Variable binding through assemblies in spiking neural networks

Robert Legenstein
Institute for Theoretical Computer Science
Graz University of Technology
A-8010 Graz, Austria
legi@igi.tugraz.at

Christos H. Papadimitriou
EECS
UC Berkeley
CA 94720, USA
christos@cs.berkeley.edu

Santosh Vempala
College of Computing
Georgia Tech
Atlanta, GA 30308, USA
vempala@gatech.edu

Wolfgang Maass
Institute for Theoretical Computer Science
Graz University of Technology
A-8010 Graz, Austria
maass@igi.tugraz.at

Abstract

We propose a model for the binding of variables to concrete fillers in the human brain. The model is based on recent experimental data about corresponding neural processes in humans. First, electrode recordings from the human brain suggest that concepts are represented in the medial temporal lobe (MTL) through sparse sets of neurons (assemblies). Second, fMRI recordings from the human brain suggest that specific subregions of the temporal cortex are dedicated to the representation of specific roles (e.g., subject or object) of concepts in a sentence or visually presented episode. We propose that quickly recruited assemblies of neurons in these subregions act as pointers to previously created assemblies that represent concepts. As a proof of principle, we performed computer simulations of a spiking neural network model that implemented the proposed paradigm for binding through assembly pointers. We show that the model supports basic operations of brain computations, such as structured recall and copying of information.

1 Introduction

Numerous electrode recordings from the human brain (see [1] for a review) suggest that concepts are represented through sparse sets of neurons that fire (more or less) whenever the corresponding concept is activated. These data confirm earlier hypotheses and models about the representation of tokens of cognitive computations through assemblies of neurons [2]. More recent data [3] suggests that assemblies should not be seen as invariant entities, but as fluent coalitions of neurons whose synaptic interconnections can be strengthened very fast, even in response to a single experience. This data also suggests that these processes on the synaptic level underlie the formation of associations.

We propose that assemblies of neurons are also instrumental for creating a transient or longer lasting binding of a variable to a filler. For example, they could bind a variable that represents a thematic role (e.g., agent or patient in an episode) to a word or concept. Information about the neural representation of semantic roles is provided through recent fMRI data, where specific subregions in the temporal cortex were shown to respond to specific semantic (thematic) roles of individuals in an episode that was communicated through a sentence [4] or a movie [5].

Here we do not assume that semantic roles are represented by fixed assemblies of neurons. Such a fixed assembly would in general not have sufficient direct synaptic connectivity to the virtually unlimited repertoire of words or concepts, each represented through assemblies in other brain regions, that could acquire this semantic role in an episode. To achieve such large potential connectivity, the size of this fixed assembly would have to be so large that its activation would not be consistent with generic sparse firing activity in each brain region. Rather, we propose that the specific subregions of the temporal cortex that were shown to be activated differentially in dependence of the specific semantic role of a concept serve as large pools of neurons (we will refer to them as neural spaces). In neural spaces, sparse assemblies can quickly be recruited from the subset of neurons that happen to have direct synaptic connections to the assemblies for the corresponding concepts involved (assembly pointers). We propose that this model can reconcile functional needs, such as being able to recall the concept from its recent thematic role, with data on the inherently sparse connectivity between brain areas [6]. One can also view this model as a direct extrapolation of data on the formation of associations between concepts from [3] to associations between thematic roles (i.e., variables) and concepts.

We propose that one well-known neurophysiological mechanism is essential for the control of this binding process: disinhibition. At least two different ways how brain areas can be selectively disinhibited have been proposed on the basis of experimental data [7]. One is neuromodulatory control (especially cholinergic), see [8]. Another one is disinhibition via the activation of VIP cells, i.e., of inhibitory neurons that primarily target other types of inhibitory neurons [9]. Firing of VIP cells is apparently often caused by top-down inputs (they are especially frequent in layer 1, where top-down and lateral distal inputs arrive). Their activation is conjectured to enable neural firing and plasticity within specific patches of the brain through disinhibition, see e.g. [7, 8, 10, 11, 12]. We propose that disinhibition plays a central role for neural computation and learning by controlling operations on assembly pointers.

