

Constraints on sequence processing speed in biological neuronal networks

Younes Bouhadjar^{1,2}, Markus Diesmann^{1,3}, Rainer Waser^{2,4}, Dirk J. Wouters⁴, Tom Tetzlaff¹

¹ Institute of Neuroscience and Medicine (INM-6) & Institute for Advanced Simulation (IAS-6) & JARA Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany

² Peter Grünberg Institute (PGI-7), Jülich Research Centre and JARA, Jülich, Germany

³ Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty & Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany

⁴ Institute of Electronic Materials (IWE 2) & JARA-FIT, RWTH Aachen University, Aachen, Germany

ABSTRACT

Sequence processing has been proposed to be the universal computation performed by the neocortex. The Hierarchical Temporal Memory (HTM) model provides a mechanistic implementation of this form of processing. While the model accounts for a number of neocortical features, it is based on networks of highly abstract neuron and synapse models updated in discrete time. Here, we reformulate the model in terms of a network of spiking neurons with continuous-time dynamics to investigate how neuronal parameters such as cell-intrinsic time constants and synaptic weights constrain the sequence-processing speed.

CCS CONCEPTS

• **Networks** → *Network dynamics*; • **Computing methodologies** → *Massively parallel algorithms*; *Temporal reasoning*.

KEYWORDS

sequence processing, neocortex, hierarchical temporal memory, spiking neurons, dendritic action potentials, spike timing dependent structural plasticity

1 INTRODUCTION

Learning and processing temporal sequences have been suggested to be *the* fundamental computation performed by the neocortex [1–3]. The Hierarchical Temporal Memory (HTM) model constitutes a mechanistic description of this type of computation [4]. It accounts for the specific anatomical structure of cortical (pyramidal) neurons, explains the functional role of dendritic action potentials, and learns continuously by means of local learning rules. The model can simultaneously learn and predict multiple sequences in streams of data and is robust with respect to failure of network elements and noise. So far, implementations of this model are based on highly abstract models of neurons and synapses with discrete-time dynamics. To foster an understanding of the sequence processing characteristics

in humans and other mammals, the model needs to be reformulated in terms of biophysical principles and parameters.

In this study, we deliver a continuous-time implementation of the temporal-memory algorithm proposed by the HTM theory [5], which comprises networks of spiking neurons, dendritic action potentials, backpropagating action potentials, lateral inhibition, and spike timing dependent structural plasticity. In the framework of this model, we investigate to what extent the sequence processing speed is constrained by neuronal parameters such as cell-intrinsic time constants and synaptic weights. We test the implementation in a task where the network learns random sequences of letters, and study the role of the inter-stimulus interval on the sequence prediction error, thereby deriving lower and upper bounds for the sequence processing speed.

2 MODEL

In the following, we will provide an overview of the network model components. A detailed description of the model and parameter values can be found in Appendix A and B.

2.1 Network model

The network consists of M minicolumns. Each minicolumn is composed of N_E excitatory (pyramidal) neurons (“E neurons”), that are recurrently connected to an inhibitory interneuron (Fig. 1A). The minicolumns are laterally connected via a sparse, random, plastic connectivity between the E neurons. Each E neuron receives K_{EE} randomly chosen connections from E neurons in other minicolumns. Initially, these K_{EE} connections are ‘potential’ connections. Only during the learning process, these potential connections can turn into effective synaptic connections (see Sec. 2.3).

Feedforward inputs representing the sequence elements are applied to the network in the form of N_{stim} external spike sources, each being connected to the E neurons of a random subset of L minicolumns (Fig. 1B).

2.2 Neuron model

The excitatory neurons are described as multicompartment models comprising a soma and N_D dendritic branches (Fig. 2). Each of the $N_D + 1$ compartments is described by a leaky integrate-and-fire (LIF) dynamics [6]. Each dendritic compartment receives excitatory $K_D = K_{EE}/N_D$ inputs from E neurons in other minicolumns. Upon threshold crossing, the dendritic compartments generate action potentials (“NMDA spikes”) that are propagated to the soma via

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Knoxville, Tennessee '19, July 23–25, 2019, Knoxville, Tennessee
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

