Hierarchical Selective Recruitment in Linear-Threshold Brain Networks Part II: Multi-Layer Dynamics and Top-Down Recruitment

Erfan Nozari Jorge Cortés

***Abstract*—Goal-driven selective attention (GDSA) is a remarkable function that allows the complex dynamical networks of the brain to support coherent perception and cognition. Part I of this two-part paper proposes a new control-theoretic framework, termed hierarchical selective recruitment (HSR), to rigorously explain the emergence of GDSA from the brain's network structure and dynamics. This part completes the development of HSR by deriving conditions on the joint structure of the hierarchical subnetworks that guarantee top-down recruitment of the task-relevant part of each subnetwork by the subnetwork at the layer immediately above, while inhibiting the activity of task-irrelevant subnetworks at all the hierarchical layers. To further verify the merit and applicability of this framework, we carry out a comprehensive case study of selective listening in rodents and show that a small network with HSR-based structure and minimal size can explain the data with remarkable accuracy while satisfying the theoretical requirements of HSR. Our technical approach relies on the theory of switched systems and provides a novel converse Lyapunov theorem for state-dependent switched affine systems that is of independent interest.**

I. Introduction

Our ability to construct a dynamic yet coherent perception of the world, despite the numerous parallel sources of information that affect our senses, is to a great extent reliant on the brain's capability to prioritize the processing of task-relevant information over task-irrelevant distractions according to one's goals and desires. This capability, commonly known as goal-driven selective attention (GDSA), has been the subject of extensive research over the past decades. Despite major advances, a fundamental understanding of GDSA and, in particular, how it emerges from the dynamics of the underlying neuronal networks, is still missing. The aim of this work is to reduce this gap by bringing tools and insights from systems and control theory into these questions from neuroscience.

In this two-part paper, we propose the novel theoretical framework of Hierarchical Selective Recruitment (HSR) for GDSA. As stated in Part I, HSR consists of a novel hierarchical model of brain organization, a set of analytical results regarding the multi-timescale dynamics of this model, and a careful translation between the properties of these dynamics and well known experimental observations about GDSA. Inspired and supported by extensive experimental research [2]–[14], HSR relies on four major assumptions about the neuronal mechanisms underlying GDSA. These include (i) the brain's hierarchical organization, so that (cognitively-)higher areas provide control inputs to the activity of lower level ones [6], [8]–[10], [12]–[14], (ii) its sparse coding, so that task-relevant and task-irrelevant stimuli is represented and processed by sufficiently distinct neuronal populations (particularly if the two stimuli differ in major or multiple properties, such as location, sensory modality, etc.) [4]–[9], [12], [14], (iii) the distributed and graded nature of GDSA, so that selective attention happens at multiple layers of the hierarchy [3], [5]–[9], [11], [12], [14], and (iv) the concurrence of the suppression and enhancement of task-irrelevant and task-relevant activity, resp. [2]–[7], [9]–[14] (formulated as *selective inhibition* and *top-down recruitment* in HSR, resp.).

The hierarchical structure of the brain plays a key role in both selective inhibition and top-down recruitment. The position of brain areas along this hierarchy is determined based on the direction in which sensory information and decisions flow, but also by the separation of timescales between the areas. As expected, the timescale of the internal dynamics of the neuronal populations increases (becomes slower) as one moves up the hierarchy [15]–[21]. Although this hierarchy of timescales is well known in neuroscience, its role in GDSA has remained, to the best of our knowledge, uncharacterized. Using tools from singular perturbation theory, we here reveal the critical role played by this separation of timescales in the top-down recruitment of the task-relevant subnetworks and provide rigorous conditions on the joint structure of all layers that guarantee such recruitment.

Literature Review: The hierarchical organization of the brain has been recognized for decades [22]–[24] and applies to multiple aspects of brain structure and function. These aspects include (i) network topology [24]–[27] (where nodes are assigned to layers based on their position on bottom-up and top-down pathways), (ii) encoding properties [28], [29] (where nodes that have larger receptive fields and/or encode more abstract stimulus properties constitute higher layers), (iii) dynamical timescale [15]–[21], [25], [27], [30]–[34] (where nodes are grouped into layers according to the timescale of their dynamics), (iv) nodal clustering [35]–[38] (where nodes only constitute the leafs of a clustering tree), and (v) oscillatory activity [39] (where layers correspond to nested oscillatory frequency bands). Note that while hierarchical layers are composed of brain regions in (i)-(iii), this is not the case for (iv) and (v). The hierarchies (i)-(iii) are remarkably similar (in terms of the assignment of brain regions to the layers of the hierarchy), and here we particularly focus on (iii) (the timescale separation between hierarchical layers) as it plays a pivotal role in HSR.

Studies of timescale separation between cortical regions are more recent. Several experimental works have demonstrated a clear increase in intrinsic timescales as one moves up the hierarchy using indirect measures such as the length of stimulus presentation that elicits a response [15], [16], resonance frequency [17], the length of the largest time window over

A preliminary version appeared as [1] at the IEEE Conference on Decision and Control

Erfan Nozari is with the Department of Electrical and Systems Engineering, University of Pennsylvania, enozari@seas.upenn.edu. Jorge Cortés is with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, cortes@ucsd.edu.

which the responses to successive stimuli interfere [18], and how quickly the activation level of any brain region can track changes in sensory stimuli [19]. Direct evidence for this hierarchical separation of timescales was indeed provided in [20] using the decay rate of spike-count autocorrelation functions. This was shown even more comprehensively in [21] using linear-threshold rate models and the concept of *continuous hierarchies* [25], [27] (whereby the layer of each node can vary continuously according to its intrinsic timescale, therefore removing the rigidity and arbitrariness of node assignment in classical hierarchical structures). Interestingly, recent studies show that this timescale variability may have roots not only in synaptic dynamics of individual neurons [30], but also in sub-neuronal genetic factors [31] as well as supra-neuronal network structures [32]. In terms of applications, computational models of motor control were perhaps the first to exploit this cortical hierarchy of timescales [33], [34]. Despite the vastness of the literature on its roots and applications, we are not aware of any theoretical analysis of the effects of this separation of timescales on the hierarchical dynamics of neuronal networks.

The accompanying Part I [40] proposes the HSR framework, which is strongly rooted in this separation of timescales. Part I analyzes the internal dynamics of the subnetworks at each layer of the hierarchy. Using the class of linear-threshold network models, it characterizes the networks that have a unique equilibrium, are locally/globally asymptotically stable, and have bounded trajectories. In Part I, we also provide a detailed account of feedforward and feedback mechanisms for selective inhibition between any two layers of the hierarchy and show that the internal dynamical properties of the task-relevant subnetwork at each layer is the sole determiner of the dynamical properties achievable under selective inhibition.

In this paper, we complete the development of the HSR framework for GDSA by analyzing the mechanisms for top-down recruitment of the task-relevant subnetwork, combining it with the feedforward and feedback mechanisms of selective inhibition, and generalizing the combination to arbitrary number of layers. Top-down recruitment is one of the most experimentally well-documented phenomena in selective attention, see, e.g., [4]–[9], [12]–[14]. While the enhancement (a.k.a. *modulation*) of activity in the task-relevant populations is the simplest form of recruitment, our model is general and thus also allows for more complex observed forms of recruitment, such as changes in the receptive fields[1] [41]–[43].

In the analysis of top-down recruitment, we use tools from singular perturbation theory to rigorously leverage this separation of timescales. The classical result on singularly perturbed ODEs goes back to Tikhonov [44], [45, Thm 11.1] and has since inspired an extensive literature, see, e.g. [46]–[49]. Tikhonov's result, however, requires smoothness of the vector fields, which is not satisfied by linear-threshold dynamics. Fortunately, several works have sought extensions to non-smooth differential equations and even differential inclusions [50]–[53], culminating in the work [54] which we use here. Similar to Tikhonov's original work, [54] only applies to finite intervals. Extensions to infinite intervals exist [55], [56] but, as expected, they require asymptotic stability of the reduced-order model (ROM) which we do not in general have.[2]

Statement of Contributions: The paper has four main contributions. First, we use the timescale separation in hierarchical brain networks and the theory of singular perturbations to provide an analytic account of top-down recruitment in terms of conditions on the network structure. These conditions guarantee the stability of the task-relevant part of a (fast) linear-threshold subnetwork towards a reference trajectory set by a slower subnetwork. This, in particular, subsumes the most classical enhancement (strengthening) of the activity of task-relevant nodes but is more general and can account for recent, complex observations such as changes in neuronal receptive fields under GDSA [1]. We further combine these results with the results of Part I to allow for simultaneous selective inhibition and top-down recruitment, as observed in GDSA. Second, we extend this combination to hierarchical structures with an arbitrary number of layers, as observed in nature, to yield a fully developed HSR framework. Here, we also derive an extension of the stability results in Part I that guarantees GES of a multi-layer multiple timescale linear-threshold network. Third, to validate the proposed HSR framework, we provide a detailed case study of GDSA in real brain networks. Using single-unit recordings from two brain regions of rodents performing a selective listening task, we provide an in-depth analysis of appropriate choices of neuronal populations in each brain region as well as the timescales of their dynamics. We propose a novel hierarchical structure for these populations, tune the parameters of the resulting network using a novel objective function, and show that the resulting structure conforms to the theoretical results and requirements of HSR while explaining more than 90% of variability in the data. As part of our technical approach, our fourth and final contribution is a novel converse Lyapunov theorem that extends the state of the art on GES for state-dependent switched affine systems. This result only requires continuity of the vector field and guarantees the existence of an infinitely smooth quadratically-growing Lyapunov function if the dynamics is GES. Because of independent interest, we formulate and prove the result for general state-dependent switched affine systems.[3]

II. Problem Statement

The problem formulation is the same as in Part I [40]. We include here a streamlined description for a self-contained

[1]The receptive field of each neuron is the area in the stimulus space where the neuron is responsive to the presence of stimuli.

