N - 1)}. \quad (19)$$

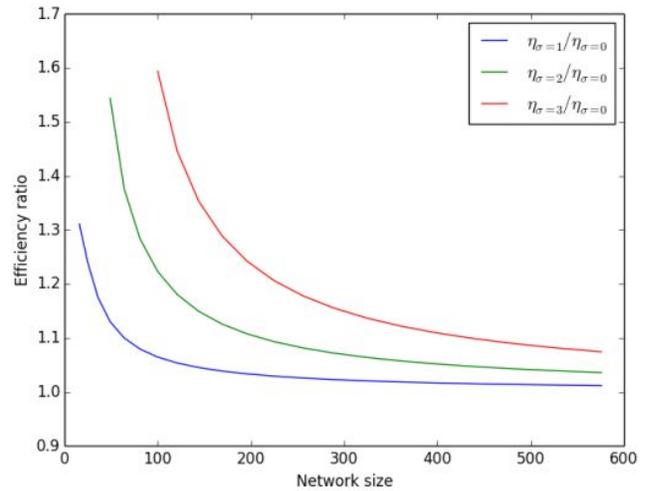


Figure 6. Evolution of the ratios of the efficiency of the networks with spatial constraints $\sigma = 1, 2,$ and 3 over the efficiency of a Willshaw network, with increasing network size, for $c = 3$.

The ratio of these two efficiencies is thus given by

$$\frac{\eta_{\sigma=1}}{\eta_{\sigma=0}} = \frac{(N - 1) \left(\log_2 \left(\frac{N^3 - 27N^2 + 194N}{6} \right) \right)}{(N - 9) \left(\log_2 \left(\frac{N^3 - 3N^2 + 2N}{6} \right) \right)}, \quad (20)$$

and is independent of the number M of stored messages. Figure 6 displays the evolution of the efficiency ratios between spatially constrained networks with $\sigma = 1, 2, 3,$ and the unconstrained Willshaw network, as a function of the number of neurons in the network, for messages made of three neurons. It shows that, although the spacing constraint can not be applied on too small networks, for sufficiently large networks, prohibiting shorter connections comes with an increase in efficiency, as the ratio is superior to one. Larger values of σ are associated with larger gain in efficiency, and for a given σ the best improvement over a Willshaw network comes for the smaller networks where this constraint applies, i.e., where it does not block the vast majority of Willshaw messages. The decay of the ratio when the network size increases is due to the fact that the constraint then prohibits a smaller proportion of connections, making the difference with a Willshaw network less noticeable. Figure 7 shows the evolution of the same efficiency ratios for cliques of order 4. The improvement is then slightly lower for constant σ and network size, in comparison to the case $c = 3$. The ratio remains superior to one however. This tends to show that the spacing constraint can be more beneficial, in terms of gained efficiency, for shorter messages.

The efficiency is a measure of the amount of information one can store for a given amount of available memory. The improvement in efficiency brought by the spacing constraint on connections means that the reduction of used material is more significant than that of the quantity of information carried by messages.

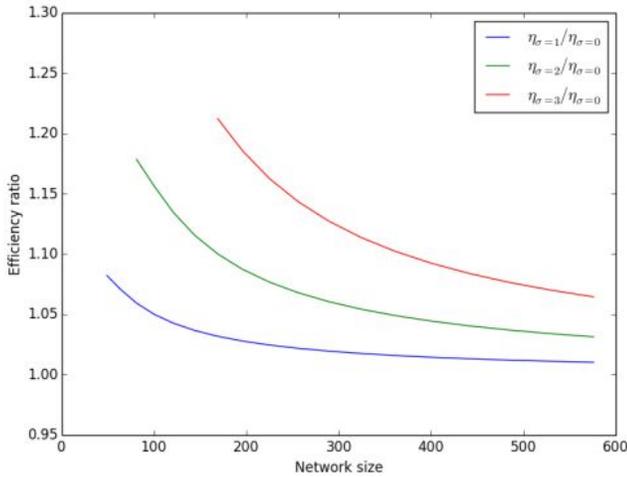


Figure 7. Evolution of the efficiency ratios for $\sigma = 1, 2,$ and 3 over the Willshaw efficiency as a function of the number of neurons in the network, for $c = 4$.