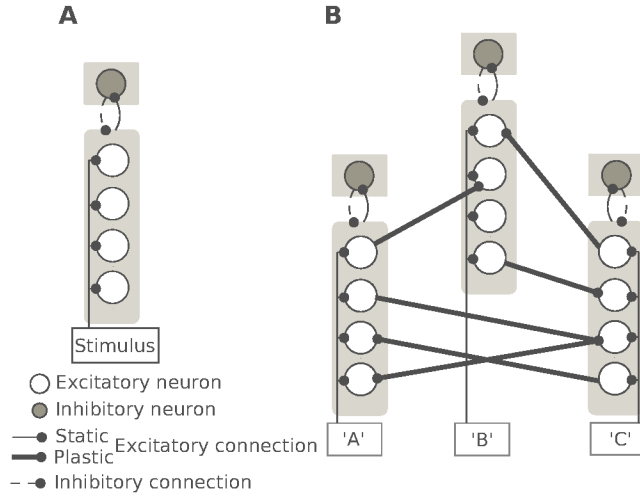


Figure 1: Sketch of the network architecture. A) Each minicolumn is composed of a number of excitatory neurons, which are recurrently connected to an inhibitory neuron, and driven by a feed-forward input (stimulus) representing an element in a sequence. B) The network is composed of a number of minicolumns, each representing an element (e.g. 'A', 'B', 'C') in a sequence. Minicolumns are interconnected via (learned) excitatory connections.

strong but subthreshold weights J_{SD} [7, 8]. To enable spike timing dependent plasticity at the dendritic synapses, somatic action potentials are backpropagated to the dendrite via suprathreshold connections with weight J_{DS} [9–11]. Inhibitory interneurons are described as simple point neuron models with LIF dynamics.

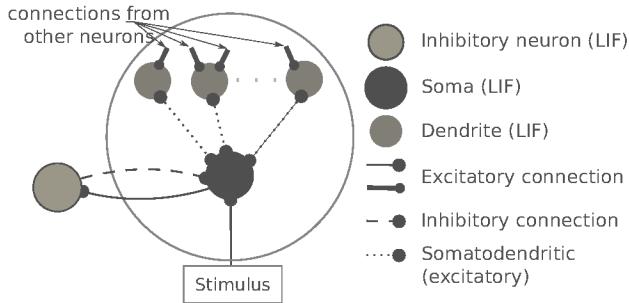


Figure 2: Sketch of the neuron model. Excitatory neurons (large white circle) are composed of a soma and a number of dendritic compartments. Each compartment is modeled as a leaky integrate-and-fire (LIF) unit. The feed-forward input ('stimulus') projects to the soma. Inputs from other minicolumns arrive at the dendritic compartments. Connections between the dendritic compartments and the soma are bidirectional to permit propagation of dendritic spikes to the soma and backpropagation of somatic spikes to the dendrites. Inhibitory (LIF) neurons are communicating with the soma.

2.3 Synapse models

For all types of connection in the network, synapses are modeled as exponential currents with amplitude $\hat{I}$, time constant τ_s and delay d . We parameterize the strength of each connection by the weight J , which corresponds to the amplitude of the membrane potential deflection caused by a single synaptic event. We distinguish between four types of connections:

EE synapses refer to connections from the somata of E neurons to the dendrites of other E neurons. The strength of these connections is determined by a Hebbian form of structural plasticity: Each EE synapse is characterized by a permanence P and a weight J_{EE} . The permanence models the level of maturity of the synapse. Premature synapses, i.e. synapses where P is below a threshold θ_P , are not effective and assigned zero weight $J_{EE} = 0$. If $P > \theta_P$, J_{EE} is set to a fixed value W , which is chosen such that the target neuron generates a dendritic spike if θ_E presynaptic neurons fire collectively. For each postsynaptic (dendritic) spike, the permanence P is updated with an increment $\Delta_f P$ or decrement $\Delta_d P$ that depends on the difference $t_{\text{post}} - t_{\text{pre}}$ between the timing of pre- and postsynaptic spikes. This form of spike timing dependent structural plasticity mimics the plasticity mechanism used in the original HTM model, and is also known from the neuroscience literature [12]. It promotes the formation of connections between neurons that are consistently activated in a sequence and the pruning of unused connections.