[2]Recall that in two-timescale dynamics, ROM results from replacing the fast variable with its equilibrium (reducing order to that of the slow variable).

[3]Throughout the paper, we use the following notation. $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, $\mathbb{R}_{\leq 0}$, and $\mathbb{R}_{>0}$ denote the set of reals, nonnegative reals, nonpositive reals, and positive reals, resp. $\mathbf{1}_n$, $\mathbf{0}_n$, $\mathbf{0}_{p\times n}$, and $\mathbf{I}_n$ stand for the n-vector of all ones, the n-vector of all zeros, the p-by-n zero matrix, and the identity n-by-n matrix, resp. The subscripts are omitted when clear from the context. When a vector $\mathbf{x}$ or matrix $\mathbf{A}$ are block-partitioned, $\mathbf{x}_i$ and $\mathbf{A}_{ij}$ refer to the ith block of $\mathbf{x}$ and (i,j)th block of $\mathbf{A}$, resp. Given $\mathbf{A}\in\mathbb{R}^{n\times n}$, its element-wise absolute value and spectral radius are $|\mathbf{A}|$ and $\rho(\mathbf{A})$, resp. $\|\cdot\|$ denotes vector 2-norm. For $x\in\mathbb{R}$ and $m\in\mathbb{R}_{>0}\cup\{\infty\}$, $[x]_0^m=\min\{\max\{x,0\},m\}$, which is the projection of x onto $[0,m]$. When $\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{m}\in\mathbb{R}^n_{>0}\cup\{\infty\}^n$, we similarly define $[\mathbf{x}]_{\mathbf{0}}^{\mathbf{m}}=[[x_1]_0^{m_1}\quad\cdots\quad[x_n]_0^{m_n}]^T$. For $\boldsymbol{\sigma}\in\{0,\ell,\mathrm{s}\}^n$, $\boldsymbol{\Sigma}^\ell=\boldsymbol{\Sigma}^\ell(\boldsymbol{\sigma})$ is a diagonal matrix with $\Sigma^\ell_{ii}=1$ if $\sigma_i=\ell$ and $\Sigma^\ell_{ii}=0$ otherwise. $\boldsymbol{\Sigma}^{\mathrm{s}}$ is defined similarly. We set the convention that $\boldsymbol{\Sigma}^{\mathrm{s}}\mathbf{m}=\mathbf{0}$ if $\boldsymbol{\Sigma}^{\mathrm{s}}=\mathbf{0}$ and $\mathbf{m}=\infty\mathbf{1}_n$. For $D\subseteq\mathbb{R}^n$ and $\mathbf{A}\in\mathbb{R}^{p\times n},\mathbf{b}\in\mathbb{R}^p$, we let $\mathbf{A}D+\mathbf{b}=\{\mathbf{A}\mathbf{x}+\mathbf{b}\mid\mathbf{x}\in D\}$.

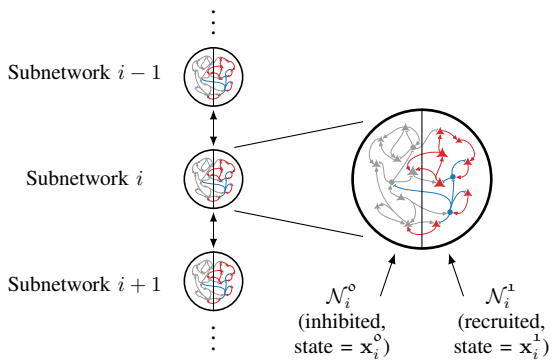

Fig. 1: Hierarchical network structure considered in this work.

exposition. We consider a hierarchical neuronal network $\mathcal{N}$, cf. Figure 1, whereby the nodes in each layer $\mathcal{N}_i$ are further decomposed into a task-irrelevant part $\mathcal{N}_i^{\circ}$ and a task-relevant part $\mathcal{N}_i^{1}$. The state evolution of each layer $\mathcal{N}_i$ is modeled with linear-threshold network dynamics of the form

$$\tau_i \dot{\mathbf{x}}_i(t) = -\mathbf{x}_i(t) + [\mathbf{W}_{i,i}\mathbf{x}_i(t) + \mathbf{d}_i(t)]_{\mathbf{0}}^{\mathbf{m}_i}, \ \mathbf{0} \le \mathbf{x}_i(0) \le \mathbf{m}_i, \quad \mathbf{m}_i \in \mathbb{R}^n_{>0} \cup \{\infty\}^n, \quad (1)$$

where $\mathbf{x}_i \in \mathbb{R}^{n_i}$, $\mathbf{W}_{i,i} \in \mathbb{R}^{n_i \times n_i}$, $\mathbf{d}_i \in \mathbb{R}^{n_i}$, and $\mathbf{m} \in \mathbb{R}^n_{>0}$ denote the state, internal synaptic connectivity, external inputs, and state upper bounds of $\mathcal{N}_i$, resp.[4]

The development of HSR is structured in four thrusts:

(i) the analysis of the relationship between structure ($\mathbf{W}_{i,i}$) and dynamical behavior for each subnetwork when operating in isolation from the rest of the network ($\mathbf{d}_i(t) \equiv \mathbf{d}_i$);

(ii) the analysis of the conditions on the joint structure of each two successive layers $\mathcal{N}_i$ and $\mathcal{N}_{i+1}$ that allows for selective inhibition of $\mathcal{N}_{i+1}^{\circ}$ by its input from $\mathcal{N}_i$, being equivalent to the stabilization of $\mathcal{N}_{i+1}^{\circ}$ to the origin (inactivity);

(iii) the analysis of the conditions on the joint structure of each two successive layers $\mathcal{N}_i$ and $\mathcal{N}_{i+1}$ that allows for top-down recruitment of $\mathcal{N}_{i+1}^{1}$ by its input from $\mathcal{N}_i$, being equivalent to the stabilization of $\mathcal{N}_{i+1}^{1}$ toward a desired trajectory set by $\mathcal{N}_i$ (activity);

(iv) the combination of (ii) and (iii) in a unified framework and its extension to the complete N-layer network of networks.

Problems (i) and (ii) are addressed in Part I [40], while problems (iii) and (iv) are the subject of this work.

We let

$$\mathbf{d}_i(t) = \mathbf{B}_i\mathbf{u}_i(t) + \tilde{\mathbf{d}}_i(t), \qquad \mathbf{u}_i \in \mathbb{R}^{p_i}_{\ge 0}, \quad (2)$$

where $\mathbf{u}_i$ is the top-down control used for inhibition of $\mathcal{N}_i^{\circ}$. While in Part I we assumed for simplicity that $\tilde{\mathbf{d}}_i(t)$ is given and constant, we here consider its complete form

$$\tilde{\mathbf{d}}_i(t) = \mathbf{W}_{i,i-1}\mathbf{x}_{i-1}(t) + \mathbf{W}_{i,i+1}\mathbf{x}_{i+1}(t) + \mathbf{c}_i,$$

where the inter-layer connectivity matrices $\mathbf{W}_{i,i-1}$ and $\mathbf{W}_{i,i+1}$ have appropriate dimensions and $\mathbf{c}_i \in \mathbb{R}^{n_i}$ captures un-modeled background activity and possibly nonzero activation thresholds. Substituting these into (1), the dynamics of each layer $\mathcal{N}_i$ is given by

$$\tau_i \dot{\mathbf{x}}_i(t) = -\mathbf{x}_i(t) + [\mathbf{W}_{i,i}\mathbf{x}_i(t) + \mathbf{W}_{i,i-1}\mathbf{x}_{i-1}(t) + \mathbf{W}_{i,i+1}\mathbf{x}_{i+1}(t) + \mathbf{B}_i\mathbf{u}_i(t) + \mathbf{c}_i]_{\mathbf{0}}^{\mathbf{m}_i}. \quad (3)$$