In this article, we briefly describe the proposed model of assembly pointers for variable binding and outline a spiking neural network that implements this model. A more detailed discussion for the model can be found in [13].

2 Results

Recent experimental data indicates that neural activity patterns in cortex can be characterized in first approximation as spontaneous and stimulus-evoked switching between the activations of different (but somewhat overlapping) subsets of neurons (see e.g. [14, 15, 16]), often referred to as assemblies of neurons. We therefore represent a specific content (a word or a concept) in our model by a specific assembly of neurons in a content space $\mathcal{C}$.

Our model for the binding of a variable that represents a syntactic role (agent, verb, patient) to a concrete word (referred to more abstractly as "content" in our model) is based on the results and hypotheses of [4]. We refer to the particular region or set of neurons that is reserved for a variable v as a neural space $\mathcal{N}_v$ for variable v. Thus each such neural space $\mathcal{N}_v$ can be viewed as functioning like a register in a computer in the terminology of [4]. But in contrast to a computer, this "register" is not used for storing content in it. Rather, assemblies in this register store "handles" or "pointers" to assemblies that store content information in the separate content space $\mathcal{C}$.

In addition our model takes into account that neurons typically do not fire just because they receive sufficiently strong excitatory input. Experimental data suggest that neurons are typically prevented from firing by an "inhibitory lock", that balances or even dominates excitatory input [17]. Thus a generic pyramidal cell is likely to fire because two events take place: its inhibitory lock is temporarily lifted ("disinhibition") and its excitatory input is sufficiently strong. Such disinhibition is apparently often caused by top-down inputs. We propose that orchestrated top-down disinhibition of neural spaces controls the formation of assembly pointers as well as the recall of content from assembly pointers and other cognitive operations.

We implemented this network structure with stochastically spiking neurons. The network consisted of a content space $\mathcal{C}$ and a neural space $\mathcal{N}_v$ for some variable v that each contained 1000 recurrently connected excitatory neurons (connection probability 0.1). To ensure sparse activity, lateral inhibition was implemented in a symbolic manner in each of the neural spaces through an inhibitory current that depended on the recent firing rate of neurons in the space. Disinhibition was modeled

through a multiplicative effect of an inhibitory input on the membrane potential of neurons. Reciprocal connections between $\mathcal{C}$ and $\mathcal{N}_v$ were introduced randomly with a connection probability of 0.1. Neurons in the content space received in addition connections from 200 input neurons. All synapses between excitatory neurons in the circuit were subject to spike-timing dependent plasticity (STDP).

First, five assemblies were induced in content space $\mathcal{C}$ by repeated presentation of 5 simple rate patterns $P_1, \ldots, P_5$ that represented concepts or words at the input. Due to these pattern presentations, an assembly $\mathcal{C}(P_i)$ emerged in content space for each of the patterns P_i (assembly sizes between 81 and 86 neurons) that showed robust firing activity whenever the corresponding pattern was presented as input. STDP of recurrent connections led to a strengthening of these synapses within each assembly, while synapses between assemblies remained weak (see [13] for details).

According to our model for variable binding, disinhibition enables the creation of an assembly pointer in some neural space $\mathcal{N}_v$ to the currently active assembly in the content space. Such disinhibition of a neural space $\mathcal{N}_v$ allows that some of neurons in it can fire, especially those that receive sufficiently strong excitatory input from a currently active assembly in the content space. Furthermore, in line with previously cited experimental reports we assume that this allowed firing of neurons in the neural space also enables plasticity of these neurons and synapses that are connected to it. To validate this hypothesis, we simulated disinhibition of the neural space $\mathcal{N}_v$ while input to content space $\mathcal{C}$ excited an assembly there. We found that STDP in the synapses that connect the content space $\mathcal{C}$ and the neural space $\mathcal{N}_v$ led to the stable emergence of an assembly in $\mathcal{N}_v$ within a second. Further, plasticity at recurrent synapses in $\mathcal{N}_v$ induced a strengthening of recurrent connections within assemblies there. Hence, disinhibition led to the rapid and stable creation of an assembly in the neural space, i.e., an assembly pointer. We denote such creation of an assembly pointer in a neural space $\mathcal{N}_v$ for a specific variable v to content P encoded in content space by CREATE(v, P).