IE synapses represent the connections from the somata of E neurons to the inhibitory neurons. The weights J_{IE} are set such that the collective firing of a certain number of θ_I E neurons in the minicolumn causes the inhibitory neuron to fire.

EI synapses refer to connections from the inhibitory neurons to the somata of E neurons in the respective minicolumn with a strong negative weight J_{EI} .

EF synapses characterize connections from the external feedforward input sources to the somata of E neurons with suprathreshold weight J_{EF} .

Apart from the EE connections, all synaptic weights are constant (static).

2.4 Simulation details

The network simulations are performed in NEST version 2.16.0 [13]. Network states are synchronously updated using exact integration of the system dynamics on a discrete time grid with step size δt [14]. For every network realization, the initial permanences and membrane potentials (at time 0) are drawn randomly and independently from uniform distributions in the interval $[P_0, \theta_P]$ and $[V_r, \theta]$, respectively.

3 RESULTS

3.1 Sequence learning and prediction in networks of spiking neurons

All neurons in a given minicolumn share the same feedforward input. Before learning or at the beginning of a new sequence, a feedforward spike will cause all neurons in the respective minicolumn to fire (Fig. 3A). In the HTM model, this is referred to as columnar bursting. If the number of inputs to a dendritic branch from a specific presynaptic minicolumn exceeds a threshold θ_E , the

simultaneous activation of these presynaptic neurons results in a dendritic spike, followed by a strong depolarization of the soma which corresponds to the “predictive” state (Fig. 3B). Predictive neurons receiving a feedforward input will fire earlier as compared to nonpredictive neurons (Fig. 3C). The advanced spike of the predictive neurons will initiate a fast and strong inhibitory feedback to the entire minicolumn, ultimately suppressing the firing of the nonpredictive neurons.

As described in Sec. 2.3, the EE connections are subject to spike timing dependent structural plasticity. Repetitive sequential activation of feedforward inputs (characters) during the learning therefore leads to the formation of a connectivity structure that corresponds to an imprint of the presented sequences in the sense that the activation of a certain minicolumn by a feedforward input will generate dendritic spikes, i.e. a “prediction”, in a specific subset of neurons in the minicolumns representing the next sequence element.

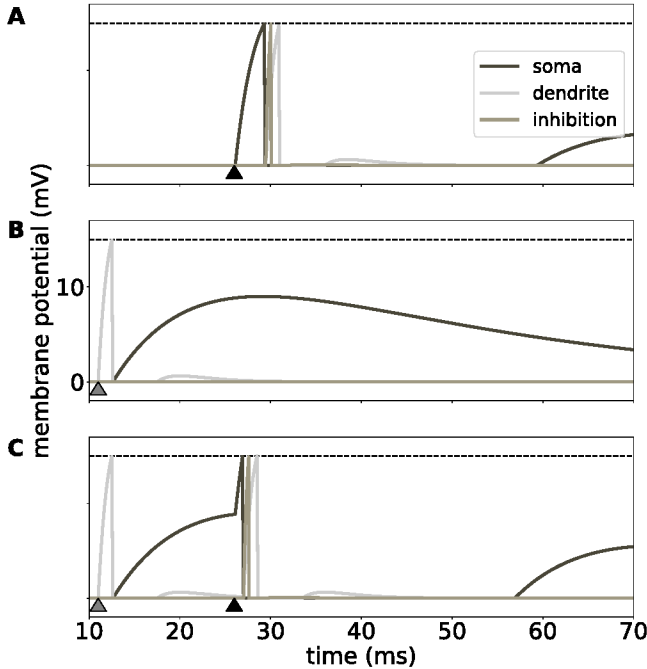


Figure 3: Membrane-potential responses to a feedforward input (black arrow; A), a suprathreshold dendritic input (gray arrow; B), and a combination of feedforward and dendritic input (C). Bars mark spike times. A dendritic spike preceding the feedforward input (as in C) can speed up somatic, and hence, inhibitory firing, provided the time interval between the dendritic spike and the feedforward input is in the right range (compare somatic and inhibitory spikes in A and C). The advanced firing of the inhibitory neuron suppresses somatic spiking in neurons of the same minicolumn that have not received suprathreshold dendritic activation. Here, we used an E neuron with $N_D = 1$ dendritic compartment. The soma of E neuron is connected to an inhibitory neuron.