Also following Part I, we partition network variables as

$$\mathbf{x}_i = \begin{bmatrix} \mathbf{x}_i^{\circ} \\ \mathbf{x}_i^{1} \end{bmatrix}, \quad \mathbf{W}_{i,j} = \begin{bmatrix} \mathbf{W}_{i,j}^{\circ\circ} & \mathbf{W}_{i,j}^{\circ 1} \\ \mathbf{W}_{i,j}^{1\circ} & \mathbf{W}_{i,j}^{11} \end{bmatrix}, \quad \mathbf{B}_i = \begin{bmatrix} \mathbf{B}_i^{\circ} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{c}_i = \begin{bmatrix} \mathbf{c}_i^{\circ} \\ \mathbf{c}_i^{1} \end{bmatrix}, \quad \mathbf{m}_i = \begin{bmatrix} \mathbf{m}_i^{\circ} \\ \mathbf{m}_i^{1} \end{bmatrix} \quad (4)$$

where $\mathbf{x}_i^{\circ} \in \mathbb{R}^{r_i}$ for all $i, j \in \{1, \dots, N\}$. By convention, $\mathbf{W}_{1,0} = \mathbf{0}$, $\mathbf{W}_{N,N+1} = \mathbf{0}$, and $r_1 = 0$ (so $\mathbf{B}_1 = \mathbf{0}$ and the first subnetwork has no inhibited part). We assume that the hierarchical layers have sufficient timescale separation, i.e.,

$$\tau_1 \gg \tau_2 \gg \cdots \gg \tau_N. \quad (5)$$

Finally, let $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_{N-1})$, with

$$\epsilon_i = \tau_{i+1}/\tau_i, \qquad i = \{1, \dots, N-1\}. \quad (6)$$

Next, we first develop the main concepts and results for the case of bilayer networks (Section III) and then extend them to the setup with N layers (Section IV).

III. Selective Recruitment in Bilayer Networks

In this section we tackle the analysis of simultaneous selective inhibition and top-down recruitment in a two-layer network. We consider the same dynamics as in (3) for the lower-level subnetwork $\mathcal{N}_2$, but temporarily allow the dynamics of $\mathcal{N}_1$ to be arbitrary. This setup allows us to study the key ingredients of selective recruitment without the extra complications that arise from the multilayer interconnections of linear-threshold subnetworks and is the basis for our later developments. Further, by keeping the higher-level dynamics arbitrary, the results presented here are also of independent interest beyond HSR, as they allow for a broader range of external inputs $\mathbf{d}(t)$ than those generated by linear-threshold dynamics. This can be of interest in, for example, direct brain stimulation applications where $\mathbf{x}_1(t)$ is generated and applied by a computer in order to control the activity $\mathbf{x}_2(t)$ of certain areas of the brain. In this view, appropriate stimulation signals $\mathbf{x}_1(t)$ may be considered as an augmentation of the natural hierarchy of the brain if they vary slow enough to satisfy (5). Section IV builds on the insights obtained here to generalize this framework to the multilayer case described in Section II.

For any $\mathbf{W} \in \mathbb{R}^{n \times n}$, define $h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n_{\ge 0}$ by

$$h(\mathbf{d}) = h_{\mathbf{W},\mathbf{m}}(\mathbf{d}) \triangleq \{\mathbf{x} \in \mathbb{R}^n_{\ge 0} \mid \mathbf{x} = [\mathbf{W}\mathbf{x} + \mathbf{d}]_{\mathbf{0}}^{\mathbf{m}}\}, \quad (7)$$

which maps any constant input $\mathbf{d} \in \mathbb{R}^n$ to the corresponding set of the equilibria of (1). Due to the switched-affine form of the dynamics, h has the piecewise-affine form

$$h(\mathbf{d}) = \{\mathbf{F}_{\boldsymbol{\sigma}}\mathbf{d} + \mathbf{f}_{\boldsymbol{\sigma}} \mid \mathbf{G}_{\boldsymbol{\sigma}}\mathbf{d} + \mathbf{g}_{\boldsymbol{\sigma}} \ge \mathbf{0}, \, \boldsymbol{\sigma} \in \{0, \ell, \mathrm{s}\}^n\}, \quad (8)$$

$$\mathbf{F}_{\boldsymbol{\sigma}} = (\mathbf{I} - \boldsymbol{\Sigma}^{\ell}\mathbf{W})^{-1}\boldsymbol{\Sigma}^{\ell}, \quad \mathbf{f}_{\boldsymbol{\sigma}} = (\mathbf{I} - \boldsymbol{\Sigma}^{\ell}\mathbf{W})^{-1}\boldsymbol{\Sigma}^{\mathrm{s}}\mathbf{m},$$

$$\mathbf{G}_{\boldsymbol{\sigma}} = \begin{bmatrix} \boldsymbol{\Sigma}^{\ell} + \boldsymbol{\Sigma}^{\mathrm{s}} - \mathbf{I} & \boldsymbol{\Sigma}^{\ell} & -\boldsymbol{\Sigma}^{\ell} & \boldsymbol{\Sigma}^{\mathrm{s}} \end{bmatrix}^T \mathbf{F}_{\boldsymbol{\sigma}},$$

$$\mathbf{g}_{\boldsymbol{\sigma}} = \begin{bmatrix} \mathbf{f}_{\boldsymbol{\sigma}}^T(\boldsymbol{\Sigma}^{\ell} + \boldsymbol{\Sigma}^{\mathrm{s}} - \mathbf{I}) & \mathbf{f}_{\boldsymbol{\sigma}}^T\boldsymbol{\Sigma}^{\ell} & (\mathbf{m} - \mathbf{f}_{\boldsymbol{\sigma}})^T\boldsymbol{\Sigma}^{\ell} & (\mathbf{f}_{\boldsymbol{\sigma}} - \mathbf{m})^T\boldsymbol{\Sigma}^{\mathrm{s}} \end{bmatrix}^T.$$

[4] We note that this is a standard and widely used model in computational neuroscience, as mentioned in Part I [40]. Please see therein for a detailed literature review of its origins and prior analysis.

The existence and uniqueness of equilibria of (1) precisely corresponds to h being single-valued on $\mathbb{R}^n$, in which case we let $h : \mathbb{R}^n \to \mathbb{R}^n_{\geq 0}$ be an ordinary function. For our subsequent analysis we need h to be Lipschitz, as stated next. The proof of this result is a special case of Lemma IV.2 and thus omitted.

Lemma III.1. ***(Lipschitzness of** h**).*** *Let h be as in (7) and single-valued[5] on $\mathbb{R}^n$. Then, it is globally Lipschitz on $\mathbb{R}^n$.*

The main result of this section is as follows.

Theorem III.2. ***(Selective recruitment in bilayer networks).*** *Consider the multilayer dynamics (3) where $N = 2$, $\mathbf{W}^{\mathtt{o}\mathtt{1}}_{2,1} = \mathbf{0}$, and $\mathbf{c}^{\mathtt{o}}_2 = \mathbf{0}$ but $\mathbf{x}_1(t)$ is generated by the dynamics*

$$\tau_1 \dot{\mathbf{x}}_1(t) = \gamma(\mathbf{x}_1(t), \mathbf{x}_2(t), t). \tag{9}$$

Let $h^{\mathtt{1}}_2 = h_{\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,2}, \mathbf{m}^{\mathtt{1}}_2}$ as in (7). If

(i) γ is measurable in t, locally bounded, and locally Lipschitz in $(\mathbf{x}_1, \mathbf{x}_2)$ uniformly in t;
(ii) (9) has bounded solutions uniformly in $\mathbf{x}_2(t)$;
(iii) $p_2 \geq r_2$;
(iv) $\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,2}$ is such that $\tau \dot{\mathbf{x}}^{\mathtt{1}}_2 = -\mathbf{x}^{\mathtt{1}}_2 + [\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,2}\mathbf{x}^{\mathtt{1}}_2 + \mathbf{d}^{\mathtt{1}}_1]^{\mathbf{m}^{\mathtt{1}}_2}_{\mathbf{0}}$ is GES towards a unique equilibrium for any constant $\mathbf{d}^{\mathtt{1}}_1$;

then there exists $\mathbf{K}_2 \in \mathbb{R}^{p_2 \times n_2}$ such that by using the feedback control $\mathbf{u}_2(t) = \mathbf{K}_2\mathbf{x}_2(t)$, one has

$$\lim_{\epsilon_1 \to 0} \sup_{t \in [\underline{t}, \bar{t}]} \left\| \mathbf{x}_2(t) - \left(\mathbf{0}_{r_2}, h^{\mathtt{1}}_2\big(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{21}\mathbf{x}_1(t) + \mathbf{c}^{\mathtt{1}}_2\big)\right) \right\| = 0, \tag{10}$$

for any $0 < \underline{t} < \bar{t} < \infty$, with ϵ_1 given in (6). Further, if the dynamics of $\mathbf{x}_2$ is monotonically bounded[6], there also exists a feedforward control $\mathbf{u}_2(t) \equiv \bar{\mathbf{u}}_2$ such that (10) holds for any $0 < \underline{t} < \bar{t} < \infty$ and arbitrary $\mathbf{W}^{\mathtt{o}\mathtt{1}}_{2,1}$ and $\mathbf{c}^{\mathtt{o}}_2$.