VI. PATTERN COMPLETION

We now pay interest specifically in the pattern retrieval ability of the modified Willshaw network, as compared to the classical model. During retrieval, only a sample from the nodes of the complete message are stimulated, the inputs are subparts of stored messages. Units that are close to elements of an input will not reach the maximum score in the network, and will therefore be ruled out after the first WTA operation. During the second phase of the algorithm, nodes in the vicinity of input neurons will also be more likely to reach a low score if they are activated, and to be discarded. Hence, the local inhibition used initially during the creation of messages impacts the retrieval process as well.

We pay interest in the network's ability to return the exact memory associated with a request. Hence every difference, even marked by a single unit, between the expected pattern and the network's output is counted as an error.

We measure the performance of the network as the ratio of the number of successfully retrieved messages over the total number of requests.

Various parameters can impact this performance, albeit to different degrees:

- the length S of the grid's side
- the number M of stored messages
- the minimal space σ between two elements of a message
- the order c of stored messages
- the number of erasures c_e applied on stored messages to obtain the corresponding request messages

The behavior of this network is interesting in relation to Willshaw networks and clustered cliques networks, as it is close to a classic Willshaw network and displays the added feature of prohibited connections as observed in CCNs. This modification can be viewed as a form of sliding-window clustering.

Figures 8 and 9 represent the matrices of allowed and forbidden connections in a clustered clique network with

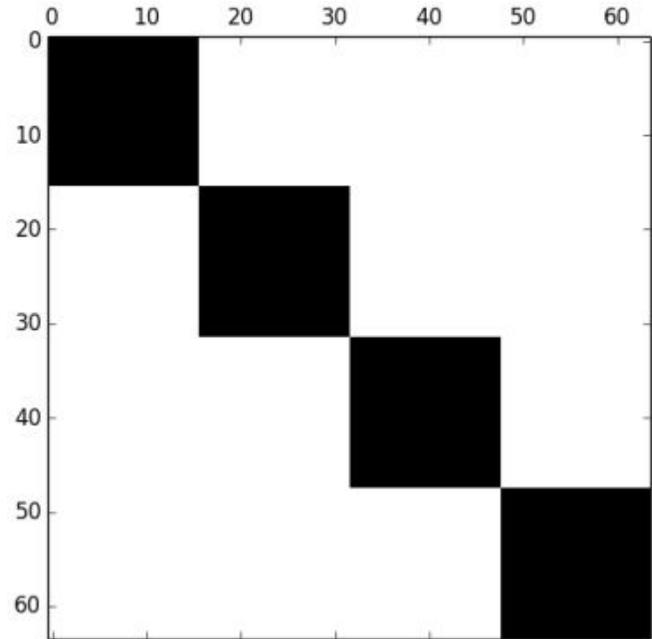


Figure 8. Matrix of the potential and forbidden connections in a clustered clique network with 4 clusters of 16 fanals each. The element at the crossing of a line i and a column j represents the connection between neurons n_i and n_j . White cells represent allowed connections, black cells correspond to forbidden ones.