3.2 Sequence prediction performance

To test the sequence prediction performance, we repetitively stimulate the network with batches (trials) of $S = 2$ sequences, each sequence consisting of $C = 10$ characters randomly chosen (with

replacement) from an alphabet of length $A = 10$. Within a sequence, the same character can occur multiple times. Each batch consists of the same set of sequences with identical order.

The prediction error is continuously monitored and assessed for every batch as follows: For each presentation of a character q (feedforward input) at time t_0 , find all dendrites that are firing in the time interval $(t_0 - \Delta T, t_0)$, with ΔT being the inter-stimulus timing. If a minicolumn contains at least θ_1 E neurons with firing dendrites, this minicolumn is considered as active. The activation of minicolumns is encoded in a binary vector z with $z_i = 1$ ($i = 1, \dots, M$) if the i th minicolumn is active, and $z_i = 0$ else. The prediction error is computed as the Euclidean distance between z and the binary target vector t representing the pattern of feedforward inputs for each character q . At the end of each batch, we compute the total prediction error by averaging over all characters within this batch. When assessing the prediction error, we discard the first character in each sequence.

After each new network instantiation, the initial normalized prediction error is at 100%. With increasing number of training episodes (presentation of a batch), the prediction error decreases as the network explores and learns the sequences, and saturates after about 300 episodes (Fig. 4). Note that the model accounts for context dependence: prediction of the next character depends not only on the current sequence element, but on all preceding characters within this sequence. For the first character in a sequence, the context is not defined. As a result, the network predicts all possible subsequent characters from all contexts (i.e. all sequences, where the same character occurs), thereby leading to a residual error, here slightly below 20%.

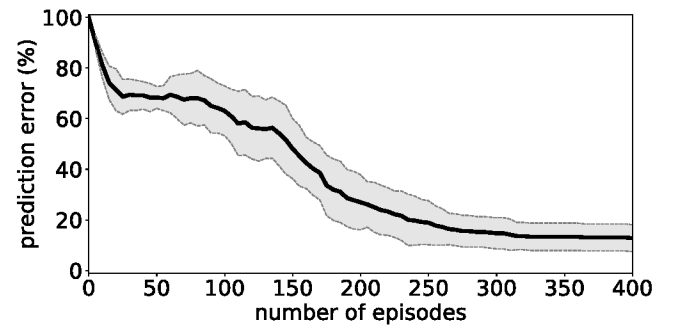


Figure 4: Dependence of the normalized sequence prediction error on the number of training episodes. Solid and dashed curves represent mean $\pm$ standard deviation of the prediction error across 15 different network realizations. Parameters: number of minicolumns $M = 10$, number of neurons per minicolumn $N_E = 30$, number of dendrites per neuron $N_D = 1$, number of synapses per dendrite $K_D = 180$, number of sequences per batch $S = 2$, number of characters per sequence $C = 10$, length of alphabet $A = 10$, inter-stimulus interval $\Delta T = 80$ ms, prediction error assessment for every 4th batch (see Appendix B for remaining parameters).

3.3 Role of the stimulus timing

In general, the timing of the feedforward inputs affects the dynamics of the network in two respects: first, reliable predictions of sequence elements can only be made if the time interval ΔT between two

consecutive stimulus presentations is such that the second input sufficiently coincides with the somatic depolarization caused by the dendritic spike triggered by the first stimulus (Fig. 3C). Second, only if the time difference between consecutive stimulus presentations matches the potentiation window of the spike timing dependent structural plasticity rule, the network can learn the sequences. Here, we restrict our investigation to the first factor. The second factor is removed by adjusting the potentiation window $[\Delta T - \Delta t_p, \Delta T + \Delta t_p]$ of the plasticity rule when varying ΔT .

The model predicts both an upper and a lower bound for the optimal inter-stimulus interval ΔT (Fig. 5). For our choice of network parameters, the optimal inter-stimulus intervals are in the range of about 20–150 ms. For inter-stimulus interval outside this range, the feedforward spike misses the somatic response to the dendritic spike, ultimately leading to a columnar burst. The optimal range of inter-stimulus intervals is determined by the neuronal and synaptic time constants, in particular by the width of the somatic postsynaptic potential caused by a dendritic spike, but also by the strength J_{SD} of the somatodendritic coupling.