Proof: First we prove the result for feedback control. By *(iii)*, there exists $\mathbf{K}_2 \in \mathbb{R}^{p_2 \times n_2}$ almost always (i.e., for almost all $(\mathbf{W}^{\mathtt{o}\mathtt{o}}_{2,2}, \mathbf{W}^{\mathtt{o}\mathtt{1}}_{2,2}, \mathbf{B}^{\mathtt{o}}_2)$) such that

$$\mathbf{W}_{2,2} + \mathbf{B}_2\mathbf{K}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{W}^{\mathtt{1}\mathtt{o}}_{2,2} & \mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,2} \end{bmatrix}. \tag{11}$$

Further, by [40, Thm IV.7(ii) & Thm V.3(ii)], all the principal submatrices of $-\mathbf{I} + (\mathbf{W}_{2,2} + \mathbf{B}_2\mathbf{K}_2)$ are Hurwitz. Therefore, by [40, Thm IV.3 & Assump. 1], $h^{\mathtt{1}}_2$ is singleton-valued almost always (i.e., for almost all $\mathbf{W}_{2,2}$). Thus, the $\mathbf{x}_2$-dynamics is

$$\begin{aligned} \tau_2 \dot{\mathbf{x}}^{\mathtt{o}}_2 &= -\mathbf{x}^{\mathtt{o}}_2, \\ \tau_2 \dot{\mathbf{x}}^{\mathtt{1}}_2 &= -\mathbf{x}^{\mathtt{1}}_2 + [\mathbf{W}^{\mathtt{1}\mathtt{o}}_{2,2}\mathbf{x}^{\mathtt{o}}_2 + \mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,2}\mathbf{x}^{\mathtt{1}}_2 + \mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\mathbf{x}_1 + \mathbf{c}^{\mathtt{1}}_2]^{\mathbf{m}^{\mathtt{1}}_2}_{\mathbf{0}}, \end{aligned} \tag{12}$$

and has a unique equilibrium for any *fixed* $\mathbf{x}_1$. Assumption *(iv)* and [40, Lemma A.2] then ensure that (12) is GES relative to $(\mathbf{0}_{r_2}, h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{21}\mathbf{x}_1(t) + \mathbf{c}^{\mathtt{1}}_2))$ for any fixed $\mathbf{x}_1$.

Based on assumption *(ii)*, let $D \subset \mathbb{R}^n$ be a compact set that contains the trajectory of the reduced-order model $\tau_1 \dot{\mathbf{x}}_1 = \gamma(\mathbf{x}_1, (\mathbf{0}_{r_2}, h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{21}\mathbf{x}_1(t) + \mathbf{c}^{\mathtt{1}}_2)), t)$. By assumption *(i)*, γ is Lipschitz in $(\mathbf{x}_1, \mathbf{x}_2)$ on compacts uniformly in t. Let L_γ be its associated Lipschitz constant on $D \times \{\mathbf{0}_{r_2}\} \times h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}D + \mathbf{c}^{\mathtt{1}}_2)$. Using (8) and Lemma III.1, for all $\mathbf{x}_1, \hat{\mathbf{x}}_1 \in D$,

$$\begin{aligned} &\|\gamma(\mathbf{x}_1, h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\mathbf{x}_1 + \mathbf{c}^{\mathtt{1}}_2), t) - \gamma(\hat{\mathbf{x}}_1, h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\hat{\mathbf{x}}_1 + \mathbf{c}^{\mathtt{1}}_2), t)\| \\ &\leq L_\gamma \|(\mathbf{x}_1 - \hat{\mathbf{x}}_1, h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\mathbf{x}_1 + \mathbf{c}^{\mathtt{1}}_2) - h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\hat{\mathbf{x}}_1 + \mathbf{c}^{\mathtt{1}}_2))\| \\ &\leq L_\gamma (\|\mathbf{x}_1 - \hat{\mathbf{x}}_1\| + \|h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\mathbf{x}_1 + \mathbf{c}^{\mathtt{1}}_2) - h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\hat{\mathbf{x}}_1 + \mathbf{c}^{\mathtt{1}}_2)\|) \\ &\leq L_\gamma (1 + L_h \|\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\|) \|\mathbf{x}_1 - \hat{\mathbf{x}}_1\|, \end{aligned}$$

so $\gamma(\cdot, h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\cdot + \mathbf{c}^{\mathtt{1}}_2), t) : \mathbb{R}^{n_1} \to \mathbb{R}^{n_1}$ is $L_\gamma(1 + L_h\|\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\|)$-Lipschitz on D. Using this, Lemma IV.2 again, and the change of variables $t' \triangleq t/\tau_1$, the claim follows from [54, Prop 1].[7]

Next, we prove the result for constant feedforward control $\mathbf{u}_2(t) \equiv \bar{\mathbf{u}}_2$. Based on assumption *(ii)*, let $\bar{\mathbf{x}}_1 \in \mathbb{R}^{n_1}_{>0}$ be the bound on the trajectories of (9) and $\bar{\mathbf{u}}_2$ be a solution of

$$\mathbf{B}^{\mathtt{o}}_2 \bar{\mathbf{u}}_2 = -[[\mathbf{W}^{\mathtt{o}\mathtt{o}}_{2,2} \ \mathbf{W}^{\mathtt{o}\mathtt{1}}_{2,2}]]^{\infty}_{\mathbf{0}} \boldsymbol{\nu}(\bar{\mathbf{x}}_1) - [\mathbf{W}^{\mathtt{o}\mathtt{1}}_{2,1}]^{\infty}_{\mathbf{0}} \bar{\mathbf{x}}_1 + [\mathbf{c}^{\mathtt{o}}_2]^{\infty}_{\mathbf{0}},$$

where $\boldsymbol{\nu}$ comes from the monotone boundedness of the dynamics of $\mathbf{x}_2$. This solution almost always exists by assumption *(ii)*. Then, the dynamics of $\mathbf{x}_2$ simplifies to (12), and [40, Lemma A.2] guarantees that it is GES relative to $(\mathbf{0}_{r_2}, h^{\mathtt{1}}_2(\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,1}\mathbf{x}_1 + \mathbf{c}^{\mathtt{1}}_2))$ for any *fixed* $\mathbf{x}_1$. The claim then follows, similar to the feedback case, from [54, Prop 1]. ■

Remark III.3. ***(Validity of the assumptions of Theorem III.2.).*** Assumption *(i)* is merely technical and satisfied by all well-known models of neuronal rate dynamics, including the linear-threshold model itself. Likewise, assumption *(ii)* is always satisfied in reality, as the firing rates of all biological neuronal networks are bounded by the inverse of the refractory period of their neurons. In theory, the verification of this assumption depends clearly on γ. If a linear-threshold model is used, we can instead use Theorem IV.3 and relax assumption *(ii)* to a less restrictive one (assumption *(i)* of Theorem IV.3), which can in turn be verified using the sufficient condition in Theorem IV.4. Assumption *(iii)* requires the existence of at least as many inhibitory control channels as the number of nodes in $\mathcal{N}_2$ that are to be inhibited. Indeed, selective inhibition is still possible without this assumption (cf. Theorem IV.3), but may require excessive inhibitory resources. The most critical requirement is assumption *(iv)*, but is not only sufficient but also necessary for inhibitory stabilization (cf. [40, Thm IV.8] for conditions on $\mathbf{W}^{\mathtt{1}\mathtt{1}}_{2,2}$ that ensure this assumption as well as [40, Thm V.2 & V.3] for its necessity for feedforward and feedback inhibitory stabilization). □

The main conclusion of Theorem III.2 is the Tikhonov-type singular perturbation statement in (10). According to this statement, the tracking error can be made arbitrarily small, i.e., for any $\theta > 0$,

$$|\mathbf{x}_2(t) - (\mathbf{0}_{r_2}, h^{\mathtt{1}}_2(\mathbf{x}_1(t)))| \leq \theta \mathbf{1}_{n_2}, \qquad \forall t \in [\underline{t}, \bar{t}], \tag{13}$$

provided that τ_2/τ_1 is sufficiently small. As discussed in Section I, this timescale separation is characteristic of biological

[5] It is possible to show, using the same proof technique, that h is Lipschitz in the Hausdorff metric even when it is multiple-valued (recall that the Hausdorff distance between two sets $S_1, S_2 \in \mathbb{R}^n$ is defined as $\max\{\sup_{\mathbf{a} \in S_1} \inf_{\mathbf{b} \in S_2} \|\mathbf{a} - \mathbf{b}\|, \sup_{\mathbf{b} \in S_2} \inf_{\mathbf{a} \in S_1} \|\mathbf{a} - \mathbf{b}\|\}$).

[6] See [40, Def V.1].

[7] [54, Prop 1] is applicable to singularly perturbed differential inclusions and thus technically involved, but for non-smooth ODEs such as (3), its assumptions can be simplified to: 1. Lipschitzness of dynamics uniformly in t, 2. Existence, uniqueness, and Lipschitzness of the equilibrium map of fast dynamics, 3. Lipschitzness and boundedness of the reduced-order model, 4. asymptotic stability of the fast dynamics uniformly in t and the slow variable, and 5. global attractivity of fast dynamics for any fixed slow variable.

neuronal networks. In general, the smaller the time constant ratio τ_2/τ_1, the smaller the tracking error θ. As shown in [20], several pairs of regions along the sensory-frontal pathways have successive time constant ratios between $1/1.5$ and $1/2.5$, which is often (more than) enough in simulations for (13) to hold with small enough θ, as shown in Example III.4 below.

