Cluster Synchronization as a Mechanism of Free Recall in Working Memory Networks

MATIN JAFARIAN [1] (Member, IEEE), DAVID CHÁVEZ HUERTA [1], GIANLUCA VILLANI [2], ANDERS LANSNER [3], AND KARL H. JOHANSSON [4] (Fellow, IEEE)

(Synchronization in Natural and Engineering Systems)

[1]RecalLab, Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands
[2]Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada
[3]Division of Computational Science and Technology, KTH Royal Institute of Technology, and Stockholm University, 114 28 Stockholm, Sweden
[4]Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, and Digital Futures, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

CORRESPONDING AUTHOR: MATIN JAFARIAN (e-mail: m.jafarian@tudelft.nl).

ABSTRACT This article studies free recall, i.e., the reactivation of stored memory items, namely *patterns*, in any order, of a model of working memory. Our free recall model is based on a biologically plausible modular neural network composed of H modules, namely *hypercolumns*, each of which is a bundle of M *minicolumns*. The coupling weights and constant bias values of the network are determined by a Hebbian plasticity rule. Using techniques from nonlinear stability theory, we show that cluster synchronization is the central mechanism governing free recall of orthogonally encoded patterns. Particularly, we show that free recall's cluster synchronization is the combination of two main mechanisms: simultaneous activities of minicolumns representing an encoded pattern, i.e., within-pattern synchronization, together with time-divided activities of minicolumns representing different patterns. We characterize the coupling and bias value conditions under which cluster synchronization emerges. We also discuss the role of heterogeneous coupling weights and bias values of minicolumns' dynamics in free recall. Specifically, we compare the behaviour of two $H \times 2$ networks with identical and non-identical coupling weights and bias values. For these two networks, we obtain bounds on couplings and bias values under which both encoded patterns are recalled. Our analysis shows that having non-identical couplings and bias values for different patterns increases the possibility of their free recall. Numerical simulations are given to validate the theoretical analysis.

INDEX TERMS Network analysis and control, stability of nonlinear systems, systems neuroscience.

I. INTRODUCTION

Synchronization is among the key collective behaviors of many complex networks, including biological networks. A prominent example is a network of biological neurons whose activities govern the underlying mechanisms of motion, cognition, and memory.

The importance of understanding human memory functioning is evident from its central role in our cognitive functions [1], and its role as the main inspiration behind developments of artificial memory networks [2]. In general, the human memory system can be categorized into short-term and long-term memory modules. Among the short-term modules is working memory (WM), which is responsible for holding and processing information in a temporary fashion in service of higher-order cognitive tasks, for example, reasoning and decision-making [3].

Among the most important types of memory tasks associated with WM is free recall: the reactivation of a stored memory item, namely pattern, in the absence of the original stimulus. A classical example is an experiment in which a subject is exposed to a set of words and is then asked to recall them in any order. Empirical studies on human subjects have established free recall as a measure of working memory capacity [4]. Nonetheless, the precise mechanisms underlying

free recall dynamics in the human brain are not yet fully understood [5]. Several abstract neuro-computational [6] as well as more detailed spiking neural network models have been built based on different neurobiological hypotheses [7], at different levels of abstractions, to account for human experimental data on WM.

Attractor neural networks, dynamical networks that evolve towards a stable pattern, have been employed for understanding the mechanisms of memory [8], [9], [10], [11], [12], [13], including WM [7], and the display of bursty limit cycle replay of encoded memories [14]. The most studied attractor neural network model is the Hopfield model, capable of storing different patterns as stable equilibria [15], [16]. The abstract Hopfield model represents a memory network with a fully connected graph, symmetric weights, and capable of storing only binary values. The model has limitations on the capacity as well as recalling previously stored patterns [6]. The latter has motivated designs with more than two states including the modular recurrent networks of Potts type [17]. The model in this paper is originated from a biologically relevant modular model of Potts type [5]. The network model in [5] is composed of H hypercolumns, each of which consists of a bundle of M elementary units, minicolumns, interacting via lateral inhibition, such that each hypercolumn module acts as a winner-take-all microcircuit.

The modular WM dynamic network, as described above, is supposed to encode, hold, and recall information. Several explanations account for WM ability to hold memory patterns during a delay period between learning and recall, either by persistent [18], [19], [20], [21] or bursty activity [14], [22], [23] of neural populations, or a combination of both [24], [25]. The model we consider [5] assumes a fast Hebbian plasticity learning rule, i.e., changes in synaptic strength, combined with slow adaptation in a recurrent neural network. This model allows us to assume constant weights during a very brief holding period for studying the immediate free recall of WM, which is among standard memory exercises.

In this article, we focus on the post learning phase aiming to show the role of synchronization in WM free recall. In fact, the role of sustained, bursty synchronization, or both in free recall or replay of WM has been intensely discussed in experimental research. In particular, it has been shown that specific areas of the human cortex synchronize when engaged in WM tasks [26], [27]. In this sense, several prior works have considered synchronization as a major player in WM functioning [28], [29], [30], [31]. Considering mathematical analysis, the similarities between attarctor and pulse-coupled neural networks has been explored in [11] where the role of temporal synchronization in associative memory feature of these networks has been shown.

A. MAIN CONTRIBUTIONS

This paper studies immediate free recall of a biologically relevant non-spiking WM model in [5]. We assume that the WM network has been trained to learn an information set composed of several memory patterns. The model assumes that all encoded patterns are orthogonal, i.e., each minicolumn is associated with only one of the patterns, and their couplings are governed by a fast Bayesian-Hebbian plasticity rule (see Section II). We assume that the network has learned a few patterns, the coupling weights and biases are constant during a very brief post-learning interval, and study the post-learning dynamics. Our contributions are formulating free recall of WM as a cluster synchronization problem, and characterizing bounds on coupling and bias values for achieving this collective behaviour. Cluster synchronization governing the free recall process consists of two main mechanisms: simultaneous activities of minicolumns representing each of the encoded patterns, and time-divided activities of minicolumns representing different patterns. First, for a network composed of H hyprcolumns and M minicolumns, using the concept of uniform and ultimate boundedness, we characterize coupling conditions under which the relative states of all minicolumns representing an arbitrary pattern are bounded and asymptotically confined within a compact set. We then show that in fact the relative state dynamics of each two minicolumns, belonging to the same pattern, is input-to-state stable, and continue by proving exponential stability of the synchronized solution for the network (Theorem 1). Second, we prove that assigning non-identical bias values for minicolumns corresponding to different patterns prevents simultaneous reactivation of the encoded patterns (Theorem 2). We then give sufficient conditions for achieving cluster synchronization in a $H \times M$ network (Corollary 1). Third, assuming that cluster synchronization is achieved for a $H \times 2$ network, we study the behaviour of the synchronized network to derive conditions under which both of the clusters, i.e., encoded patterns, are activated in turn (Theorem 3). Specifically, we compare the behaviour of two $H \times 2$ networks with identical and non-identical coupling weights and bias values. We obtain bounds on couplings and bias values under which both encoded patterns are recalled. Our analysis shows that having non-identical couplings and bias values for different patterns enhance free recall. Numerical simulations are given to validate our theoretical analysis.

To the best of our knowledge, revealing the mechanisms of WM free recall in the view of synchronization, as well as applications of nonlinear stability theory for this purpose have not been addressed in the literature. A preliminary result considering a network of H hypercolumns, each composed of 2 minicolumns, communicating over an all-to-all homogeneous network, i.e., equal coupling weights and bias terms, was presented in our conference paper [32]. Comparably, this article elaborates on the selected model, its biological relevance, considers a $H \times M$ heterogeneous network over a regular graph, and uses a new set of techniques, leading to new results and insights.

The rest of this article is organized as follows. Section II reviews the model, explains its biological justifications, presents the problem formulation, and gives the definition of free recall in view of synchronization. The analyses for synchronization

are presented in Section III. Section IV presents simulation results, and Section V summarizes the concluding remarks.