Our model for variable binding based on assembly pointers further assumes that strengthened synaptic connections between assemblies in a neural space $\mathcal{N}_v$ and content space $\mathcal{C}$ enable the recall RECALL(v) of the variables' content, i.e., the activation of the assembly for content P in content space that was active at the most recent CREATE(v, P) operation (e.g., representing the word "truck"). It has been shown that the excitability of pyramidal cells can be changed in a very fast but transient manner through fast depression of GABA-ergic synapses onto pyramidal cells [18]. Using such a mechanism, a RECALL(v) can be initiated by disinhibition of the neural space $\mathcal{N}_v$ while the content space does not receive any bottom up input. The increased excitability of recently activated neurons in $\mathcal{N}_v$ ensures that the most recently active assembly is activated which in turn activates the corresponding content through its (previously potentiated) feedback connections to content space $\mathcal{C}$. The viability of this model for the recall of previously bound content was confirmed in simulations of the spiking neural network model described above, see [13] for details.

Apart from the creation of assembly pointers and recall of content, two further operations have been postulated to be essential for many higher cognitive functions [19]. The first is COPY(u, v) that copies the content of variable u to variable v. In our model, the copy operation creates an assembly pointer in neural space $\mathcal{N}_v$ for variable v to the content to which the assembly pointer in neural space $\mathcal{N}_u$ for variable u refers to. This operation can be implemented in our model simply by disinhibiting $\mathcal{N}_u$ in order to activate the corresponding content in $\mathcal{C}$ followed by a disinhibition of $\mathcal{N}_v$ in order to create an assembly pointer there. A final fundamental operation considered in [19] is COMPARE(u, v) which compares whether the content of u equals the content of v. One possible implementation of this operation in our model is a readout neuron that receives depressing synaptic connections from the content space. Then, when the content for $\mathcal{N}_u$ and $\mathcal{N}_v$ is recalled in sequence, readout synapses will be depressed for the content of $\mathcal{N}_v$ if and only if the content of $\mathcal{N}_u$ equals the content of $\mathcal{N}_v$. Such a "change detecting" readout thus exhibits high activity if the contents of $\mathcal{N}_u$ and $\mathcal{N}_v$ are different. We confirmed in computer simulations that these operations can be implemented in a spiking neural network model of assembly pointers, see [13].

Reproducing experimental data on the binding of agents to roles: Two experiments were performed in [4] that provided new insights in how variables may be encoded in cortex. Sentences were shown to participants where individual words (like "truck" or "ball") can occur as the agent or as the patient. In a first experiment, the authors aimed to identify cortical regions that encode the meaning of such sentences. Four example sentences with the words "truck" and "ball" are "The truck hit the ball" (S1), "The ball was hit by the truck" (S2), "The truck was hit by the ball" (S3), and "The ball hit the truck" (S4). Here, S1 and S2 (and S3 and S4 respectively) have the same meaning. Indeed,

the authors showed that a linear classifier is able to classify the meaning of such sentences from the fMRI signal of left mid-superior temporal cortex (lmSTC). Using our model for assembly pointers, we can model such situations by binding words either to an agent variable ("who did it") or to a patient variable ("to whom it was done"). Under the assumption that lmSTC hosts neural spaces (with assembly pointers) for the role of words, it is expected that the meaning of a sentence can be decoded from the activity there, but not from the activity in content space where the identities are encoded independently of their role. This conjecture was verified through computer simulations of our spiking neural network model for assembly pointers, see [13] for details.