An important observation regarding (13) is that the equilibrium map $h_2^{\mathbb{1}}$ does not have a closed-form expression, so the reference trajectory $h_2^{\mathbb{1}}(\mathbf{x}_1(t))$ of the lower-level network is only implicitly known for any given $\mathbf{x}_1(t)$. However, if a desired trajectory $\boldsymbol{\xi}_2^{\mathbb{1}}(t) \in \prod_{j=r_2+1}^{n_2}[0, m_{2,j}]$ for $\mathbf{x}_2^{\mathbb{1}}$ is known a priori, one can specify the appropriate γ such that $h_2^{\mathbb{1}}(\mathbf{x}_1(t)) = \boldsymbol{\xi}_2^{\mathbb{1}}(t)$. To show this, let the dynamics of $\boldsymbol{\xi}_2^{\mathbb{1}}(t)$ be

$$\tau_1 \dot{\boldsymbol{\xi}}_2^{\mathbb{1}}(t) = \gamma_\xi(\boldsymbol{\xi}_2^{\mathbb{1}}(t), t).$$

Then, choosing $\mathbf{x}_1(t) = (\mathbf{W}_{2,1}^{\mathbb{1}\mathbb{1}})^{-1}\big((\mathbf{I} - \mathbf{W}_{2,2}^{\mathbb{1}\mathbb{1}})\boldsymbol{\xi}_2^{\mathbb{1}}(t) - \mathbf{c}_2^{\mathbb{1}}\big)$,

$$[\mathbf{W}_{2,2}^{\mathbb{1}\mathbb{1}}\boldsymbol{\xi}_2^{\mathbb{1}}(t) + \mathbf{W}_{2,1}^{\mathbb{1}\mathbb{1}}\mathbf{x}_1(t) + \mathbf{c}_2^{\mathbb{1}}]_{\mathbf{0}}^{\mathbf{m}_2^{\mathbb{1}}} = [\boldsymbol{\xi}_2^{\mathbb{1}}(t)]_{\mathbf{0}}^{\mathbf{m}_2^{\mathbb{1}}} = \boldsymbol{\xi}_2^{\mathbb{1}}(t),$$

which, according to (7), implies $\boldsymbol{\xi}_2^{\mathbb{1}}(t) = h_2^{\mathbb{1}}(\mathbf{x}_1(t))$.

We use this result to illustrate the core concepts of the bilayer HSR in a synthetic but biologically-inspired example, where a inhibitory subnetwork generates oscillations which are selectively induced on a lower-level excitatory subnetwork.

Example III.4. ***(HSR of an excitatory subnetwork by inhibitory oscillations).*** Consider the dynamics (3) with $N = 2$, a 3-dimensional excitatory subnetwork at the lower level, a 3-dimensional inhibitory subnetwork at the higher level, and $\mathbf{m}_1 = \mathbf{m}_2 = \infty\mathbf{1}_3$ (Figure 2). Let

$$\mathbf{W}_{1,1} = \begin{bmatrix} 0 & -0.8 & -1.7 \\ -1 & 0 & -0.5 \\ -0.7 & -1.8 & 0 \end{bmatrix}, \quad \mathbf{c}_1 = \begin{bmatrix} 11 \\ 10 \\ 10 \end{bmatrix},$$

$$\mathbf{W}_{2,2} = \begin{bmatrix} 0 & 0.9 & 1.2 \\ 0.7 & 0 & 1 \\ 0.8 & 0.2 & 0 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 2 \\ 3.5 \\ 2.5 \end{bmatrix},$$

$$\mathbf{W}_{1,2} = \mathbf{0}, \quad \mathbf{W}_{2,1} = -\mathbf{I}, \quad u_2 = 5. \tag{14}$$

This example satisfies all the assumptions of Theorem III.2, so we expect the actual $\mathbf{x}_2$-trajectory to be close to the *desired* $\mathbf{x}_2$-trajectory $(0, h_2^{\mathbb{1}}(\mathbf{x}_1(t)))$ provided that $\epsilon_1 \ll 1$. As shown in Figure 2, $\mathbf{x}_2(t)$ and $(0, h_2^{\mathbb{1}}(\mathbf{x}_1(t)))$ are remarkably close even with a mild separation of timescales, $\epsilon_1 = 0.5$.

This example further illustrates the complementary roles of selective inhibition and selective recruitment. The complete $\mathbf{x}_2$-subsystem is unstable by itself, but any two-dimensional subnetwork of it is stable. Therefore, $\mathcal{N}_1$ can selectively inhibit any single node of $\mathcal{N}_2$ while simultaneously recruiting (e.g., by inducing oscillations in) the remaining two. Thus, as suggested earlier in [40, Rem V.7], different "tasks" can be accomplished at different times by varying the selectively recruited subnetwork over time. Generalizing this to more complex networks allows for more flexible selective recruitment schemes of larger neuronal subnetworks, as observed in nature. □

Remark III.5. ***(Biological relevance of Example III.4).*** In addition to providing a simple illustration of the HSR framework developed here, Example III.4 has interesting similarities with well-known aspects of selective attention in the brain. Extensive studies have demonstrated a robust correlation between oscillatory activity, particularly in the gamma range ($\sim 30-100^{\text{Hz}}$), and selective attention [57]–[60]. Furthermore, gamma oscillations in the cortex are shown to be primarily generated by networks of inhibitory neurons, which then recruit the excitatory populations (see [61] and the references therein), as captured by the network structure of Figure 2. Interestingly, the oscillations generated by the higher-level inhibitory subnetwork fall within the gamma band by setting $\tau_1 \sim 3^{\text{ms}}$ which lies within the decay time constant range of GABA_A inhibitory receptors[8] (the major type of inhibitory synapse in the central nervous system). □

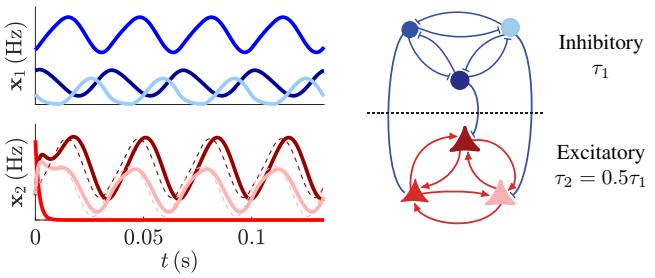

Fig. 2: The network structure (right) and trajectories (left) of the two-timescale network in (14) for $\tau_1 = 3.3^{\text{ms}}$. The red pyramids and blue circles depict excitatory and inhibitory nodes, resp., and the trajectory colors on the left correspond to node colors on the right. The dashed lines show the desired reference trajectories $\big(0, h_2^{\mathbb{1}}(\mathbf{W}_{2,1}^{\mathbb{1}\mathbb{1}}\mathbf{x}_1(t) + \mathbf{c}_2^{\mathbb{1}})\big)$.

IV. Selective Recruitment in Multilayer Networks

We tackle here the problem of Section II in its general form and consider an N-layer hierarchical structure of subnetworks with linear-threshold dynamics. Given (3), let

$$h_i^{\mathbb{1}} : \mathbf{c}_i^{\mathbb{1}} \rightrightarrows \{\mathbf{x}_i^{\mathbb{1}} \mid \mathbf{x}_i^{\mathbb{1}} = [\mathbf{W}_{i,i+1}^{\mathbb{1}\mathbb{1}} h_{i+1}^{\mathbb{1}}(\mathbf{W}_{i+1,i}^{\mathbb{1}\mathbb{1}}\mathbf{x}_i^{\mathbb{1}} + \mathbf{c}_{i+1}^{\mathbb{1}}) + \mathbf{W}_{i,i}^{\mathbb{1}\mathbb{1}}\mathbf{x}_i^{\mathbb{1}} + \mathbf{c}_i^{\mathbb{1}}]_{\mathbf{0}}^{\mathbf{m}_i^{\mathbb{1}}}\}, \ i = 2, \dots, N-1,$$

with $h_N^{\mathbb{1}} = h_{\mathbf{W}_{N,N}^{\mathbb{1}\mathbb{1}}, \mathbf{m}_N^{\mathbb{1}}}$, be the recursive definition of the (set-valued) equilibrium maps of the task-relevant parts of the layers $\{2, \dots, N\}$. These maps play a central role in the multiple-timescale dynamics of (3). Therefore, we begin by characterizing their piecewise-affine nature. The proof of the following result can be found in Appendix B.