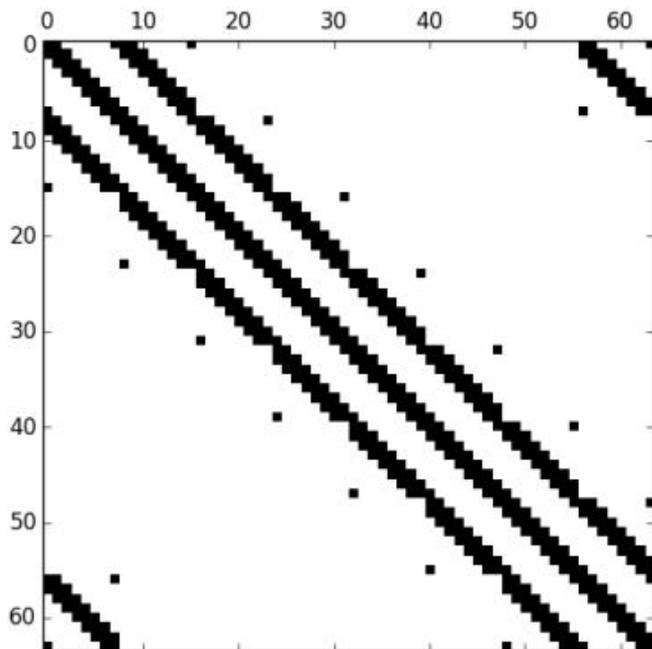


Figure 9. Connection matrix for a modified Willshaw network with side length $S = 8$ and $\sigma = 1$.

$\chi = 4$ clusters comprising $l = 16$ neurons each, and in a modified Willshaw network of side length $S = 8$ under constraint $\sigma = 1$, respectively. The two networks have the same number of neurons, and comparable numbers of allowed and forbidden connections. Indeed, the number of potential non-oriented connections in the clustered clique network is given by

$$Q = \frac{\chi(\chi - 1)l^2}{2} [\text{bits}], \quad (21)$$

which here gives 1,536 allowed connections, against 480 ones forbidden due to the clustering constraint. On the other hand, from (12) we have 1,760 allowed connections in the spatially constrained Willshaw networks, and 256 forbidden ones. The major difference is the potential overlapping between the inhibition areas around different neurons in the modified Willshaw network. It follows that in this network, two neurons n_i and n_j can be both prohibited from connecting to a third one n_k , and yet be allowed to connect together. To the contrary, if in a clustered clique network, connection weights w_{ik} and w_{jk} are forced to remain at 0, then necessarily w_{ij} will be as well.

In a first series of experiments, we focus on the ability of the network to first store independent, identically distributed messages, and then complete them properly when probed with partial cues. For every configuration of the network, messages and requests we test, we store a set of thousands of messages in the network. These messages are generated randomly following the local inhibition pattern described in section IV. We then request it with the full set of queries associated with stored messages.

For each network size, we observe that there is an optimal value of the minimal distance σ , that lowers the most significantly the error rate, as compared to the corresponding Willshaw network without constraint on local connections. For a given minimal distance, the reduction in error rate depends on the number of stored messages, with an optimal number of messages which is a function of the network size. For cliques of order four and with two erasures, the maximal reachable improvement is close to 15%, and seems to be the same for all network sizes. In this configuration, the minimal distance bringing the best performance is approximately the third of the network side.

The evolution of the retrieval error rate as a function of the number of stored messages is slower with the appropriate constraint on connections than for a classic Willshaw network, as can be seen in Figure 10.

Figure 11 shows a similar comparison, this time between the modified Willshaw network with constraint $\sigma = 5$ made of 400 neurons, and an unconstrained Willshaw network with 335 neurons. Because of the reduction in the number of connections when $\sigma = 5$, the two networks have almost the same binary resource. Indeed, the Willshaw network has 55,945 connections while the constrained one has 55,800 possible connections, despite having more neurons. Even though the modified network has a slightly lower footprint, the improvement is even more noticeable than for the comparison with equal size of the neurons sets. In fact for 1,500 stored