A second experiment in [4] revealed that subregions of lmSTC also contain information about the current value of the variables for the agent and the patient. More specifically, the authors showed that one is able to predict from the fMRI signal of one subregion of lmSTC the identity of the agent and from the signal in another subregion the identity of the patient (generalizing over all identities of other roles and over different verbs). We confirmed through computer simulations that this is also the case in the proposed model since the assemblies that are formed in the neural spaces $\mathcal{N}_{\text{agent}}$ and $\mathcal{N}_{\text{patient}}$ are typically specific to the bound content. Note that such classification would fail if each neural space consisted of only a single assembly that is activated for all possible fillers [19], since in this case no information about the identity of the role is available in the neural space for the variable.

3 Discussion

It has often been emphasized (see e.g. [20, 21]) that there is a need to understand brain mechanisms for variable binding, and several models for variable binding had been proposed in the literature. These models fall into one of the general classes of pointer-based binding, binding by synchrony, or convolutional binding. Pointer-based models (e.g., [19, 22]) assume that pointers are implemented by single neurons or populations of neurons which are activated as a whole group. In contrast, our model is based on the assumption that distributed assemblies of neurons are the fundamental tokens for encoding symbols and content in the brain, and also for pointers. We propose that these assembly pointers can be created on the fly in some neural spaces for variables and occupy only a sparse subset of neurons in these spaces. It has been shown in [4] that the filler of a thematic role (e.g. the actor) can be predicted from the fMRI signal of a subregion in temporal cortex when a person reads a sentence. As shown above, this finding is consistent with assembly pointers. It is however inconsistent with models where a variable engages a population of neurons that is independent of the bound content, such as traditional pointer-based models. In comparison to traditional pointer models, the assembly pointer model could also give rise to a number of functional advantages. In a neural space $\mathcal{N}_v$ for a variable v, several instantiations of the variable can coexist at the same time, since they can be represented there by increased excitabilities of different assemblies. These contents could be recalled as different possibilities in a structured recall and combined in content space $\mathcal{C}$ with the content of other variables to in order to answer more complex questions.

Some data shows that the relation between spiking activity and the phases of underlying oscillatory population activity may play a role in hippocampus and for working memory [23], indicating a possible role of synchrony in the binding process. Still, the reliability and capacity of binding by synchrony is currently unclear. We note that, while our model is not based on precise synchronization of spikes in different neural spaces, the synaptic coupling between these spaces together with lateral inhibition leads to some synchronized oscillations of interacting neural spaces in our simulations. This is consistent with recent experimental data which suggest that common rhythms in two brain areas support the flow of excitation between these two areas, and also the potentiation of synapses between activated neurons in both areas [24].

Convolutional binding (see e.g., [25]) uses mathematical operations on high-dimensional vectors for variable binding. It had been used in the semantic pointer architecture of Eliasmith [26] where spiking neural networks were constructed to perform these rather complex operations. Similarly, the neural blackboard architecture (NBA, see e.g. [27]) relies on a number of neural circuits that were constructed for example to gate activity or to memorize associations. In contrast to these models, the assembly pointer model focuses on the emergence of binding operations, using assumptions on the fundamental level of assembly coding, network connectivity statistics, and plasticity processes.

We have presented in this article a model for variable binding through assemblies based on "assembly pointers". The model is consistent with recent findings on cortical assemblies and the encoding of

sentence meaning in cortex [4]. It provides a direct link between information processing on the level of symbols and sentences and processes on the level of neurons and synapses. The resulting model for brain computation allows top down structuring of incoming information, thereby laying the foundation of goal oriented „willful" information processing rather than just input-driven processing.

Acknowledgments

Written under partial support by the European Union project #604102 (Human Brain Project).

References

[1] R. Q. Quiroga. Neuronal codes for visual perception and memory. *Neuropsychologia*, 83:227–241, 2016.

[2] D. O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949.

[3] M. J. Ison, R. Q. Quiroga, and I. Fried. Rapid encoding of new memories by individual neurons in the human brain. *Neuron*, 87(1):220–230, 2015.