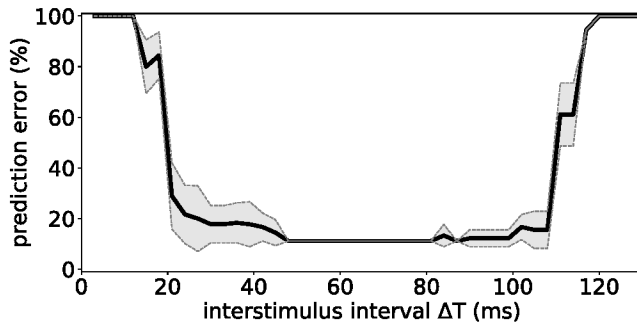


Figure 5: Dependence of the normalized sequence prediction error on the interstimulus interval ΔT (after 400 training episodes). Same parameters as in Fig. 4.

4 DISCUSSION

In the context of our study, the sequence processing speed is determined by the time interval ΔT between subsequent presentations of sequence elements that guarantee a successful prediction, i.e. a coincidence between the feedforward input and the somatic depolarization caused by dendritic spikes. Our model demonstrates that the optimal range of inter-stimulus intervals is solely determined by low-level neuronal and synaptic parameters, but not by the task complexity such as the sequence or alphabet length, the number of learned sequences, or the ambiguity level [5].

Our model suggests a number of biological mechanisms by which the processing speed could be modulated. Candidates are variations in synaptic or membrane time constants, e.g. controlled by the level of synaptic background input, or a modulation in the effective strength of feedforward inputs or somatodendritic coupling, e.g. through synaptic or intrinsic plasticity.

Note that an HTM implementation based on spiking neurons with continuous-time dynamics was already presented in [15]. This implementation mainly served the purpose of porting the HTM model to an analog-digital neuromorphic hardware system, and

differs from our model in several respects. The study did not address the role of neuronal parameters for the sequence processing speed.

The network model underlying our study is still very simplistic. Lateral inhibition within a minicolumn, for example, is mediated by a single interneuron with unrealistically strong and fast connections to and from the pool of excitatory neurons. In future versions, we will replace this interneuron by a recurrently connected network of inhibitory neurons [16], thereby permitting more realistic weights and simultaneously speeding up the interaction between inhibitory and excitatory cells by virtue of the fast-tracking property of such networks [17]. Another limiting factor of the current implementation is the network size: the number of minicolumns, the number of neurons within minicolumns, the number of dendritic branches per neuron, as well as the number of synapses per neuron are far from realistic [5]. The number of sequences that can be successfully learned in this network is hence small. Note, however, that high network capacity is not the central aim of this study. The model is minimalistic on purpose and serves as a means to understand what network characteristics determine the sequence processing speed. In future work, however, we aim at upscaling the network, thereby improving its sequence processing capacity and performance.

ACKNOWLEDGMENTS

This project was funded by the Helmholtz Association Initiative and Networking Fund (project number SO-092, Advanced Computing Architectures), and the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2).