B. PRELIMINARIES

Here, we review some preliminaries from the nonlinear systems and stability theory [33].

Definition 1 (Definition 4.6 [33]): The solutions of $\dot{x} = f(x, t)$ are: (1) uniformly ultimately bounded with ultimate bound b if there exist positive constants b and c, independent of $t_0 \geq 0$, and for every $a \in (0, c)$, there is $T = T(a, b) > 0$, independent of t_0, such that

$$\|x(t_0)\| \leq a \Rightarrow \|x(t)\| \leq b, \forall t \geq t_0 + T,$$

(2) globally uniformly ultimately bounded if the above holds for arbitrarily large a.

Lemma 1 (Lemma 4.6 [33]): Suppose $f(t, x, u)$ is continuously differentiable and globally Lipschitz in (x, u), uniformly in t. If the unforced system $\dot{x} = f(t, x, 0)$ has a globally exponentially stable equilibrium point at the origin $x = 0$, then the system $\dot{x} = f(t, x, u)$ is input-to-state stable.

Lemma 2 (Lemma 2.1 [33]). Poincaré-Bendixson Criterion: For $\dot{x} = f(x)$, let $\mathcal{M}$ be a closed bounded subset of the plane such that

- $\mathcal{M}$ contains no equilibrium points, or contains only one equilibrium point such that the Jacobian matrix $\delta f / \delta x$ at this point has eigenvalues with positive real parts (Hence, the equilibrium point is unstable focus or unstable node),
- Every trajectory starting in $\mathcal{M}$ stays in $\mathcal{M}$ for all future time.

Then $\mathcal{M}$ contains a periodic orbit of $\dot{x} = f(x)$.

II. MODEL AND PROBLEM FORMULATION

We consider an abstract model of working memory based on a simplification of a (non-spiking) attractor neural network originated from a modular recurrent neural network model [5]. This network is composed of H modules, hypercolumns, each of which consists of M elementary units, (minicolumns). In this model, each memory item, pattern, is encoded by a mechanism into H attributes (represented by hypercolumns) each of which is composed of M intervals (represented by minicolumns). Fig. 1 shows a configuration of a network composed of 3 hypercolumns, each containing 3 minicolumns.

The dynamics of each minicolumn s_{ij} obeys [5]

$$\tau_m \dot{s}_{ij} = \left(\sum_{\substack{r=1 \\ r \neq i}}^{H} \sum_{l=1}^{M} \omega_{rl,ij} o_{rl} \right) - a_{ij} - s_{ij} + \beta_{ij} + I_{ij} \quad (1)$$

$$\tau_a \dot{a}_{ij} = g_a o_{ij} - a_{ij} \quad (2)$$

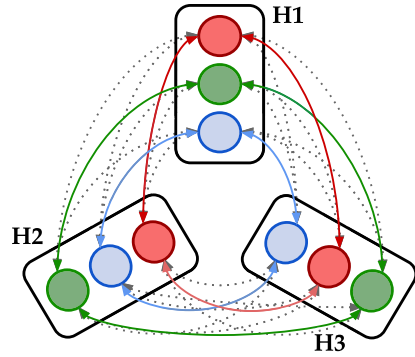$$o_{ij} = \frac{e^{s_{ij}}}{\sum_{h=1}^{M} e^{s_{ih}}}, \quad (3)$$

FIGURE 1. Network of three hypercolumns, each composed of three minicolumns, depicted as circles. Memory patterns are color-coded in blue, green, and red. Each minicolumn is positively (excitatory) coupled to the minicolumns in other hypercolumns if their activities are correlated, depicted using colored lines. Negative (inhibitory) couplings are shown as grey, dashed lines.

where $s_{ij} \in \mathbb{R}$, $a_{ij} \in \mathbb{R}$, $o_{ij} \in (0, 1)$ represent the activation, adaptation, and output of minicolumn ij, respectively. Constant parameters $\tau_m > 0$, $\tau_a > 0$, and $g_a > 0$ denote time constants and the adaptaion gain.

We assume that WM network has been trained to learn an information set composed of several patterns using an exogenous input I_{ij}. In the learning phase [5] (which is outside the scope of our paper), the weight $\omega_{rl,ij}$ and bias value β_{ij} are determined in terms of estimated probabilities of pre-synaptic activation p_{rl}, post-synaptic activation p_{ij}, and co-activation p as

$$w_{rl,ij} = \log_\varepsilon \frac{\hat{p}\hat{p}_{rl,ij}}{\hat{p}_{rl}\hat{p}_{ij}}, \quad \beta_{ij} = g_\beta \log_\varepsilon \hat{p}_{ij}, \quad (4)$$

where g_β is a gain parameter, ε is a low cut off value to avoid $\log(0)$ and $\log_\varepsilon x = \log \max(\varepsilon, x)$. It is assumed that for a very short time period the weights and bias values, $\omega_{rl,ij}$ and β_{ij} respectively, will stay constant. Moreover, the external current I_{ij}, which is used for training, is set to zero during the free recall process. That is, during the recall, it holds that,

$$\dot{\omega}_{rl,ij} = 0, \ \dot{\beta}_{ij} = 0, \ I_{ij} = 0. \quad (5)$$

In this paper, we study the post-leaning dynamics, i.e., the model represented by (1)–(3) subject to the modifications in (5). Following the learning rule in (4), we assume that the minicolumns whose activities are correlated are connected with positive (excitatory) coupling, while the minicolumns whose activities are un-correlated are connected with negative (inhibitory) coupling as described in Assumption 3.

We here introduce two terms: *homogenous* and *hetergenous* networks. The former refers to a network where all coupling weights are equal. Then, based on the Bayesian-Hebbian learning rule (4), all bias values are also equal. The latter, a heterogeneous network, models a more general setting by assuming that minicolumns corresponding to different patterns are coupled with different coupling weights, and bias values.

A. ASSUMPTIONS AND BIOLOGICAL RATIONALE OF THE MODEL

Here, we describe the notions of *hypercolumn and* minicolumn based on some biological justifications.

Minicolumns have been identified as the basic unit of the neocortex: narrow chains of 80 to 250 neurons extending vertically across cellular layers [34]. The influential research of Hubel & Wiesel [35] describes the hypercolumn as a complete set of minicolumns working together as a small machine [36]. From a morphological perspective, it is a set of neighboring cortical minicolumns, competing dynamically via lateral inhibition and exhibiting an orderly progression of their respective parameters [37].

Following [38], we consider orthogonal encoding, that is, no minicolumns are shared between distinct patterns. Orthogonality between memory patterns serves as a useful abstraction of the very sparse activity in higher order cortex [40], for instance in studying storing many random binary patterns in cortical associative memory [41]. Besides, sparse orthogonal encoding is considered a key element in hippocampal neural activity impacting the episodic memory capacity important in navigation. Examples include experimental findings confirming sparse orthogonal patterns in mice's retrosplenial cortex involved in spatial navigation [42].

Assumption 1: The stored patterns are orthogonal. That is, each memory pattern engages only one minicolumn in every hypercolumn.

In the rest of the paper, without loss of generality, it is assumed that each pattern ξ_j, is represented by the activation of minicolumn j of each hypercolumn i.

B. PROBLEM FORMULATION

We now provide a definition for WM free recall using the notion of network synchronization. Given the post-learning WM model in (1)–(5), under the assumptions in Section II-A, the encoded patterns are recalled in an arbitrary fashion, provided that the three following properties are satisfied:

P1 the activities of minicolumns representing pattern ξ_j are synchronized, i.e., their corresponding error dynamic is exponentially stable,

P2 the activities of minicolumns representing different patterns do not achieve exact synchronization,

P3 the synchronized minicolumns in different clusters alter the role of being the winner, i.e., having the maximum level of activity, in their corresponding hypercolumns.