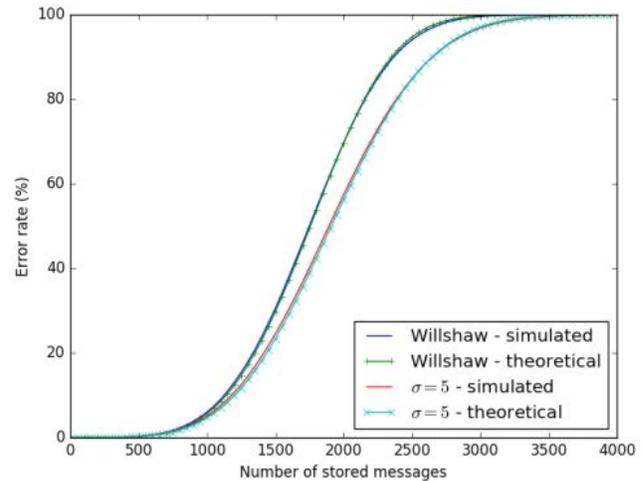


Figure 10. Evolution of the retrieval error rate with and without constraint $\sigma = 5$ in a network of side length 20 with 400 neurons, stored messages of order 6 and 1 erasure applied to form corresponding requests, with 1 iteration.

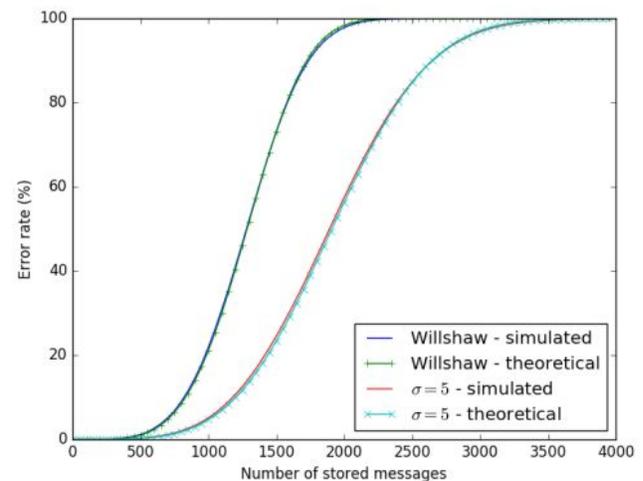


Figure 11. Evolution of the retrieval error rate after one iteration as a function of the number of stored messages, in a classic Willshaw network with 335 neurons, and in a modified Willshaw network of side length 20 with 400 neurons, with constraint $\sigma = 5$. Stored messages have order $c = 6$ and associated queries are obtained by erasing one vertex. The two networks have close numbers of possible connections.

messages, the modified network gains around 50% in error rate over the Willshaw network. Like the CCNs approach, this shows the interest of spreading the binary resource across a larger set of units with constrained connectivity, as opposed to allowing any two neurons to link.

For a constant number of stored messages, the graph of the error rate as a function of σ is characterized by a progressive decay down to a minimum, followed by a rapid growth for upper values of σ , as shown in Figure 12.

This can be explained by two phenomena. On the one hand, the prohibition of a growing part of the possible connections gradually decreases the probability of a "false message", characterized by the intrusion of a spurious node in the output. The existence of a node that is connected to all elements in

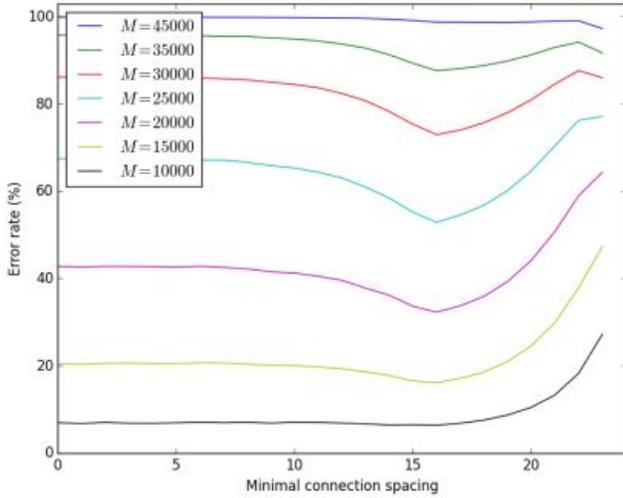


Figure 12. Minimal connection distance effect on performance in a modified Willshaw network with local inhibition, made of 2500 neurons. Stored messages are of order $c = 4$, and $c_e = 2$ erasures are applied to form corresponding requests. Seven different numbers of stored messages are tested. The case where minimal spacing $\sigma = 0$ corresponds to a classic Willshaw network.