Lemma IV.1. ***(Piecewise affinity of equilibrium maps is preserved along layers of hierarchical linear-threshold network).*** *Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a piecewise affine function,*

$$h(\mathbf{c}) = \mathbf{F}_\lambda \mathbf{c} + \mathbf{f}_\lambda, \qquad \forall \mathbf{c} \in \Psi_\lambda \triangleq \{\mathbf{c} \mid \mathbf{G}_\lambda \mathbf{c} + \mathbf{g}_\lambda \geq \mathbf{0}\}, \ \forall \lambda \in \Lambda,$$

where Λ is a finite index set and $\bigcup_{\lambda\in\Lambda} \Psi_\lambda = \mathbb{R}^n$. Given matrices $\mathbf{W}_\ell, \ell = 1, 2, 3$ and a vector $\bar{\mathbf{c}}$, assume

$$\mathbf{x} = [\mathbf{W}_1\mathbf{x} + \mathbf{W}_2 h(\mathbf{W}_3\mathbf{x} + \bar{\mathbf{c}}) + \mathbf{c}']_{\mathbf{0}}^{\mathbf{m}}, \tag{15}$$

[8] See, e.g., the Neurotransmitter Time Constants database of the CNRGlab at the University of Waterloo, http://compneuro.uwaterloo.ca/research/constants-constraints/neurotransmitter-time-constants-pscs.html.

is known to have a unique solution $\mathbf{x} \in \mathbb{R}^{n'}$ *for all* $\mathbf{c}' \in \mathbb{R}^{n'}$ *and let* $h'(\mathbf{c}')$ *be this unique solution. Then, there exists a finite index set* Λ' *and* $\{(\mathbf{F}'_{\lambda'}, \mathbf{f}'_{\lambda'}, \mathbf{G}'_{\lambda'}, \mathbf{g}'_{\lambda'})\}_{\lambda' \in \Lambda'}$ *such that*

$$h'(\mathbf{c}') = \mathbf{F}'_{\lambda'}\mathbf{c}' + \mathbf{f}'_{\lambda'}, \quad \forall \mathbf{c}' \in \Psi'_{\lambda'} \triangleq \{\mathbf{c}' \mid \mathbf{G}'_{\lambda'}\mathbf{c}' + \mathbf{g}'_{\lambda'} \geq \mathbf{0}\}, \quad \forall \lambda' \in \Lambda', \tag{16}$$

and $\bigcup_{\lambda' \in \Lambda'} \Psi'_{\lambda'} = \mathbb{R}^{n'}$. □

A special case of Lemma IV.1 is when $\mathbf{W}_2 = \mathbf{0}$, where h' becomes, like $h_N^{\mathbf{1}}$, the standard equilibrium map (7). Next, we characterize the global Lipschitzness property of the equilibrium maps. The proof is in Appendix B.

Lemma IV.2. ***(Piecewise affine equilibrium maps are globally Lipschitz).*** *Let* $h : \mathbb{R}^n \to \mathbb{R}^n$ *be a piecewise affine function of the form*

$$h(\mathbf{c}) = \mathbf{F}_\lambda \mathbf{c} + \mathbf{f}_\lambda, \quad \forall \mathbf{c} \in \Psi_\lambda \triangleq \{\mathbf{c} \mid \mathbf{G}_\lambda \mathbf{c} + \mathbf{g}_\lambda \geq \mathbf{0}\}, \quad \forall \lambda \in \Lambda,$$

where Λ *is a finite index set and* $\bigcup_{\lambda \in \Lambda} \Psi_\lambda = \mathbb{R}^n$. *Then,* h *is globally Lipschitz.* □

We are now ready to generalize Theorem III.2 to an N-layer architecture while at the same time relaxing several of its simplifying assumptions in favor of generality.

Theorem IV.3. ***(Selective recruitment in multilayer networks).*** *Consider the dynamics* (3)*. If*

(i) The reduced-order model (ROM)

$$\tau_1 \dot{\bar{\mathbf{x}}}_1^{\mathbf{1}} = -\bar{\mathbf{x}}_1^{\mathbf{1}} + [\mathbf{W}_{1,1}^{\mathbf{11}} \bar{\mathbf{x}}_1^{\mathbf{1}} + \mathbf{W}_{1,2}^{\mathbf{11}} h_2^{\mathbf{1}}(\mathbf{W}_{2,1}^{\mathbf{11}} \bar{\mathbf{x}}_1^{\mathbf{1}} + \mathbf{c}_2^{\mathbf{1}}) + \mathbf{c}_1^{\mathbf{1}}]_{\mathbf{0}}^{\mathbf{m}_1^{\mathbf{1}}},$$

of the first subnetwork has bounded solutions (recall $\mathbf{x}_1 \equiv \mathbf{x}_1^{\mathbf{1}}$ *since* $r_1 = 0$*);*

(ii) For all $i = 2, \ldots, N$,

$$\tau_i \dot{\mathbf{x}}_i^{\mathbf{1}}(t) = -\mathbf{x}_i^{\mathbf{1}}(t) + [\mathbf{W}_{i,i}^{\mathbf{11}} \mathbf{x}_i^{\mathbf{1}}(t) + \mathbf{W}_{i,i+1}^{\mathbf{11}} h_{i+1}^{\mathbf{1}}(\mathbf{W}_{i+1,i}^{\mathbf{11}} \mathbf{x}_i^{\mathbf{1}}(t) + \mathbf{c}_{i+1}^{\mathbf{1}}) + \mathbf{c}_i^{\mathbf{1}}]_{\mathbf{0}}^{\mathbf{m}_i^{\mathbf{1}}},$$

is GES towards a unique equilibrium for any $\mathbf{c}_{i+1}^{\mathbf{1}}$ *and* $\mathbf{c}_i^{\mathbf{1}}$*;*

then there exists $\mathbf{K}_i \in \mathbb{R}^{p_i \times n_i}$ *and* $\bar{\mathbf{u}}_i : \mathbb{R}_{\geq 0} \to \mathbb{R}^{p_i}_{\geq 0}, i \in \{2, \ldots, N\}$ *such that using the feedback-feedforward control*

$$\mathbf{u}_i(t) = \mathbf{K}_i \mathbf{x}_i(t) + \bar{\mathbf{u}}_i(t), \qquad i \in \{2, \ldots, N\}, \tag{17}$$

we have, for any $0 < \underline{t} < \bar{t} < \infty$,

$$\lim_{\boldsymbol{\epsilon} \to \mathbf{0}} \sup_{t \in [\underline{t}, \bar{t}]} \|\mathbf{x}_i^{\mathbf{o}}(t)\| = \mathbf{0}, \qquad \forall i \in \{2, \ldots, N\}, \tag{18a}$$

and

$$\lim_{\boldsymbol{\epsilon} \to \mathbf{0}} \sup_{t \in [0, \bar{t}]} \|\mathbf{x}_1^{\mathbf{1}}(t) - \bar{\mathbf{x}}_1^{\mathbf{1}}(t)\| = 0, \tag{18b}$$

$$\lim_{\boldsymbol{\epsilon} \to \mathbf{0}} \sup_{t \in [\underline{t}, \bar{t}]} \|\mathbf{x}_2^{\mathbf{1}}(t) - h_2^{\mathbf{1}}(\mathbf{W}_{2,1}^{\mathbf{11}} \mathbf{x}_1^{\mathbf{1}}(t) + \mathbf{c}_2^{\mathbf{1}})\| = 0, \tag{18c}$$

$$\vdots$$

$$\lim_{\boldsymbol{\epsilon} \to \mathbf{0}} \sup_{t \in [\underline{t}, \bar{t}]} \|\mathbf{x}_N^{\mathbf{1}}(t) - h_N^{\mathbf{1}}(\mathbf{W}_{N,N-1}^{\mathbf{11}} \mathbf{x}_{N-1}^{\mathbf{1}}(t) + \mathbf{c}_N^{\mathbf{1}})\| = 0. \tag{18d}$$

Proof: For any 2×2 block-partitioned matrix $\mathbf{W}$, we introduce the convenient notation $\mathbf{W}^{\ell,\mathrm{all}} \triangleq [\mathbf{W}^{\ell\mathbf{o}} \ \mathbf{W}^{\ell\mathbf{1}}]$ and $\mathbf{W}^{\mathrm{all},\ell} \triangleq [(\mathbf{W}^{\mathbf{o}\ell})^T \ (\mathbf{W}^{\mathbf{1}\ell})^T]^T$ for $\ell = \mathbf{o}, \mathbf{1}$. Further, for any $i \in \{2, \ldots, N\}$, let $\mathbf{x}_{1:i} = [\mathbf{x}_1^T \ \ldots \ \mathbf{x}_i^T]^T$. To begin with, let $\mathbf{K}_N$ and $\bar{\mathbf{u}}_N$ be such that

$$\mathbf{B}_N^{\mathbf{o}} \mathbf{K}_N \leq -\mathbf{W}_{N,N}^{\mathbf{o},\mathrm{all}}, \tag{19a}$$

$$\bar{\mathbf{u}}_N(t) \leq -\mathbf{W}_{N,N-1}^{\mathbf{o},\mathrm{all}} \mathbf{x}_{N-1}(t) - \mathbf{c}_N^{\mathbf{o}}, \qquad \forall t, \tag{19b}$$