REFERENCES

- [1] Lashley, K. S. (1951). *The problem of serial order in behavior*, Volume 21. Bobbs-Merrill.
- [2] Hawkins, J., & Blakeslee, S. (2007). *On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines*. Macmillan.
- [3] Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature neuroscience* 17(5), 732.
- [4] Hawkins, J., Ahmad, S., & Dubinsky, D. (2011). Cortical learning algorithm and hierarchical temporal memory. *Numenta Whitepaper*, 1–68.
- [5] Hawkins, J., & Ahmad, S. (2016). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits* 10.
- [6] Lapique, L. (1907). Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.* 9, 620–635.
- [7] Antic, S. D., Zhou, W.-L., Moore, A. R., Short, S. M., & Ikonomu, K. D. (2010). The decade of the dendritic nmda spike. *Journal of neuroscience research* 88(14), 2991–3001.
- [8] Major, G., Larkum, M., & Schiller, J. (2013). Active properties of neocortical pyramidal neuron dendrites. *Annu. Rev. Neurosci.* 36, 1–24.
- [9] Stuart, G. J., & Häusser, M. (2001). Dendritic coincidence detection of epsps and action potentials. *Nature neuroscience* 4(1), 63.
- [10] Chklovskii, D., Mel, B., & Svoboda, K. (2004). Cortical rewiring and information storage. *Nature* 431, 782–788.
- [11] Losonczy, A., Makara, J. K., & Magee, J. C. (2008). Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature* 452(7186), 436.
- [12] Deger, M., Helias, M., Rotter, S., & Diesmann, M. (2012). Spike-timing dependence of structural plasticity explains cooperative synapse formation in the neocortex. *PLoS Comput. Biol.* 8(9), e1002689.
- [13] Linssen, C., Lepperød, M. E., Mitchell, J., Pronold, J., Eppler, J. M., Keup, C., Peyser, A., Kunkel, S., Weidel, P., Nodden, Y., Terhorst, D., Deepu, R., Deger, M., Hahne, J., Sinha, A., Antonietti, A., Schmidt, M., Paz, L., Garrido, J., Ippen, T., Riquelme, L., Serenko, A., Kühn, T., Kitayama, I., Mørk, H., Spreizer, S., Jordan, J., Krishnan, J., Senden, M., Hagen, E., Shusharin, A., Vennemo, S. B., Rodarie, D., Morrison, A., Graber, S., Schuecker, J., Diaz, S., Zajzon, B., & Plesser, H. E. (2018). NEST 2.16.0.
- [14] Rotter, S., & Diesmann, M. (1999). Exact digital simulation of time-invariant linear systems with applications to neuronal modeling. *Biol. Cybern.* 81(5–6), 381–402.
- [15] Billautelle, S., & Ahmad, S. (2015). Porting htm models to the heidelberg neuromorphic computing platform. *arXiv preprint arXiv:1505.02142*.

- [16] Hu, H., Gan, J., & Jonas, P. (2014). Fast-spiking, parvalbumin+ gabaergic interneurons: From cellular design to microcircuit function. *Science* 345(6196), 1255263.
- [17] van Vreeswijk, C., & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 10(6), 1321–1371.

A NETWORK MODEL

Summary		
Populations	ensemble of minicolumns, each composed of an excitatory and an inhibitory subpopulation	
Connectivity	• sparse random connectivity between excitatory neurons subject to spike-timing-dependent structural plasticity • local recurrent connectivity between excitatory and inhibitory neurons (static)	
Neuron model	• excitatory neurons: multi-compartment model with one soma compartment and one compartment per dendritic branch, each compartment modeled by leaky integrate-and-fire (LIF) dynamics • inhibitory neurons: point neuron (LIF)	
Synapse model	additive (linear) exponential postsynaptic currents (PSCs)	
Plasticity	spike-timing-dependent structural plasticity in excitatory-to-excitatory connections	
Input	external spike sources, connected to excitatory neurons	
Populations		
Ensemble of minicolumns $\{i i = 1, \dots, M\}$		
Name	Elements	Size
$\mathcal{E}_i$	excitatory neurons in minicolumn i	n_E
$\mathcal{I}_i$	inhibitory neurons in minicolumn i	n_I
Connectivity		
Source	Target	Pattern
$\mathcal{E}_i (\forall i)$	$\mathcal{E}_j (\forall j)$	random, fixed in-degrees K_{EE} and delays d_{EE} , plastic weights $J_{EE,ji}$ (EE connections)
$\mathcal{E}_i (\forall i)$	$\mathcal{I}_i (\forall i)$	all-to-all, fixed delays d_{IE} , fixed weights J_{IE} (suprathreshold) (IE connections)
$\mathcal{I}_i (\forall i)$	$\mathcal{E}_i (\forall i)$	all-to-all, fixed delays d_{EI} , fixed weights J_{EI} (EI connections)
$\mathcal{I}_i (\forall i)$	$\mathcal{I}_i (\forall i)$	none (II connections)
no self-connections (“autapses”), no multiple connections (“multapses”)		
Neuron		
Excitatory neurons	multi-compartment model with one soma compartment and N_D dendritic branches • soma compartment: – leaky-integrate-and-fire (LIF) dynamics (see below) • dendritic compartments: – LIF dynamics (see below) – K_D (excitatory) synapses per dendritic branch	
Inhibitory neurons	point neuron (LIF)	
Dynamics of single compartments (LIF)	dynamics of membrane potential $V_c(t)$ of compartment c : • spike emission at time t_k^c if $V_c(t_k^c) \geq \theta$ • subthreshold dynamics: $\tau_m \dot{V}_c = -V_c + R_m I_c(t) \quad \forall k, \forall t \notin \left[t_k^c, t_k^c + \tau_{\text{ref}}\right)$ • reset and refractoriness: $V_c(t) = V_r \quad \forall k, \forall t \in \left(t_k^c, t_k^c + \tau_{\text{ref}}\right]$	
Coupling between compartments of excitatory neurons	• pulse-coupled interaction (spikes) • dendrite $\rightarrow$ soma coupling: linear exponential post-synaptic current, fixed weight J_{SD} (PSP amplitude), time constant τ_{SD}, delay d_{SD} • soma $\rightarrow$ dendrite coupling (action-potential backpropagation): linear exponential post-synaptic current, fixed weight J_{DS} (PSP amplitude, suprathreshold), time constant τ_{DS}, delay d_{DS}	