a request, yet is not part of the corresponding message, will potentially cause an error. In fact, forbidding some connections has the effect of reducing the number of concurrent nodes susceptible to cause errors. We can estimate the mean number of concurrent nodes remaining after the choice of k neurons of a message:

$$N_c(\sigma, k) = N \left(1 - \left(\frac{(2\sigma + 1)^2}{N} \right)^k \right). \quad (22)$$

The corresponding number of nodes blocked by the constraint on connections is, on average:

$$N_b(\sigma, k) = N \left(1 - \left(1 - \left(\frac{(2\sigma + 1)^2}{N} \right)^k \right) \right). \quad (23)$$

This explains the decay phase in error rate observed for the first values of σ . Let us note that it comes with a decrease in the diversity of messages, namely the total number of different messages that can be stored in the network. Following this decay, the decrease in the number of concurrent nodes has another effect: the reuse of connections by different messages becomes more frequent as the choice for possible connections gets reduced. This comes to counteract the former phenomenon and raises the error rate.

The density of the modified network after the storage of n messages can be calculated by

$$d = 1 - \left(1 - \left(\frac{c}{Q} \right)^n \right), \quad (24)$$

that is

$$d = 1 - \left(1 - \frac{c(c-1)}{N(N-(2\sigma+1)^2)} \right)^n. \quad (25)$$

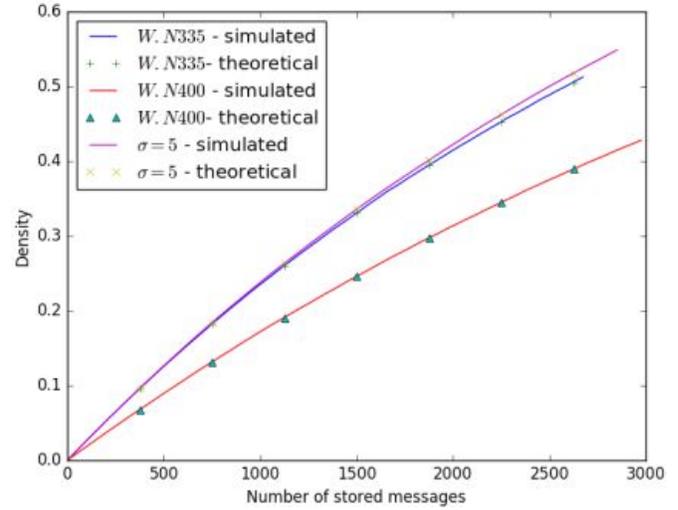


Figure 13. Evolution of the density for ordinary Willshaw networks with 335 and 400 neurons, and for a modified Willshaw network of side length 20 with 400 neurons and $\sigma = 5$, with stored messages of order 6.

As for the Willshaw network, we can calculate the probability of error of the modified network after one iteration of decoding. Given queries where c_e nodes have been removed, the probability of retrieval after one iteration can be estimated by:

$$P_{retrieve} = (1 - d^{c-c_e})^{N_c(\sigma, c-c_e)-c_e}. \quad (26)$$

One can then deduce the probability of error:

$$P_e = 1 - (1 - d^{c-c_e})^{N_c(\sigma, c-c_e)-c_e}. \quad (27)$$

Now referring to (24), this leads to:

$$P_e = 1 - \left(1 - \left[1 - \left(1 - \left(\frac{c}{Q} \right)^n \right)^{c-c_e} \right]^{N_c(\sigma, c-c_e)-c_e} \right), \quad (28)$$

which holds for lower values of σ . For higher values, however, the global density is no longer a proper estimator of the probability of spurious connections. Given $c - c_e$ message elements, the local connection density between these nodes and the restricted ensemble of allowed neighbors they can all be potentially connected to, is then higher than the average density over the whole network.