Note that, if $p_N \geq r_N$, then (19a) can be satisfied with equality. Otherwise, (19a) can still be satisfied since all the rows of $\mathbf{B}_N^{\mathbf{o}}$ are nonzero, but may require excessive amounts of inhibition. Also, notice that $\bar{\mathbf{u}}_N$ is set by the subnetwork $N-1$, which has access to $\mathbf{x}_{N-1}(t)$ and can thus fulfill (19b). As a result, the nodes in $\mathbf{x}_N^{\mathbf{o}}$ are fully inhibited and evolve according to $\tau_N \dot{\mathbf{x}}_N^{\mathbf{o}} = -\mathbf{x}_N^{\mathbf{o}}$. The overall dynamics become

$$\begin{aligned}
\tau_1 \dot{\mathbf{x}}_1 &= -\mathbf{x}_1 + [\mathbf{W}_{1,1}\mathbf{x}_1 + \mathbf{W}_{1,2}\mathbf{x}_2 + \mathbf{c}_1]_{\mathbf{0}}^{\mathbf{m}_1}, \\
&\vdots \\
\tau_{N-1} \dot{\mathbf{x}}_{N-1} &= -\mathbf{x}_{N-1} + [\mathbf{W}_{N-1,N-1}\mathbf{x}_{N-1} + \mathbf{B}_{N-1}\mathbf{u}_{N-1} \\
&\qquad + \mathbf{W}_{N-1,N}\mathbf{x}_N + \mathbf{W}_{N-1,N-2}\mathbf{x}_{N-2} + \mathbf{c}_{N-1}]_{\mathbf{0}}^{\mathbf{m}_{N-1}}, \\
\epsilon_{N-1}\tau_{N-1} \dot{\mathbf{x}}_N^{\mathbf{o}} &= -\mathbf{x}_N^{\mathbf{o}}, \\
\epsilon_{N-1}\tau_{N-1} \dot{\mathbf{x}}_N^{\mathbf{1}} &= -\mathbf{x}_N^{\mathbf{1}} + [\mathbf{W}_{N,N}^{\mathbf{1},\mathrm{all}} \mathbf{x}_N + \mathbf{W}_{N,N-1}^{\mathbf{1},\mathrm{all}} \mathbf{x}_{N-1} + \mathbf{c}_N^{\mathbf{1}}]_{\mathbf{0}}^{\mathbf{m}_N^{\mathbf{1}}}.
\end{aligned}$$

Letting $\epsilon_{N-1} \to 0$, we get our first separation of timescales between $\mathbf{x}_N$ and $\mathbf{x}_{1:N-1}$, as follows. For any constant $\mathbf{x}_{N-1}$, the $\mathbf{x}_N$ dynamics is GES by assumption *(ii)* and [40, Lemma A.2]. Further, the equilibrium map $h_N = (\mathbf{0}_{r_N}, h_N^{\mathbf{1}})$ of the N'th subnetwork is globally Lipschitz by Lemmas IV.1 and IV.2, and the entire vector field of network dynamics is globally Lipschitz due to the Lipschitzness of $[\cdot]_0^m$. Therefore, it follows from [54, Prop 1] that for any $0 < \underline{t} < \bar{t} < \infty$,

$$\begin{aligned}
&\lim_{\epsilon_{N-1} \to 0} \sup_{t \in [\underline{t}, \bar{t}]} \|\mathbf{x}_N^{\mathbf{o}}(t)\| = 0, \\
&\lim_{\epsilon_{N-1} \to 0} \sup_{t \in [\underline{t}, \bar{t}]} \|\mathbf{x}_N^{\mathbf{1}}(t) - h_N^{\mathbf{1}}(\mathbf{W}_{N,N-1}^{\mathbf{1},\mathrm{all}} \mathbf{x}_{N-1}(t) + \mathbf{c}_N^{\mathbf{1}})\| = 0, \\
&\lim_{\epsilon_{N-1} \to 0} \sup_{t \in [0, \bar{t}]} \|\mathbf{x}_{1:N-1}(t) - \mathbf{x}_{1:N-1}^{(1)}(t)\| = 0.
\end{aligned}$$

Here, $\mathbf{x}_{1:N-1}^{(1)}$ is the solution of the "first-step ROM"

$$\begin{aligned}
\tau_1 \dot{\mathbf{x}}_1^{(1)} &= -\mathbf{x}_1^{(1)} + [\mathbf{W}_{1,1}\mathbf{x}_1^{(1)} + \mathbf{W}_{1,2}\mathbf{x}_2^{(1)} + \mathbf{c}_1]_{\mathbf{0}}^{\mathbf{m}_1}, \\
&\vdots \\
\tau_{N-1} \dot{\mathbf{x}}_{N-1}^{(1)} &= -\mathbf{x}_{N-1}^{(1)} + [\mathbf{W}_{N-1,N-1}\mathbf{x}_{N-1}^{(1)} \\
&\qquad + \mathbf{W}_{N-1,N}^{\mathrm{all},\mathbf{1}} h_N^{\mathbf{1}}(\mathbf{W}_{N,N-1}^{\mathbf{1},\mathrm{all}} \mathbf{x}_{N-1}^{(1)}(t) + \mathbf{c}_N^{\mathbf{1}}) \\
&\qquad + \mathbf{W}_{N-1,N-2}\mathbf{x}_{N-2}^{(1)} + \mathbf{B}_{N-1}\mathbf{u}_{N-1} + \mathbf{c}_{N-1}]_{\mathbf{0}}^{\mathbf{m}_{N-1}},
\end{aligned}$$

which results from replacing $\mathbf{x}_N$ with its equilibrium value. Except for technical adjustments, the remainder of the proof essentially follows by repeating this process $N-2$ times. In particular, for $i = N-1, \ldots, 2$, let $\mathbf{K}_i$ and $\bar{\mathbf{u}}_i$ be such that

$$\begin{aligned}
\mathbf{B}_i^{\mathbf{o}} \mathbf{K}_i &\leq -|\mathbf{W}_{i,i}^{\mathbf{o},\mathrm{all}}| - |\mathbf{W}_{i,i+1}^{\mathbf{o}\mathbf{1}}| \bar{\mathbf{F}}_{i+1} |\mathbf{W}_{i+1,i}^{\mathbf{1},\mathrm{all}}|, \\
\bar{\mathbf{u}}_i(t) &\leq -\mathbf{W}_{i,i-1}^{\mathbf{o}:} \mathbf{x}_{i-1}(t) - \mathbf{c}_i^{\mathbf{o}}, \qquad \forall t,
\end{aligned}$$

where $\bar{\mathbf{F}}_i \in \mathbb{R}^{(n_i - r_i) \times (n_i - r_i)}$ is the entry-wise maximal gain of the map $h_i^{\mathbf{1}}$ over $\mathbb{R}^{n_i - r_i}$ (cf. Theorem IV.4). This results in

the "$(N-i)$'th-step ROM"

$$
\begin{aligned}
\tau_1\dot{\mathbf{x}}_1^{(N-i)} &= -\mathbf{x}_1^{(N-i)} + [\mathbf{W}_{1,1}\mathbf{x}_1^{(N-i)} + \mathbf{W}_{1,2}\mathbf{x}_2^{(N-i)} + \mathbf{c}_1]_{\mathbf{0}}^{\mathbf{m}_1},\\
&\vdots\\
\tau_{i-1}\dot{\mathbf{x}}_{i-1}^{(N-i)} &= -\mathbf{x}_{i-1}^{(N-i)} + [\mathbf{W}_{i-1,i-1}\mathbf{x}_{i-1}^{(N-i)}\\
&\qquad + \mathbf{W}_{i-1,i}\mathbf{x}_i^{(N-i)} + \mathbf{W}_{i-1,i-2}\mathbf{x}_{i-2}^{(N-i)}\\
&\qquad + \mathbf{B}_{i-1}\mathbf{u}_{i-1} + \mathbf{c}_{i-1}]_{\mathbf{0}}^{\mathbf{m}_{i-1}},\\
\epsilon_{i-1}\tau_{i-1}\dot{\mathbf{x}}_i^{(N-i)^{\mathbf{o}}} &= -\mathbf{x}_i^{(N-i)^{\mathbf{o}}},\\
\epsilon_{i-1}\tau_{i-1}\dot{\mathbf{x}}_i^{(N-i)^{\mathbf{1}}} &= -\mathbf{x}_i^{(N-i)^{\mathbf{1}}} + [\mathbf{W}_{i,i}^{\mathbf{1},\text{all}}\mathbf{x}_i^{(N-i)^{\mathbf{1}}}\\
&\qquad + \mathbf{W}_{i,i+1}^{\text{all},\mathbf{1}}h_{i+1}^{\mathbf{1}}(\mathbf{W}_{i+1,i}^{\mathbf{1},\text{all}}\mathbf{x}_i^{(N-i)}(t) + \mathbf{c}_{i+1}^{\mathbf{1}})\\
&\qquad + \mathbf{W}_{i,i-1}^{\mathbf{1},\text{all}}\mathbf{x}_{i-1}^{(N-i)} + \mathbf{c}_i^{\mathbf{1}}]_{\mathbf{0}}^{\mathbf{m}_i^{\mathbf{1}}}.
\end{aligned}
$$