The main goal of this paper is to derive sufficient conditions, in the form of lower and upper bounds, on the coupling weights and bias values such that the solutions of the WM network model under analysis satisfy properties P1–P3.

Assumption 2: The underlying graph of the network with the node dynamics in (1)–(3) is connected and undirected, i.e., $\omega_{rl,ij} = \omega_{ij,rl}$. The graph connecting all minicolumns of each pattern ξ_j is regular, i.e., all nodes in that graph have the same degree.

III. ANALYSIS: SYNCHRONIZATION FOR FREE RECALL

This section presents stability analysis for the network modeled in (1)–(5). First, we derive sufficient conditions for the synchronization of minicolumns representing a pattern, e.g. ξ_j, in a $H \times M$ network (property P1). We then continue by deriving sufficient conditions under which M synchronized patterns form M separate clusters. The latter guarantees property P2, and prevents simultaneous reactivation of two of the encoded patterns. Last, we focus on a $H \times 2$ network, and we assume that the minicolumns of each of the two patterns are synchronized and study conditions for their recall. Based on properties P2 and P3, the recall translates to alternating possession of the maximum level of activation, being the winner, between minicolumns representing pattern ξ_1 with those representing pattern ξ_2. To this aim, we study conditions for existence of a limit cycle behaviour for the two minicolumns within a hypercolumn in both homogeneous and heterogeneous networks.

Assumption 3: For the network with the node dynamics in (1)–(3), the absolute value of the coupling weight of each of the inhibitory (negative) connections is equal to $\underline{\omega} > 0$. All minicolumns representing pattern ξ_j are coupled with the weight $\tilde{\omega}_j > 0$, and the bias value for their dynamics is equal to $\tilde{\beta}_j > 0$.

Before we start with the analysis, for the brevity of notation, define the relative states $s_{rl,ij} = s_{rl} - s_{ij}$, $a_{rl,ij} = a_{rl} - a_{ij}$, $o_{rl,ij} = o_{rl} - o_{ij}$, and $\tilde{\beta}_{j\ell} = \tilde{\beta}_j - \tilde{\beta}_\ell$.

Lemma 3: Consider a heterogeneous network with the node dynamics in (1)–(3) under Assumptions 1–3. The dynamics corresponding to the relative state of each two minicolumns representing pattern ξ_j obey

$$\tau_m \, \dot{s}_{ij,kj} = - \left(\underline{\omega} + \tilde{\omega}_j \right) o_{ij,kj} - a_{ij,kj} - s_{ij,kj}, \quad (6)$$

$$\tau_a \, \dot{a}_{ij,kj} = - a_{ij,kj} + g_a o_{ij,kj}. \quad (7)$$

Theorem 1: Consider a heterogeneous network with node dynamics in (1)–(3) under Assumptions 1–3. The relative state dynamics corresponding to each two minicolumns representing pattern ξ_j, given in (6) and (7):

1) are globally uniformly and ultimately bounded in the set

$$S_u = \{ s_{ij,kj} \in \mathbb{R}, a_{ij,kj} \in \mathbb{R} : |a_{ij,kj}| \leq 2g_a,$$
$$|s_{ij,kj}| \leq 2(\underline{\omega} + \tilde{\omega}_j); \forall i, k \in \{1, \ldots H\},$$
$$\text{and } j, \ell \in \{1, \ldots M\}, \quad (8)$$

2) are input-to-state stable,

3) in S_u, possess an exponentially stable equilibrium at the origin.

REACTIVATION OF MULTIPLE PATTERNS

Theorem 1 proves achieving synchronization of minicolumns representing pattern ξ_j. However, the recall of multiple encoded patterns requires satisfaction of properties P2 and P3, as well. In what follows, we first consider a $H \times M$ network, and prove that property P2 is satisfied if the bias values corresponding to different patterns are non-identical. As mentioned

before, based on the learning rule in (4), the latter requires that coupling weights are different for different encoded patterns. We then combine the results of Theorems 1 and 2, and present sufficient conditions to have cluster synchronization for a $H \times M$ network.

Theorem 2: Consider a heterogeneous network with node dynamics in (1)–(3) under Assumptions 1–3. The relative state dynamics, $\dot{s}_{ij,k\ell}(t)$, corresponding to two minicolumns ij and $k\ell$ representing two different patterns ξ_j and ξ_ℓ do not possess an equilibrium at the origin, provided that $\min |\tilde{\omega}_j - \tilde{\omega}_\ell| < \frac{\max |\tilde{\beta}_j - \tilde{\beta}_\ell|}{\eta}$ and $\tilde{\beta}_j \neq \tilde{\beta}_\ell, \forall j, \ell$, with $\eta = H - 1$.

Corollary 1: Consider a heterogeneous network with node dynamics in (1)–(3) under Assumptions 1 and 2. Sufficient conditions guaranteeing that the solution to the network system converges to M locally and exponentially stable synchronous clusters are: (1) the coupling weights of all inhibitory weights are equal to $-\omega < 0$, (2) for the network corresponding to each pattern ξ_j, it holds that the positive coupling weights are equal to $\tilde{\omega}_j > 0$ and the bias values are equal to $\tilde{\beta}_j > 0$, (3) the coupling weights and bias values corresponding to the minicolumns of each two patterns ξ_j and ξ_ℓ satisfy $\tilde{\beta}_j \neq \tilde{\beta}_\ell$ and $\min |\tilde{\omega}_j - \tilde{\omega}_\ell| < \frac{\max |\tilde{\beta}_j - \tilde{\beta}_\ell|}{H-1}$.

Corollary 1 presents sufficient conditions for cluster synchronization, properties P1 and P2, of multiples encoded patterns in WM network modeled in (1)–(3). We now study conditions under which a $H \times 2$ network satisfies all properties P1-P3. We denote the two encoded patterns by ξ_1 and ξ_2, and assume that minicolumns representing pattern ξ_1 (ξ_2) are synchronized. In order to verify whether or not each of the two encoded patterns are recalled, we write the relative state dynamics between two minicolumns $i1$ and $i2$ in an arbitrary hypercolumn i, representing patterns ξ_1 and ξ_2. The relative dynamics across two patterns obey

$$\tau_m \dot{s}_{i1,i2} = v_1 o_{i1} - v_2 o_{i2} - a_{i1,i2} - s_{i1,i2} + \tilde{\beta}_{1,2}, \quad (9)$$

$$\tau_a \dot{a}_{i1,i2} = -a_{i1,i2} + g_a o_{i1,i2}, \quad (10)$$

where $v_1 = \eta(\omega + \tilde{\omega}_1)$, $v_2 = \eta(\omega + \tilde{\omega}_2)$, $\tilde{\beta}_{1,2} = \tilde{\beta}_1 - \tilde{\beta}_2$, and $\eta = H - 1$. Notice that based on the definition of the output o_{ij} in (3), i.e., $o_{ij} = \frac{e^{s_{ij}}}{\sum_h e^{s_{ih}}}$, we can write

$$o_{i1} = \frac{e^{s_{i1}}}{e^{s_{i1}} + e^{s_{i2}}} = \frac{e^{s_{i1,i2}}}{1 + e^{s_{i1,i2}}} \quad (11)$$

$$o_{i2} = \frac{e^{s_{i2}}}{e^{s_{i1}} + e^{s_{i2}}} = \frac{e^{-s_{i1,i2}}}{1 + e^{-s_{i1,i2}}} \quad (12)$$

$$o_{i1,i2} = \frac{e^{s_{i1}} - e^{s_{i2}}}{e^{s_{i1}} + e^{s_{i2}}} = \frac{e^{s_{i1,i2}} - 1}{1 + e^{s_{i1,i2}}}. \quad (13)$$

Replacing the above in (9) and (10), gives

$$\tau_m \dot{s}_{i1,i2} = F(s_{i1,i2}) - a_{i1,i2} - s_{i1,i2} + \tilde{\beta}_{1,2}, \quad (14)$$

$$\tau_a \dot{a}_{i1,i2} = -a_{i1,i2} + g_a G(s_{i1,i2}), \quad (15)$$

where $F(s_{i1,i2}) = v_1 o_{i1} - v_2 o_{i2}$ and $G(s_{i1,i2}) = o_{i1,i2}$. In what follows, we study the properties of the equilibria of the system

in (14) and (15), considering both homogeneous and heterogeneous networks (see Section II).