Figure 13 shows how the network density grows faster as messages are stored in the network, than for a classic Willshaw network with equal number of neurons. This is because of the decrease in number of possible connections due to the spacing constraint. When the number of connections of the compared classic and modified networks is close, the two densities evolve at a similar rate however.

Besides, we observe that the maximal improvement in performance, for given values of c and c_e , does not considerably vary as a function of the network size. This can be explained by the fact that the minimal distance giving the

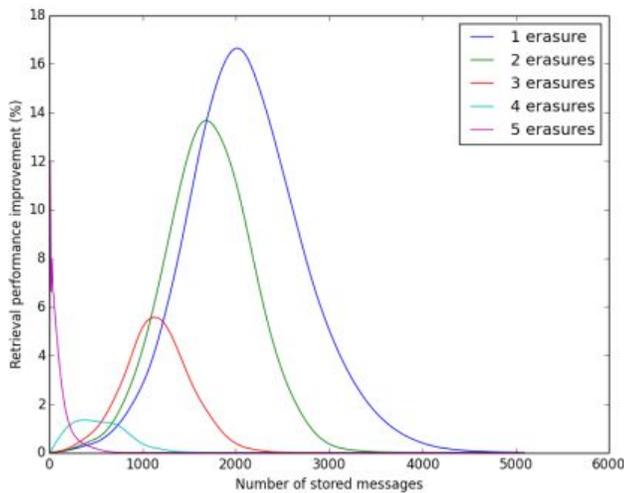


Figure 14. Maximal improvement obtained over a classic Willshaw network made of 400 units with messages of order 6, using a single iteration of decoding.

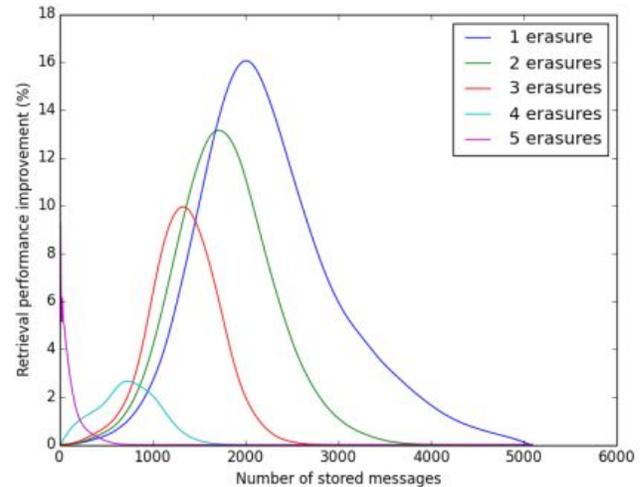


Figure 15. Maximal improvement obtained over a classic Willshaw network of 400 neurons with messages of order 6, using a maximum of 3 iterations.

best performance is approximately proportional to the size of the network. Consequently, the proportion of neurons in the network that cannot be connected to the $c - c_e$ neurons in the request remains more or less the same for different network sizes, with the optimal minimal distance.

The benefits brought by the constraint on connections tends to be stronger for smaller numbers of erasures. For erasures of about half the units of the messages, the maximum gain will be lower, yet for a high amount of erasures the performance may be more noticeably enhanced by the added constraint. The performance improvement over a classic Willshaw network also depends on the number of messages stored in the network. It reaches a peak for a certain number of stored messages, and then decays when additional messages get stored. The maximal improvement tends to be reached earlier during storage for higher numbers of erasures, as illustrated by Figures 14 and 15 for a network with 400 neurons, messages of order $c = 6$, and numbers of erasures c_e ranging from one to five. Figure 16 shows that for a larger network with 900 neurons, this arrangement is respected for the most part, with the exception of the case where $c_e = 1$, for which the peak in performance improvement occurs for a lower number of stored messages than for $c_e = 2$ or $c_e = 3$. Figures 14 and 15 also show that the maximum number of iterations applied during retrieval has a varying effect on performance improvement, depending on the number of erasures applied to form requests. With 400 neurons, increasing the number of iterations has a clear effect on performance for $c_e = 3$ and $c_e = 4$, more so than for $c_e = 1$ and $c_e = 2$.