Synapse	
Model	current based synapses with additive exponential post-synaptic currents (PSCs): $I_i(t) = \sum_{j=1}^N \hat{I}_{ij}(\text{PSC} * s_j)(t)$ with $\text{PSC}(t) = e^{-t/\tau_s} \Theta(t)$ and Heaviside function $\Theta(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{else} \end{cases}$ $\leadsto$ post-synaptic potential $\text{PSP}_{ij}(t) = \hat{I}_{ij} \frac{R_m \tau_s}{\tau_s - \tau_m} (e^{-t/\tau_s} - e^{-t/\tau_m}) \Theta(t)$
Plasticity	spike-timing dependent structural plasticity for EE connections; each synapse described by permanence P (maturity) and synaptic weight J_{EE}; update of P_{ij} and $J_{EE,ij}$ for each postsynaptic spike at time t_{post}: <pre> if $\Delta T - \Delta t_p < t_{\text{post}} - t_{\text{pre}} < \Delta T + \Delta t_p$ then $P \leftarrow P + \Delta_f P$ else $P \leftarrow P - \Delta_d P$ end if if $P > P_{\text{max}}$ then $P = P_{\text{max}}$ end if if $P > P_\theta$ then $J_{EE} \leftarrow W$ else $J_{EE} \leftarrow 0$ end if </pre> here, t_{pre} denotes the spike timing of the presynaptic neuron. ΔT is the interstimulus timing, Δt_p the range from ΔT at which the permanence is incremented and ΔP is the amount used to increment the permanence.
Input	
• N_{stim} spike-generating sources, each representing a character in the sequences • each spike source connected to a random subset of L minicolumns via a synaptic weight J_{EF} and a delay of d_{EF}	
Training	
• repetitive stimulation of network (k times) using the same batch of data. • a batch of data contains S sequences, each sequence is composed of C characters	
Initial conditions	
• for every network realization: initial permanences and membrane potentials (at time 0) randomly and independently drawn from uniform distributions in intervals $[P_0, \theta_p]$ and $[V_r, \theta]$, respectively	
Simulation details	
• network simulations performed in NEST version 2.16.0 [13] • synchronous update using exact integration of the system dynamics on a discrete time grid with step size Δt [14]	