Theorem 3: Consider a synchronized network of H hypercolumns and $M = 2$ minicolumns with the relative dynamics in (14) and (15). The set

$$\mathcal{M} = \Big\{ s_{i1,i2} \in \mathbb{R}, a_{i1,i2} \in \mathbb{R} : |a_{i1,i2}| \leq g_a,$$

$$|s_{i1,i2}| \leq \max \{|\gamma_1|, |\gamma_2|, \frac{\tau_a}{\tau_m} g_a\} \Big\} \quad (16)$$

where $\gamma_1 = v_2 + g_a + \tilde{\beta}_{1,2}$, and $\gamma_2 = v_1 + g_a + \tilde{\beta}_{1,2}$, with $v_1 = \eta(\omega + \tilde{\omega}_1)$, $v_2 = \eta(\omega + \tilde{\omega}_2)$, and $\eta = H - 1$, is forward invariant for the system. Moreover, the system in (14) and (15) has at most three equilibria, unless both $v_1 - v_2 = -2\tilde{\beta}_{1,2}$ and $v_1 + v_2 = 2g_a + 4$ hold. Furthermore,

1) for the homogeneous network: if $v_1 = v_2 > 2(1 + \frac{\tau_m}{\tau_a})$ and $g_a > \frac{2\tau_m}{\tau_a}$, the relative state dynamics in (14) and (15) exhibit a limit cycle behaviour,

2) for the heterogeneous network: given the network parameters, $v_1, v_2, \tau_m, \tau_a, g_a$, there exists $\tilde{\beta}_{1,2} \neq 0$, such that the relative state dynamics in (14) and (15) have a unique equilibrium, $s^*_{1,2}$, in $\mathcal{M}$. Moreover, if the Jacobian Matrix J (17) at the equilibrium, $s^*_{1,2}$, has at least one unstable eigenvalue, then system (14) and (15) exhibits a limit cycle behaviour.

$$J = \begin{bmatrix} \frac{v_1}{\tau_m} \frac{e^{s^*_{1,2}}}{(e^{s^*_{1,2}}+1)^2} + \frac{v_2}{\tau_m} \frac{e^{-s^*_{1,2}}}{(e^{-s^*_{1,2}}+1)^2} - \frac{1}{\tau_m} & \frac{-1}{\tau_m} \\ \frac{g_a}{\tau_a} \frac{2e^{s^*_{1,2}}}{(e^{s^*_{1,2}}+1)^2} & -\frac{1}{\tau_a} \end{bmatrix} \quad (17)$$

Based on the above result, non-identical bias values change the location of the equilibrium of the system (14), create a unique equilibrium which can be made unstable by tuning the coupling weights and other network's parametres. Based on the results of Theorem 3, we expect that conditions in Theorem 2 together with Theorem 3 guarantee the recall of at least two patterns for a $H \times M$ network.

A. DISCUSSION: LIMITATIONS AND FUTURE WORK

The biological plausibility of the model, originally derived to explain human behavioral experiments in WM immediate free recall tasks [23], is discussed in Section II-A and references therein. This article studies a non-spiking model. Among important future directions are the analysis of more-detailed network dynamics, for instance, spiking models as well as models derived from measured data in WM tasks.

This article assumes that the coupling weights, learned during the encoding phase, stay constant during the recall task. This assumption is justified owing to the immediate aspect of the recall task considered in this article. In our mathematical analysis, we have assumed that the network topology is a regular graph. We envision that the relaxation of the aforementioned assumption to the general undirected graphs, which serves as a future work, will not affect the main message

TABLE 1. Network parameters.

Parameter	Description	Value
H	# hypercolumns	4
M	# minicolumns	3
$[\tau_m, \tau_a]$	Time constants	$[0.05 \text{ s}, 2.70 \text{ s}]$
g_a	Adaptation gain	97.0

of this article, i.e., cluster synchronization as the WM's free recall mechanism.

This article has studied two possible learning settings: homogeneous and heterogeneous networks. For the latter, we have assumed that the coupling weights and bias terms corresponding to each encoded pattern are equal. The latter assumption is inspired by the Bayesian-Hebbian learning rule (4), which depends on the probability of simultaneous activation of minicolumns corresponding to a pattern. We have approximated equal coupling weights for excitatory connections between minicolumns corresponding to an encoded pattern. However, analysis of a more general heterogeneous network, in particular in the presence of uncertainties [13], is an interesting future direction.

An important validation approach for the model and our analysis is through experiments. A closely related, but spiking, WM model capable of capturing the temporal dynamics of a realistic attractor memory network based on experimental data has been discussed in [43]. For the model of this paper, experimental validation is an open future direction.

In a nutshell, our mathematical analysis sheds light on the WM model's immediate free recall mechanism. We have shown that the adaptation dynamics, size and sign of the weights, bias values, and properties of the coupling function determine the network's behavior. Stability analysis reveals which assumptions and dynamics are important in achieving a particular behavior. This understanding opens a future avenue for tuning and control.

IV. SIMULATION RESULTS

To validate our main theoretical results, we consider a network of $H = 4$ hypercolumns, each composed by $M = 3$ minicolumns. The minicolumns representing each pattern are interconnected with an all-to-all topology. The model parameters are provided in Table 1. The time constants and the adaptation gain are according to [5].

In Fig. 2 the results of Theorems 1 and 2 are verified. The coupling weights are set to $\tilde{\omega}_1 = 1.3$, $\tilde{\omega}_2 = 1.1$, $\tilde{\omega}_3 = 1.2$, $\underline{\omega} = 0.2$ and the bias values $\tilde{\beta}_1 = 110$, $\tilde{\beta}_2 = 112$, $\tilde{\beta}_3 = 115$. It is assumed that three distinct patterns have been presented to the network during the training phase. The simulations in Fig. 2(a)–(c) show that the positively correlated minicolumns corresponding to each of the three patterns asymptotically synchronize. In Fig. 2(d) we report the time-evolution of states of a single hypercolumn: every minicolumn is activated in turn. The winner-take-all behaviour is clear at the output level with each minicolumn having the highest level of activation for a short period of time (≈ 150 ms) after which its level of

(a) Minicolumn 1   (b) Minicolumn 2
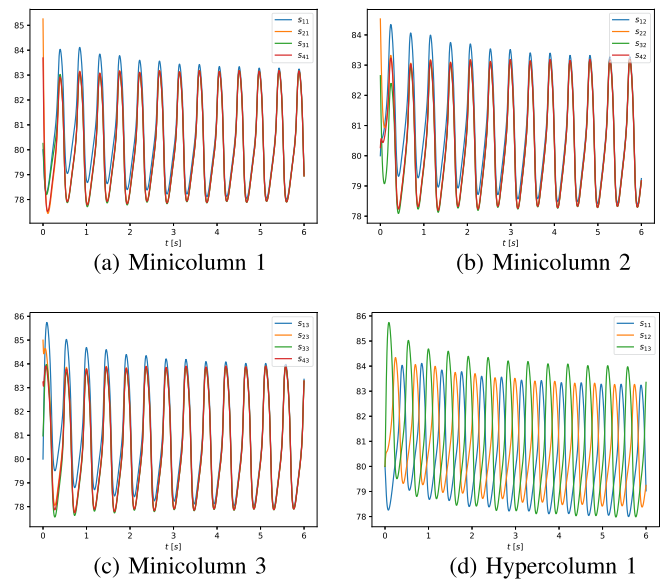
(c) Minicolumn 3   (d) Hypercolumn 1