When comparing networks with the same number of neurons, the greatest performance improvement is most often observed over a classic Willshaw network and a number of stored messages originally giving an error rate ranging from about 40%, up to 70%. The performance gain is then often close to 15%.

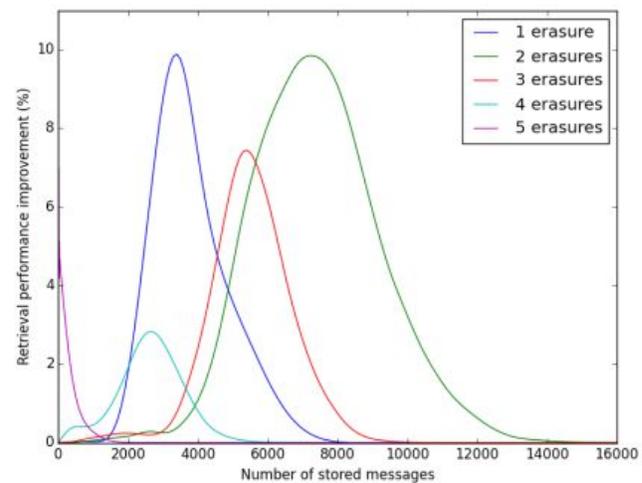


Figure 16. Maximal improvement obtained over a classic Willshaw network of 900 neurons with messages of order 6, using a maximum of 3 iterations.

VII. CONCLUSION AND PERSPECTIVES

We introduced a modified version of Willshaw neural networks that has interesting properties regarding storage capacity and retrieval performance. By prohibiting certain types of connections in the network, we observe that the retrieval ability can be enhanced, and that the value of the threshold on inter-neuron connection spacing has a direct impact on performance. As a result, this constraint can be used to enhance pattern completion performance without modifying the size of the pool of neurons. In addition, the applied constraint comes with an improved efficiency, as the quantity of information carried by each single connection is made higher. Also, given a fixed amount of allotted memory to store connection weights, a constrained network with more neurons but a comparable number of connections can be created, which will display even better performances.

This is relevant to observations on clustered cliques neural

networks, in that it shows constraining connections in a Willshaw network modifies its capacity and efficiency in a way that depends on the nature of the applied constraint. It is a step forward in understanding why the use of clusters in CCNs brings significantly higher capacity as compared to Willshaw networks. To some extent, it also emulates biological observations of lateral inhibition in the brain and sensory channels, as we prevent neighbor neurons from connecting and therefore let them compete for activity. This makes sense with a framework in which close-by neurons encode patterns that differ only in degree and where only one unit that resonates most with input stimuli must activate.

Future work may involve experimenting with other constraints on connections based on the relative locations of neurons, such as variable values of the spacing constraint applied in different subregions of the graph, or preventing random subsets of connections. It may also focus on theoretical advances in the understanding of the relationship between the nature of a connectivity constraint and its influence on efficiency and performance.

ACKNOWLEDGEMENTS

This work was supported by the European Research Council under Grant ERC-AdG2011 290901 NEUCOD.

The authors would like to thank NVIDIA for providing us with a free graphics card allowing to speed up computations for the experiments performed during this work.