Table 1: Description of the network model

B PARAMETERS

Name	Value	Description
Network		
n_E	30	number of excitatory neurons per minicolumns
n_I	1	number of inhibitory neurons per minicolumns
M	10	number of minicolumns
(Potential) Connectivity		
K_{EE}	180	number of excitatory inputs per excitatory neuron (EE in-degree)
K_{EI}	1	number of inhibitory inputs per excitatory neuron (EI in-degree)
K_{IE}	n_E	number of excitatory inputs per inhibitory neuron (IE in-degree)
K_{II}	0	number of inhibitory inputs per inhibitory neuron (II in-degree)
Excitatory neurons		
Soma		
τ_m	10 ms	membrane time constant
τ_{ref}	30 ms	absolute refractory period
C_m	250 pF	membrane capacity
V_r	0.0 mV	reset potential
θ	15 mV	spike threshold
Dendrite		
N_D	1	number of dendritic branches per excitatory neuron
K_D	180 ($K_D = K_{EE}$)	number of synapses per dendritic branch
τ_m	3 ms	membrane time constant
τ_{ref}	2 ms	absolute refractory period
C_m	250 pF	membrane capacity
V_r	0.0 mV	reset potential
θ	15 mV	spike threshold
Somato-dendritic coupling		
J_{SD}	9.0 mV	dendrite $\rightarrow$ soma coupling strength (amplitude of PSP evoked by dendritic spike)
J_{DS}	8.25 mV	soma $\rightarrow$ dendrite coupling strength (amplitude of dendritic PSP evoked by somatic spike)
τ_{SD}	30 ms	time constant of dendrite $\rightarrow$ soma coupling (post-synaptic current)
τ_{DS}	2 ms	time constant of soma $\rightarrow$ dendrite coupling (post-synaptic current)
d_{SD}	0.1 ms	delay of dendrite $\rightarrow$ soma coupling
d_{DS}	0.1 ms	delay of soma $\rightarrow$ dendrite coupling
Inhibitory neurons		
τ_m	5 ms	membrane time constant
τ_{ref}	2 ms	absolute refractory period
C_m	250 pF	membrane capacity
V_r	0.0 mV	reset potential
θ	15 mV	spike threshold

Name	Value	Description
Synapse		
$J_{EE,ij}$	$\{0, W\}$	E-neuron $\rightarrow$ E-neuron connections strength (amplitude of dendritic PSP of neuron i evoked by somatic spike of neuron j)
W	7.5 mV	amplitude of PSP evoked by somatic spike in the case of a mature synapse, $\theta_E = \frac{\theta}{W} = 2$
J_{IE}	3.75 mV	E-neuron $\rightarrow$ I-neuron connections strength (amplitude of the inhibitory neuron i evoked by somatic spike of neuron i), $\theta_I = \frac{\theta}{J_{IE}} = 4$
J_{EI}	-22.5 mV	I-neuron $\rightarrow$ E-neuron connections strength (amplitude of somatic PSP of neuron i evoked by the inhibitory neuron i)
$\tau_{s,EE}$	2 ms	time constant of E-neuron $\rightarrow$ E-neuron connection (post-synaptic current)
$\tau_{s,IE}$	0.5 ms	time constant of E-neuron $\rightarrow$ I-neuron connection (post-synaptic current)
$\tau_{s,EI}$	2 ms	time constant of I-neuron $\rightarrow$ E-neuron connection (post-synaptic current)
d_{EE}	5 ms	delay of E-neuron $\rightarrow$ E-neuron connection
d_{IE}	0.1 ms	delay of E-neuron $\rightarrow$ I-neuron connection
d_{EI}	0.1 ms	delay of I-neuron $\rightarrow$ E-neuron connection
Plasticity		
P_{ij}	$[P_{\min}, P_{\theta}]$	synaptic permanence
$\Delta_f P$	3	permanence increment
$\Delta_d P$	1	permanence decrement
W	9.0 mV	weight (J_{EE}) of mature synapses
P_{θ}	2300	synapse maturity threshold
$P_{\max}$	4000	upper permanence bound
$P_{\min}$	2000	lower permanence bound
Δt_p	5	interval for synaptic increment
Input		
N_{stim}	10	number of external spike sources
L	1	number of minicolumns connected to the same spike source
S	2	number of sequences per batch
C	10	number of characters per sequence
A	10	length of the total alphabet
J_{EF}	16.5 mV	Feedforward input $\rightarrow$ E-neuron connections strength (amplitude of somatic PSP of neuron i evoked by feedforward spike)
$\tau_{s,EF}$	3.16 ms	time constant of Feedforward input $\rightarrow$ E-neuron connection (post-synaptic current)
d_{EF}	0.1 ms	delay of Feedforward input $\rightarrow$ E-neuron connection
Simulation		
ΔT	[3, 180] ms	interstimulus timing
Δt	0.1 ms	time resolution
k	400	training steps

Table 2: Network and simulation parameters