FIGURE 2. Simulation results of a heterogenous network of $H = 4$ hypercolumns, each composed by $M = 3$ minicolumns. Each plot (a), (b), (c) corresponds to all the correlated minicolumns belonging to the same pattern across all the hypercolumns. Plot (d) shows the time-evolution of all minicolumns of one hypercolumn.
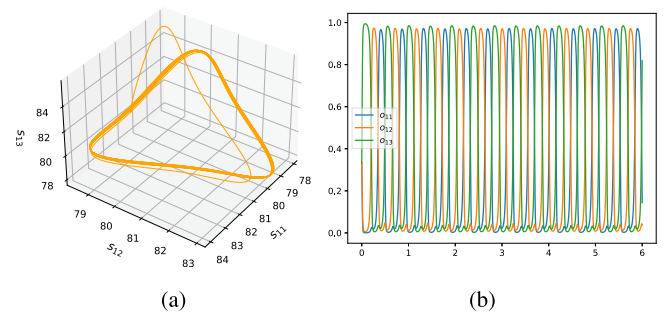

(a)   (b)

FIGURE 3. Simulation results of a heterogeneous network of four hypercolumns, each composed by three minicolumns. The left plot represents the limit cycle in the s-plane corresponding to the states of hypercolumn 1. The right plot represents the output, o_{1j}, corresponding to each minicolumn in hypercolumn 1.

activation decreases and the other minicolumns are activated (Fig. 3(b)). These simulations confirm that all the three different patterns are effectively stored in the network and recalled in the form of a limit cycle (Fig. 3(a)).

We now verify the results of Theorem 3 for a 4×2 network, assuming that two synchronized clusters have been formed, representing two patterns. Considering the dynamics in (14) and (15), define $y_1 = s_{i1,i2}$ and $y_2 = v_1 o_{i1} - v_2 o_{i2} - g_a o_{i1,i2} + \tilde{\beta}_{1,2}$. The equilibria of the system (14) and (15) are located at the intersections of y_1 and y_2. For the heterogeneous network, the pair (v_1, v_2) is set to $(11, 12)$, and $\tilde{\beta}_{1,2} = 6$. The unique equilibrium is at $s^*_{i1,i2} = 0.13$. Computation of eigenvalues of the Jacobian matrix in (17) gives $(-20.4, 64.7)$, hence an unstable equilibrium. Figs. 4(a) and 5(a) show the system's unique
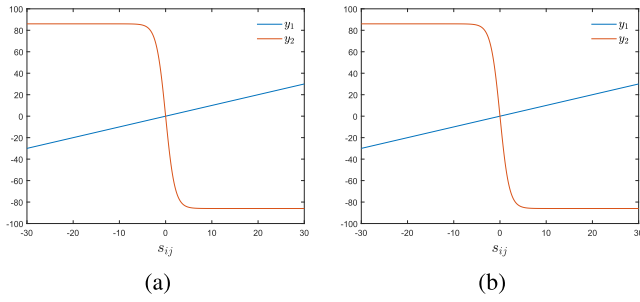
FIGURE 4. System in (14) and (15): The equilibria are at the intersection of y_1 and y_2. Plots in (a) and (b) illustrate the location of the equilibrium for the heterogeneous and homogeneous networks, respectively.
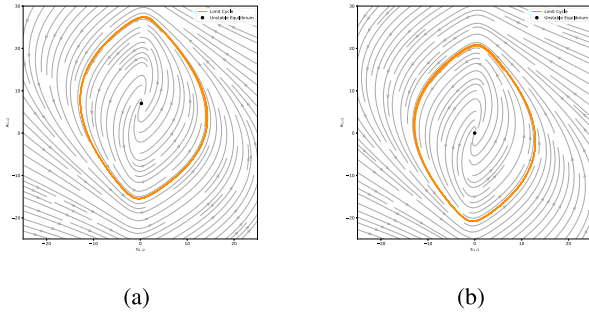

FIGURE 5. Phase portrait for dynamics in (14) and (15) for the heterogeneous (a), and homogeneous (b) networks.
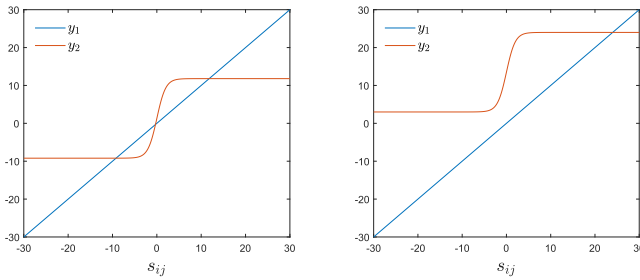

FIGURE 6. System in (14) and (15) for a heterogeneous network with three equilibria (left plot). By increasing the relative bias value $\beta_{1,2}$, a unique equilibrium is obtained (right plot).

equilibrium and its corresponding limit cycle. For a homogeneous network, we set $v = 11$ and $\tilde{\beta}_{1,2} = 0$. We verify the presence of a unique equilibrium in Fig. 4(b) and the resulting limit cycles in Fig. 5(b).

Fig. 6 further illustrates the effect of tuning the relative bias term $\tilde{\beta}_{1,2}$ on the location of the equilibrium point for the heterogenous network. For the system represented in the left figure, the parameter set $(v_1, v_2, g_a, \tilde{\beta}_{1,2})$ is set to $(11, 12, 1, 1.8)$. Since this setting gives three equilibria, $\tilde{\beta}_{1,2}$ is then increased to 14. The latter results in a unique equilibrium at $s^*_{i1,i2} = 24$. Computation of eigenvalues of the Jacobian matrix in (17) gives $(0, 3.13)$. Thus, we obtain an unstable equilibrium leading to a limit cycle behaviour.

V. CONCLUSION

We studied free recall of an abstract, yet biologically plausible, dynamic network model of working memory (WM). We have shown that WM's free recall is a cluster synchronization problem, i.e., simultaneous activities of minicolumns representing an encoded pattern, together with time-divided activities of minicolumns representing different patterns. Using nonlinear systems and stability theory, bounds on coupling and bias values were derived that guarantee M synchronized clusters for a $H \times M$ network. Furthermore, for a $H \times 2$ network, we studied cluster synchronization assuming both identical and non-identical coupling and bias values, and derived conditions under which both of the encoded patterns are recalled. The latter result provides insights on network tuning for the recall of multiple patterns in a $H \times M$ network. Our analysis also shows that having non-identical couplings and bias values for different patterns increases the chance of their free recall. Future avenues include extending the mathematical framework for the heterogeneous network, as well as establishing connections with experimental results. Further, the effects of uncertainties for the given model deserves investigations.