[4] S. M. Frankland and J. D. Greene. An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences*, 112(37):11732–11737, 2015.

[5] J. Wang, V. L. Cherkassky, Y. Yang, K. K., Chang, R. Vargas, N. Diana, and M.A. Just. Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. *Cognitive Neuropsychology*, 33(3-4):257–264, 2016.

[6] X. J. Wang and H. Kennedy. Brain structure and dynamics across scales: in search of rules. *Current opinion in neurobiology*, 37:92–98, 2016.

[7] J. J. Letzkus, S. B. E. Wolff, and A. Lüthi. Disinhibition, a circuit mechanism for associative learning and memory. *Neuron*, 88:264–276, 2015.

[8] R. C. Froemke and C. E. Schreiner. Synaptic plasticity as a cortical coding scheme. *Current Opinion in Neurobiology*, 35:185–199, 2015.

[9] K. D. Harris and G. M. G. Shepherd. The neocortical circuit: themes and variations. *Nature Neuroscience*, 18(2):170–181, 2015.

[10] P. Caroni. Inhibitory microcircuit modules in hippocampal learning. *Current Opinion in Neurobiology*, 35:66–73, 2015.

[11] C. K. Pfeffer. Inhibitory neurons: VIP cells hit the brake on inhibition. *Current Biology*, 24(1):R18–R20, 2014.

[12] Y. Fu, M. Kaneko, Y. Tang, a. Alvarez-Buylla, and M. P. Stryker. A cortical disinhibitory circuit for enhancing adult plasticity. *Elife*, 4:e05558, 2015.

[13] R. Legenstein, C. H. Papadimitriou, S. Vempala, and W. Maass. Assembly pointers for variable binding in networks of spiking neurons. *arXiv:1611.03698*, 2016.

[14] G. Buzsaki. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362–385, 2010.

[15] B. Bathellier, L. Ushakova, and S. Rumpel. Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron*, 76(2):435–449, 2012.

[16] A. Luczak and J. N. MacLean. Default activity patterns at the neocortical microcircuit level. *Frontiers in Integrative Neuroscience*, 6(30):doi: 10.3389/fnint.2012.00030, 2012.

[17] B. Haider, M. Häusser, and M. Carandini. Inhibition dominates sensory responses in the awake cortex. *Nature*, 493(7430):97–100, 2013.

[18] D. M. Kullmann, A. W. Moreau, Y. Bakiri, and E. Nicholson. Plasticity of inhibition. *Neuron*, 75(6):951–962, 2012.

[19] A. D. Zylberberg, L. Paz, P. R. Roelfsema, S. Dehaene, and M. Sigman. A neuronal device for the control of multi-step computations. *Papers in Physics*, 5:050006, 2013.

[20] G. F. Marcus. The algebraic mind: Integrating connectionism and cognitive science. *MIT Press*, 2003.

[21] G. F. Marcus, A. Marblestone, and T. Dean. The atoms of neural computation - does the brain depend on a set of elementary, reusable computations? *Science*, 346(6209):551–552, 2014.

[22] T. Kriete, D. C. Noelle, J. D. Cohen, and R. C. O'Reilly. Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. *Proceedings of the National Academy of Sciences*, 110(41):16390–16395, 2013.

[23] M. Siegel, M. R. Warden, and E. K. Miller. Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences*, 106(50):21341–21346, 2009.

[24] A. D. Friederici and W. Singer. Grounding language processing on basic neurophysiological principles. *Trends in Cognitive Sciences*, 19(6):329–338, 2015.

[25] T. A. Plate. Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6(3):623–641, 1995.

[26] C. Eliasmith, T. C. Stewart, X. Choo, T. Bekolay, T. DeWolf, Y. Tang, and D. Rasmussen. A large-scale model of the functioning brain. *science*, 338(6111):1202–1205, 2012.

[27] F. Van der Velde and M. De Kamps. Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, 29(01):37–70, 2006.