REFERENCES

- [1] P. Tigreat, V. Gripon, and P. H. Horrein, "Improved Willshaw networks with local inhibition," Proceedings of the Eighth International Conference on Advanced Cognitive Technologies and Applications, 2016, pp. 96–101.
- [2] H. Jarollahi *et al.*, "Algorithm and architecture for a multiple-field context-driven search engine using fully-parallel clustered associative memories," In IEEE Workshop on Signal Processing Systems (SiPS), 2014, pp. 1–6.
- [3] H. Gazit *et al.*, "Network device architecture to support algorithmic content addressable memory (CAM)," U.S. Patent No. 9,367,645. Washington, DC: U.S. Patent and Trademark Office, 2016.
- [4] L. Y. Liu *et al.*, "CAM-based VLSI architectures for dynamic Huffman coding," IEEE Trans. Consum. Electron., vol. 40, no. 3, 1994, pp. 282–289.
- [5] M. E. Valle, "A class of sparsely connected autoassociative morphological memories for large color images," IEEE Transactions on Neural Networks, vol. 20, no. 6, 2009, pp. 1045–1050.
- [6] S. A. Shamma, "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," The Journal of the Acoustical Society of America, vol. 78, no. 5, 1985, pp. 1622–1632.
- [7] H. Ozeki *et al.*, "Inhibitory stabilization of the cortical network underlies visual surround suppression," Neuron, vol. 62, no. 4, 2009, pp. 578–592.
- [8] D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins, "Non-holographic associative memory," Nature, 1969, pp. 960–962.
- [9] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the national academy of sciences, vol. 79, no. 8, 1982, pp. 2554–2558.
- [10] P. Chossat, and M. Krupa, "Consecutive and non-consecutive heteroclinic cycles in Hopfield networks," Dynamical Systems, 2016, pp. 1–15.
- [11] R. J. McEliece *et al.*, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the national academy of sciences, vol. 33, no. 4, 1987, pp. 461–482.
- [12] M. Laiho *et al.*, "High-dimensional computing with sparse vectors," Biomedical Circuits and Systems Conference (BioCAS), 2015, pp. 1–4.
- [13] H. C. Longuet-Higgins, D. J. Willshaw, and O. P. Buneman, "Theories of associative recall," Quarterly reviews of biophysics, 3(02), 1970, pp. 223–244.
- [14] J. D. Keeler, "Comparison between kanerva's sdm and hopfield-type neural networks," Cognitive Science, vol. 12, no. 3, 1988, pp. 299–329.
- [15] D. Willshaw and P. Dayan, "Optimal plasticity from matrix memories: What goes up must come down," Neural Computation, vol. 2, no. 1, 1990, pp. 85–93.
- [16] A. Knoblauch, G. Palm, and F. T. Sommer, "Memory capacities for synaptic and structural plasticity," Neural Computation, vol. 22, no. 2, 2010, pp. 289–341.
- [17] V. Gripon and C. Berrou, "Sparse neural networks with large learning diversity," Neural Networks, IEEE Transactions on, vol. 22, no. 7, 2011, pp. 1087–1096.
- [18] A. A. Mofrad *et al.*, "Clique-based neural associative memories with local coding and precoding," Neural Computation, 2016
- [19] —, "A simple and efficient way to store many messages using neural cliques," Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on, 2011, pp. 1–5.
- [20] —, "Nearly-optimal associative memories based on distributed constant weight codes," Information Theory and Applications Workshop (ITA), 2012, pp. 269–273.
- [21] B. K. Aliabadi, C. Berrou, and V. Gripon, "Storing sparse messages in networks of neural cliques," Proceedings of the national academy of sciences, vol. 25, no. 5, 2014, pp. 461–482.
- [22] G. Palm, "Neural associative memories and sparse coding," Neural Networks, vol. 37, 1987, pp. 165–171.
- [23] J. R. Cavanaugh, W. Bair, and J. A. Movshon, "Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons," Journal of neurophysiology, vol. 88, no. 5, 2002, pp. 2530–2546.
- [24] M. A. Heller and E. Gentaz, "Psychology of touch and blindness," Psychology Press, 2013.
- [25] T. Kohonen, "The self-organizing map," Neurocomputing, vol. 21, no. 1, 1998, pp. 1–6.
- [26] A. Aboudib, V. Gripon, and X. Jiang, "A study of retrieval algorithms of sparse messages in networks of neural cliques," Proceedings of the Sixth International Conference on Advanced Cognitive Technologies and Applications, 2014, pp. 140–146.
- [27] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "Tree-based reparameterization framework for analysis of sum-product and related algorithms," IEEE Transactions on information theory, vol. 49, no. 5, 2003, pp. 1120–1146.