APPENDIX A
PROOF OF LEMMA 3

Proof: Recall (1), and (5). Computing $\dot{s}_{ij,kj}$ gives

$$\tau_m \dot{s}_{ij,kj} = \left(\sum_{\substack{p=1 \\ p \neq i}}^{H} \sum_{l=1}^{M} \omega_{pl,ij} o_{pl} \right) - \left(\sum_{\substack{p=1 \\ p \neq k}}^{H} \sum_{l=1}^{M} \omega_{pl,kj} o_{pl} \right)$$
$$- a_{ij,kj} - s_{ij,kj} + \beta_{ij} - \beta_{kj}. \qquad (18)$$

Based on Assumption 3, all inhibitory (negative) coupling has the same weight, and also all minicolumns representing pattern ξ_j are coupled with the weight $\tilde{\omega}_j$, also $\beta_{ij} = \beta_{kj} = \tilde{\beta}_j$. That is,

$$(\omega_{ij,rj} - \omega_{kj,rj}) o_{rj} = 0; \quad r \neq i, k, \qquad (19)$$

$$(-\omega_{ij,r\ell} + \omega_{kj,r\ell}) o_{r\ell} = 0; \quad r \neq i, k, \text{ and}, \ell \neq j, \qquad (20)$$

where $\omega_{ij,rj} = \omega_{kj,rj} = \tilde{\omega}_j$, and $\omega_{ij,r\ell} = \omega_{kj,r\ell} = -\underline{\omega}$. Considering the above modification to the dynamics of $\dot{s}_{ij,kj}$ gives

$$\tau_m \dot{s}_{ij,kj} = \underline{\omega} \left(\sum_{\substack{\ell=1 \\ \ell \neq j}}^{M} o_{i\ell} - \sum_{\substack{\ell=1 \\ \ell \neq j}}^{M} o_{k\ell} \right) - \tilde{\omega}_j o_{ij,kj} - a_{ij,kj} - s_{ij,kj},$$
$$\qquad (21)$$

$$\tau_a \dot{a}_{ij,kj} = -a_{ij,kj} + g_a o_{ij,kj}. \qquad (22)$$

Based on the definition of the output in (3), we have

$$\sum_{\substack{\ell=1 \\ \ell \neq j}}^{M} o_{il} = 1 - o_{ij}, \quad \sum_{\substack{\ell=1 \\ \ell \neq j}}^{M} o_{k\ell} = 1 - o_{kj}. \quad (23)$$

Substituting (23) in (21) gives (6), which completes the proof. ∎

APPENDIX B
PROOF OF THEOREM 1

Proof: Following Definition 4.6, and Theorem 4.18 in [33], define $V(s, a)$, a continuous, positive, and radially unbounded function, as follows:

$$V(s, a) = \sum_{i=1}^{H} \sum_{j=1}^{M} \frac{\tau_m}{2} (s_{ij,kj})^2 + \sum_{i=1}^{H} \sum_{j=1}^{M} \frac{\tau_a}{2} (a_{ij,kj})^2. \quad (24)$$

First, we rewrite $V(s, a) = \sum_{i,k,j} V_{ij,kj}$, where

$$V_{ij,kj} = \frac{\tau_m}{2} s_{ij,kj}^2 + \frac{\tau_a}{2} a_{ij,kj}^2. \quad (25)$$

Calculating $\dot{V}_{ij,kj} = \tau_m \dot{s}_{ij,kj} s_{ij,kj} + \tau_a \dot{a}_{ij,kj} a_{ij,kj}$, gives

$$\dot{V}_{ij,kj} = -s_{ij,kj}^2 - (\underline{\omega} + \tilde{\omega}_j) o_{ij,kj} s_{ij,kj}$$
$$- s_{ij,kj} a_{ij,kj} + g_a a_{ij,kj} o_{ij,kj} - a_{ij,kj}^2. \quad (26)$$

Since $o_{ij} \in (0, 1)$, then $-1 < o_{ij,kj} < 1$, and it holds that

$$\dot{V}_{ij,kj} \leq -\frac{1}{2}(s_{ij,kj} + a_{ij,kj})^2 - \frac{1}{2}(s_{ij,kj} \pm \hat{\omega})^2 + \frac{1}{2}\hat{\omega}^2$$
$$- \frac{1}{2}(a_{ij,kj} \pm g_a)^2 + \frac{1}{2}g_a^2, \quad (27)$$

with $\hat{\omega} = \underline{\omega} + \tilde{\omega}_j$. Therefore, provided that $|s_{ij,kj}| > 2\hat{\omega}$ and $|a_{ij,kj}| > 2g_a$ are satisfied, $\dot{V}_{ij,kj} < 0$ holds. Since $\dot{V}_{ij,kj} < 0$, then $\dot{V}(s, a) \leq 0$ is also satisfied. Hence the solutions to (6) and (7) are ultimately bounded w.r.t to S_u. (2) System (6) and (7) can be written as $\dot{x} = Ax + Bu$, with $u = o_{ij,kj}$, $|u| < 1$. If $u = 0$, the origin is globally exponentially stable, and since u is bounded then ISS stability is concluded (see Lemma 4.18 [33]). The solutions' bound is exactly as the one characterized above.

(3) Considering the dynamics in (6), the nonlinear term is $o_{ij,kj} = o_{ij} - o_{kj}$, which takes values in the interval $(-1, 1)$. In fact, it holds that $|o_{ij,kj}| \leq |\tanh(s_{ij,kj})|$. Replacing $\tanh(s_{ij,kj})$ in (6) and (7), and linearizing the system around the equilibrium $(0, 0)$, gives

$$\dot{\tilde{x}} = \begin{bmatrix} -(\frac{1+\underline{\omega}+\tilde{\omega}_j}{\tau_m}) & -\frac{1}{\tau_m} \\ \frac{g_a}{\tau_a} & -\frac{1}{\tau_a} \end{bmatrix} \tilde{x}, \quad (28)$$

where $\tilde{x} = \begin{bmatrix} s_{ij,kj}, a_{ij,kj} \end{bmatrix}$. The linearized system is exponentially stable, if the matrix in (28) is Hurwitz (see Theorem 4.13 [33]), which is guaranteed since network's parameters are positive. ∎

APPENDIX C
PROOF OF THEOREM 2

Proof: First, write the relative state dynamics between two minicoloumns ij and $k\ell$, in hypercolumns i and k, representing patterns ξ_j and ξ_ℓ, we have

$$\tau_m \dot{s}_{ij,k\ell} = (\underline{\omega} + \tilde{\omega}_j) \sum_i o_{ij} - (\underline{\omega} + \tilde{\omega}_\ell) \sum_k o_{k\ell}$$
$$- a_{ij,k\ell} - s_{ij,k\ell} + \tilde{\beta}_{j\ell}, \quad (29)$$
$$\tau_a \dot{a}_{ij,k\ell} = -a_{ij,k\ell} + g_a o_{ij,k\ell}. \quad (30)$$

Now, assume exact synchronization for minicolumns in pattern ξ_j, as well as for pattern ξ_ℓ. The aim is to show that if the minicolumns of one of these patterns are the winners at time t, they are the only winners. That is, no two patterns are recalled simultaneously. Assuming exact synchronization for patterns j and ℓ, (29) and (30) are reduced to

$$\tau_m \dot{s}_{ij,k\ell} = \eta[(\underline{\omega} + \tilde{\omega}_j) o_{ij} - (\underline{\omega} + \tilde{\omega}_\ell) o_{k\ell}] - a_{ij,k\ell} - s_{ij,k\ell} + \tilde{\beta}_{i\ell} \quad (31)$$
$$\tau_a \dot{a}_{ij,k\ell} = -a_{ij,k\ell} + g_a o_{ij,k\ell}, \quad (32)$$

where $\eta = H - 1$. In case a simultaneous reactivation of these two patterns occur, $o_{ij,k\ell} = 0$ should hold. Thus, the equilibrium of the second dynamics is $a_{ij,k\ell}^* = 0$. Now, since $o_{ij} = o_{k\ell}$, if it holds that $\eta|\tilde{\omega}_j - \tilde{\omega}_\ell| \neq |\tilde{\beta}_{i\ell}|$, the equilibrium for the first dynamics cannot be the origin. Thus, $s_{ij,k\ell}$ is stabilized at a non-zero value. Since j and ℓ are arbitrary, this holds for any two minicolumns in different patterns, including each two minicolumns in the same hypercolumn (due to orthogonality). Thus, $s_{ij,k\ell} \neq 0$ gives $o_{ij,k\ell} \neq 0$, which contradicts the assumption of $o_{ij,k\ell} = 0$, and completes the proof. ∎

APPENDIX D
PROOF OF THEOREM 3

Proof: The first part, we show that the trajectories converge to a forward invariant set. Consider the Lyapunov function

$$V(s_{i1,i2}, a_{i1,i2}) = \frac{1}{2}(\tau_m s_{i1,i2} - \tau_a a_{i1,i2})^2 + \frac{\tau_a}{2} a_{i1,i2}^2.$$

Calculating the derivative of V gives

$$\dot{V} = -\underbrace{(\tau_m s_{i1,i2} - \tau_a a_{i1,i2})}_{I_1} \underbrace{(s_{i1,i2} - \gamma - \tilde{\beta}_{1,2})}_{I_2}$$
$$- a_{i1,i2} \underbrace{(a_{i1,i2} - g_a o_{i1,i2})}_{I_3}, \quad (33)$$

with $\gamma = v_1 o_{i1} - v_2 o_{i2} - g_a o_{i1,i2}$. Since $-1 < o_{i1,i2} < 1$, the sign of I_3 in (33) is equal to sign$(a_{i1,i2})$ if $|a_{i1,i2}| > g_a$ holds. Also, since $0 < o_{i1} < 1, 0 < o_{i2} < 1$, we have $-v_2 - g_a < \gamma < v_1 + g_a$. The latter, considering I_2, gives a bound on $|s_{i1,i2}|$. Replacing $|a_{i1,i2}| \leq g_a$ in I_1, gives another bound on $|s_{i1,i2}|$. Combining the bounds from I_1 and I_2 gives set $\mathcal{M}$ in (16), which is forward invariant [33].

We now discuss the existence and stability of equilibria within this set. Consider the dynamics in (14) and (15). The

right hand side is the sum of two functions $y_1 = s_{i1,i2}$ and $y_2 = v_1 o_{i1} - v_2 o_{i2} - g_a o_{i1,i2} + \tilde{\beta}_{1,2}$. The system may have infinite number of equilibria if it has an equilibrium at the origin, i.e., $y_1(0) = y_2(0)$ and it holds that $\frac{dy_1}{ds_{i1,i2}}(0) = \frac{dy_2}{ds_{i1,i2}}(0) = 1$. These conditions are satisfied if both $v_1 - v_2 = -2\tilde{\beta}_{1,2}$ and $v_1 + v_2 - 2g_a = 4$ hold. Otherwise, since y_2 is a monotone function (see below for the proof) and has two bounded limits, similar to a $tanh(\cdot)$ function, system (14) and (15) has at most three equilibria.

Now we focus on the conditions which gives a limit cycle (LC) as an attractor for the system for two cases: homogeneous and heterogeneous networks. For the homogeneous network, we refer to our result in [32]. In brief, the given conditions in part (1) of Theorem 3 guarantee that the linearized system has a conjugate pair of imaginary eigenvalues, leading to a Hopf bifurcation.

For the heterogeneous network, we first argue that for a given set of network parameters, there exits $\tilde{\beta}_{1,2} \neq 0$ such that the system has only one equilibrium. Calculating the derivative of function $y_2 = v_1 o_{i1} - v_2 o_{i2} - g_a o_{i1,i2} + \tilde{\beta}_{1,2}$, we obtain

$$\frac{dy_2}{ds_{i1,i2}} = \frac{e^{s_{i1,i2}}}{(1 + e^{s_{i1,i2}})^2}(v_1 - 2g_a + v_2). \tag{34}$$

Given v_1, v_2, g_a, the above derivative is either positive or negative, hence y_2 is a monotone function, which crosses the line $y_1 = s_{i1,i2}$ at most at three points. Thus, there exists a bias term $\tilde{\beta}_{1,2}$, such that the number of equilibria is reduced to one by moving y_2. Now, calculating the Jacobian of the system (14) and (15) around this equilibrium point, gives the matrix in (17). Provided that at least one of the eigenvalues of the Jacobian Matrix J has a positive real part, based on the Poincaré-Bendixson Criterion [33], the system has a periodic solution, hence a limit cycle, which ends the proof. ∎

REFERENCES

[1] M. D'Esposito and B. Postle, "The cognitive neuroscience of working memory," *Annu. Rev. Psychol.*, vol. 66, pp. 115–142, 2015.

[2] A. Graves et al., "Hybrid computing using a neural network with dynamic external memory," *Nature*, vol. 538, no. 7626, pp. 471–476, 2016.

[3] A. D. Baddeley, "Working memory: Theories, models, and controversies," *Annu. Rev. Psychol.*, vol. 63, pp. 1–29, 2012.

[4] N. Unsworth and R. W. Engle, "The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory," *Psychol. Rev.*, vol. 114, no. 1, pp. 104–32, 2007.

[5] A. Lansner, P. Marklund, S. Sikström, and L. Nilsson, "Reactivation in working memory: An attractor network model of free recall," *PLoS One*, vol. 8, no. 8, 2013, Art. no. e73776.

[6] F. Fiebig, "Active memory processing: On multiple time-scalesin simulated cortical networks with Hebbian plasticity," Ph.D. dissertation, KTH Royal Inst. Technol., Stockholm, Sweden, 2018.

[7] A. Lansner, "Associative memory models: From the cell-assembly theory to biophysically detailed cortex simulations," *Trends Neurosci.*, vol. 32, no. 3, pp. 178–186, 2009.

[8] M. Camperi and X.-J. Wang, "A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability," *J. Comput. Neurosci.*, vol. 5, pp. 383–405, 1998.

[9] D. J. Amit and N. J.-B. Brunel, "Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex," *Cereb. Cortex*, vol. 7, no. 3, pp. 237–52, 1997.

[10] M. Jafarian, X. Yi, M. Pirani, H. Sandberg, and K. H. Johansson, "Synchronization of Kuramoto oscillators in a bidirectional frequency-dependent tree network," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 4505–4510.

[11] E. M. Izhikevich, "Weakly pulse-coupled oscillators, FM interactions, synchronization, and oscillatory associative memory," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 508–526, May 1999.

[12] J. F. Mejías and X.-J. Wang, "Mechanisms of distributed working memory in a large-scale network of macaque neocortex," *eLife*, vol. 11, 2022, Art. no. e72136.

[13] M. Jafarian, M. H. Mamduhi, and K. H. Johansson, "Stochastic stability of discrete-time phase-coupled oscillators over uncertain and random networks," *IEEE Trans. Autom. Control*, early access, Aug. 07, 2023, doi: 10.1109/TAC.2023.3303102.

[14] M. Lundqvist, A. M. Bastos, and E. K. Miller, "Preservation and changes in oscillatory dynamics across the cortical hierarchy," *J. Cogn. Neurosci.*, vol. 32, no. 10, pp. 2024–2035, 2020.

[15] J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci.*, vol. 79, no. 8, pp. 2554–2558, 1982.

[16] V. Centorrino, F. Bullo, and G. Russo, "Contraction analysis of hopfield neural networks with Hebbian learning," in *Proc. IEEE 61st Conf. Decis. Control*, 2022, pp. 622–627.

[17] I. Kanter, "Potts-glass models of neural networks," *Phys. Rev. A*, vol. 37, no. 7, 1988, Art. no. 2739.

[18] C. E. Curtis and T. C. Sprague, "Persistent activity during working memory from front to back," *Front. Neural Circuits*, vol. 15, 2021, Art. no. 696060.

[19] A. Compte, N. Brunel, P. S. Goldman-Rakic, and X. J. Wang, "Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model," *Cereb. Cortex*, vol. 10, no. 9, pp. 910–923, 2000.

[20] P. S. Goldman-Rakic, "Architecture of the prefrontal cortex and the central executive," *Ann. New York Acad. Sci.*, vol. 769, pp. 71–83, 1995.

[21] E. Ghazizadeh and S. Ching, "Slow manifolds within network dynamics encode working memory efficiently and robustly," *PLoS Comput. Biol.*, vol. 17, 2021, Art. no. e1009366.

[22] F. Fiebig and A. Lansner, "A spiking working memory model based on Hebbian short-term potentiation," *J. Neurosci.*, vol. 37, no. 1, pp. 83–96, 2017.

[23] A. Sandberg, J. N. Tegnér, and A. Lansner, "A working memory model based on fast Hebbian learning," *Network*, vol. 14, no. 4, pp. 789–802, 2003.

[24] J. Barbosa et al., "Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory," *Nature Neurosci.*, vol. 23, pp. 1016–1024, 2020.

[25] V. Würzbauer, K. Lenk, and M. Jafarian, "Astrocytic gliotransmission as a pathway for stable stimulation of post-synaptic spiking: Implications for working memory," in *Proc. Int. Federation Autom. Control*, 2023.

[26] F. Roux and P. J. Uhlhaas, "Working memory and neural oscillations: Alpha–gamma versus theta–gamma codes for distinct WM information?," *Trends Cogn. Sci.*, vol. 18, pp. 16–25, 2014.

[27] H. A. ElShafei, Y. J. Zhou, and S. Haegens, "Shaping information processing: The role of oscillatory dynamics in a working memory task," *eNeuro*, Eneuro, 2022, 9.5: ENEURO. 0489-21.2022.

[28] G. Mongillo, O. Barak, and M. Tsodyks, "Synaptic theory of working memory," *Science*, vol. 319, pp. 1543–1546, 2008.

[29] J. Barbosa, V. Babushkin, A. Temudo, K.K. Sreenivasan, and A. Compte, "Across-area synchronization supports feature integration in a biophysical network model of working memory," *Front. Neural Circuits*, vol. 15, 2021, doi: 10.3389/fncir.2021.716965.

[30] M. van Gerven et al., "Decoding the memorization of individual stimuli with direct human brain recordings," *NeuroImage*, vol. 70, pp. 223–232, 2013.

[31] Y. Yuan, X. Pan, and R. Wang, "Biophysical mechanism of the interaction between default mode network and working memory network," *Cogn. Neurodynamics*, vol. 15, pp. 1101–1124, 2021.

[32] G. Villani, M. Jafarian, A. Lansner, and K. Johansson, "Analysis of free recall dynamics of an abstract working memory model," in *Proc. IEEE Amer. Control Conf.*, 2020, pp. 2562–2567.

[33] H. K. Khalil, *Nonlinear Systems*. Prentice-Hall, New Jersey, 1996.

[34] V. B. Mountcastle, "The columnar organization of the neocortex," *Brain, J. Neurol.*, vol. 120, no. 4, pp. 701–722, 1997.

[35] D. H. Hubel and T. N. Wiesel, "Functional architecture of macaque monkey visual cortex," *Proc. Roy. Soc. Lond. B*, vol. 198, 1-59, 1977.

[36] D. H. Hubel and T. N. Wiesel, "Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor," *J. Comp. Neurol.*, vol. 158, pp. 295–305, 1974.

[37] D. P. Buxhoeveden and M. F. Casanova, "The minicolumn hypothesis in neuroscience," *Brain, J. Neurol.*, vol. 125, no. 5, pp. 935–951, 2002.

[38] M. A. Petrovici et al., "Characterization and compensation of network-level anomalies in mixed-signal neuromorphic modeling platforms," *PLoS ONE*, vol. 9, 2014, Art. no. e108590.

[39] P. Krishnamurthy, G. Silberberg, and A. Lansner, "A cortical attractor network with Martinotti cells driven by facilitating synapses," *PLoS ONE*, vol. 7, 2012, Art. no. e30752.

[40] R. Q. Quiroga, G. Kreiman, C. Koch, and I. Fried, "Sparse but not 'grandmother-cell' coding in the medial temporal lobe," *Trends Cogn. Sci.*, vol. 12, pp. 87–91, 2008.

[41] C. Meli and A. Lansner, "A modular attractor associative memory with patchy connectivity and weight pruning," *Netw., Computation Neural Syst.*, vol. 24, pp. 129–150, 2013.

[42] D. Mao, S. Kandler, B. L. McNaughton, and V. Bonin, "Sparse orthogonal population representation of spatial context in the retrosplenial cortex," *Nature Commun.*, vol. 8, no. 1, 2017, Art. no. 243.

[43] F. Fiebig, P. Herman, and A. Lansner, "An indexing theory for working memory based on fast Hebbian plasticity," *eNeuro*, eNeuro 7:ENEURO.0374-19.2020.

MATIN JAFARIAN (Member, IEEE) received the B.Sc. degree in electrical engineering, majoring in systems and control from the K. N. Toosi University of Technology, Tehran, Iran, in 2003, the M.Sc. degree in electrical engineering, majoring in systems and control from the University of Twente, Enschede, The Netherlands, in 2010, and the Ph.D. degree in systems and control theory from ENTEG Institute, University of Groningen, The Netherlands, in 2015. She is currently an Assistant Professor with the Delft Center for Systems and Control, Delft University of Technology (TU Delft), The Netherlands, and a Research Fellow with the Donders Institute for Brain, Cognition, and Behavior, The Netherlands. She has held a Postdoctoral position with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. Her research interests include analysis and control of nonlinear and stochastic dynamic networks and brain memory. She was a Marie Skłodowska-Curie Fellow jointly with TU Delft and Donders Institute.

DAVID CHÁVEZ HUERTA received the B.Sc. degree in physics from BUAP University, Puebla, Mexico, and the M.Sc. degree in control and dynamical systems from the IPICYT Institute, Mexico. He is currently working toward the Ph.D. degree with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. His research interests include developing mathematical models and control-theory-based analyses for working memory networks.

GIANLUCA VILLANI received the B.Sc. degree in electrical and computer engineering from Politecnico di Torino, Turin, Italy, and the M.Sc. degree in electrical and computer engineering from the KTH Royal Institute of Technology, Stockholm, Sweden. He is currently working toward the Ph.D. degree with the the Edward S. Rogers Sr., Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. His research focuses on the analysis and control of networked systems, with a specific emphasis on multistability in biological systems.

ANDERS LANSNER received the M.Sc. degree in engineering chemistry from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1974, and the Ph.D. degree in computer science from KTH Royal Institute of Technology, in 1986. He is currently a research active Professor Emeritus with the KTH Royal Institute of Technology. After five years as a civil servant, he returned to KTH to study computer science. He was appointed as an Assistant Professor of computer science in 1987, Associate Professor in 1992, and Professor at KTH in 1999. Since 2002, he has held a Professor's chair in computer science with Stockholm University, Stockholm, Sweden, with affiliation and most research activities at KTH from 2015. His research interests include neuroinformatics, computational neuroscience, and brain-like computing. Since 2008, he has been a Member of the Royal Swedish Academy of Engineering Sciences (IVA), Division XII —Information Technology.

KARL H. JOHANSSON (Fellow, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in automatic control from Lund University, Lund, Sweden, in 1992 and 1997, respectively. He is currently a Swedish Research Council Distinguished Professor of electrical engineering and computer science with KTH Royal Institute of Technology, Stockholm, Sweden, where he is also the Director of Digital Futures. He has held visiting positions at UC Berkeley, Caltech, NTU among other institutions. He has coauthored more than 800 journal and conference papers in his research areas which include networked control systems and cyber-physical systems with applications in transportation, energy, and automation networks, and supervised almost 100 postdocs and Ph.D. students. He is the President of European Control Association and Member of IFAC Council, and has served on the IEEE Control Systems Society Board of Governors and the Swedish Scientific Council for Natural Sciences and Engineering Sciences. He was the recipient of several best paper awards and other distinctions from IEEE, IFAC, and ACM. He has been awarded Distinguished Professor by Swedish Research Council, Wallenberg Scholar with the Knut and Alice Wallenberg Foundation, Future Research Leader from the Swedish Foundation for Strategic Research, the triennial IFAC Young Author Prize, and IEEE Control Systems Society Distinguished Lecturer. He is Fellow of the Royal Swedish Academy of Engineering Sciences.