Similarly to above, invoking [54, Prop 1] then ensures that

$$
\begin{aligned}
&\lim_{\boldsymbol{\epsilon}\to\mathbf{0}} \sup_{t\in[\underline{t},\bar{t}]} \|\mathbf{x}_i^{(N-i)^{\mathbf{o}}}(t)\| = 0,\\
&\lim_{\boldsymbol{\epsilon}\to\mathbf{0}} \sup_{t\in[\underline{t},\bar{t}]} \|\mathbf{x}_i^{(N-i)^{\mathbf{1}}}(t) - h_i^{\mathbf{1}}(\mathbf{W}_{i,i-1}^{\mathbf{1},\text{all}}\mathbf{x}_{i-1}^{(N-i)}(t) + \mathbf{c}_i^{\mathbf{1}})\| = 0,\\
&\lim_{\boldsymbol{\epsilon}\to 0} \sup_{t\in[0,\bar{t}]} \|\mathbf{x}_{1:i-1}^{(N-i)}(t) - \mathbf{x}_{1:i-1}^{(N-i+1)}(t)\| = 0.
\end{aligned}
$$

Note that, after every invocation of [54, Prop 1], the super-index inside the parenthesis increases by 1, showing one more replacement of a fast dynamics by its equilibrium state. In particular, after the $(N-1)$'th invocation of [54, Prop 1], we reach $\mathbf{x}_1^{(N-1)^{\mathbf{1}}}$, which is the same as $\bar{\mathbf{x}}_1^{\mathbf{1}}$ in the statement. Together, these results (and sufficiently many applications of the triangle inequality and Lemma IV.2) ensure (18). ∎

An instructive difference, by design, between Theorems III.2 and IV.3 is the separate treatment of feedforward and feedback inhibition in the former versus the combination of the two in the latter. While the separate treatment in Theorem III.2 is conceptually simpler and highlights the theoretical difference between the two inhibitory mechanisms, the combination in Theorem IV.3 results in more flexibility and less conservativeness: in pure feedforward inhibition, countering local excitations requires monotone boundedness and a sufficiently large $\bar{\mathbf{u}}$ providing inhibition under the worst-case scenario, a goal that is achieved more efficiently using feedback. On the other hand, pure feedback inhibition needs to dynamically cancel local excitations at all times and is also unable to counter the effects of constant background excitation, limitations that are easily addressed when combined with feedforward inhibition.

Similar to Theorem III.2 (cf. Remark III.3), assumption *(ii)* of Theorem IV.3 is its only critical requirement which we next relate to the joint structure of the subnetworks.

Theorem IV.4. ***(Sufficient condition for existence and uniqueness of equilibria and GES in multilayer linear-threshold networks).*** *Let $h:\mathbb{R}^n\to\mathbb{R}^n$ be a piecewise affine function of the form*

$$
h(\mathbf{c}) = \mathbf{F}_\lambda\mathbf{c} + \mathbf{f}_\lambda, \qquad \forall \mathbf{c}\in\Psi_\lambda \triangleq \{\mathbf{c} \mid \mathbf{G}_\lambda\mathbf{c} + \mathbf{g}_\lambda \geq \mathbf{0}\}, \quad \forall\lambda\in\Lambda, \tag{20}
$$

where Λ is a finite index set and $\bigcup_{\lambda\in\Lambda}\Psi_\lambda = \mathbb{R}^n$. Further, let $\bar{\mathbf{F}} \triangleq \max_{\lambda\in\Lambda}|\mathbf{F}_\lambda|$ be the matrix whose elements are the maximum of the corresponding elements from $\{|\mathbf{F}_\lambda|\}_{\lambda\in\Lambda}$. For arbitrary matrices $\mathbf{W}_\ell$, $\ell = 1,2,3$, if $\rho\big(|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|\big) < 1$, then the linear-threshold dynamics

$$
\tau\dot{\mathbf{x}}(t) = -\mathbf{x}(t) + [\mathbf{W}_1\mathbf{x}(t) + \mathbf{W}_2 h(\mathbf{W}_3\mathbf{x}(t) + \bar{\mathbf{c}}) + \mathbf{c}]_{\mathbf{0}}^{\mathbf{m}},
$$

is GES towards a unique equilibrium for all $\bar{\mathbf{c}}$ and $\mathbf{c}$.

Proof: We use the same proof technique as in [62, Prop. 3]. For simplicity, assume that $|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|$ is irreducible (i.e., the network topology induced by it is strongly connected)[9]. Then, the left Perron-Frobenius eigenvector $\boldsymbol{\alpha}$ of $|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|$ has positive entries [63, Fact 4.11.4], making the map $\|\cdot\|_{\boldsymbol{\alpha}} : \mathbf{v}\to\|\mathbf{v}\|_{\boldsymbol{\alpha}} \triangleq \boldsymbol{\alpha}^T|\mathbf{v}|$ a norm on $\mathbb{R}^n$. Further, it can be shown, similar to the proof of Lemma IV.2, that for all $\mathbf{c}_1,\mathbf{c}_2\in\mathbb{R}^n$, $|h(\mathbf{c}_1) - h(\mathbf{c}_2)| \leq \bar{\mathbf{F}}|\mathbf{c}_1 - \mathbf{c}_2|$, where the inequality is entrywise. Thus, for any $\mathbf{x},\hat{\mathbf{x}}\in\mathbb{R}^n$,

$$
\begin{aligned}
&\big\|[\mathbf{W}_1\mathbf{x} + \mathbf{W}_2 h(\mathbf{W}_3\mathbf{x} + \mathbf{w}) + \mathbf{c}]_{\mathbf{0}}^{\mathbf{m}}\\
&\qquad - [\mathbf{W}_1\hat{\mathbf{x}} + \mathbf{W}_2 h(\mathbf{W}_3\hat{\mathbf{x}} + \mathbf{w}) + \mathbf{c}]_{\mathbf{0}}^{\mathbf{m}}\big\|_{\boldsymbol{\alpha}}\\
&= \boldsymbol{\alpha}^T\big|[\mathbf{W}_1\mathbf{x} + \mathbf{W}_2 h(\mathbf{W}_3\mathbf{x} + \mathbf{w}) + \mathbf{c}]_{\mathbf{0}}^{\mathbf{m}}\\
&\qquad - [\mathbf{W}_1\hat{\mathbf{x}} + \mathbf{W}_2 h(\mathbf{W}_3\hat{\mathbf{x}} + \mathbf{w}) + \mathbf{c}]_{\mathbf{0}}^{\mathbf{m}}\big|\\
&\leq \boldsymbol{\alpha}^T\big|\mathbf{W}_1(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{W}_2(h(\mathbf{W}_3\mathbf{x} + \mathbf{w}) - h(\mathbf{W}_3\hat{\mathbf{x}} + \mathbf{w}))\big|\\
&\leq \boldsymbol{\alpha}^T\big(|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|\big)|\mathbf{x} - \hat{\mathbf{x}}|\\
&= \rho\big(|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|\big)\boldsymbol{\alpha}^T|\mathbf{x} - \hat{\mathbf{x}}|\\
&= \rho\big(|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|\big)\|\mathbf{x} - \hat{\mathbf{x}}\|_{\boldsymbol{\alpha}}.
\end{aligned}
$$

This proves that $\mathbf{x}\mapsto[\mathbf{W}_1\mathbf{x} + \mathbf{W}_2 h(\mathbf{W}_3\mathbf{x} + \mathbf{w}) + \mathbf{c}]_{\mathbf{0}}^{\mathbf{m}}$ is a contraction (on $\mathbb{R}^n_{\geq 0}$ if $\mathbf{m} = \infty\mathbf{1}_n$ or on $\prod_i[0,m_i]$ if $\mathbf{m} < \infty\mathbf{1}_n$) and has a unique fixed point, denoted $\mathbf{x}^*$, by the Banach Fixed-Point Theorem [64, Thm 9.23].

To show GES, let $\boldsymbol{\xi}(t) \triangleq (\mathbf{x}(t) - \mathbf{x}^*)e^t$, satisfying

$$
\tau\dot{\boldsymbol{\xi}}(t) = \mathbf{M}(t)\mathbf{W}\boldsymbol{\xi}(t), \tag{21}
$$

where $\mathbf{M}(t)$ is a diagonal matrix with diagonal entries

$$
M_{ii}(t) \triangleq \frac{\big([\mathbf{W}_1\mathbf{x}(t) + \mathbf{W}_2 h(\mathbf{W}_3\mathbf{x}(t) + \mathbf{w}) + \mathbf{c}]_{\mathbf{0}}^{\mathbf{m}} - \mathbf{x}^*\big)_i}{\xi_i(t)}
$$

if $\xi_i(t)\neq 0$ and $M_{ii}(t) \triangleq 0$ otherwise. Then

$$
|\mathbf{M}(t)| \leq |\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|, \qquad \forall t\geq 0,
$$

where the inequality is entry-wise. Then, by using [65, Lemma] (which is essentially a careful application of Gronwall-Bellman's Inequality [45, Lemma A.1] to (21)),

$$
\begin{aligned}
&\|\boldsymbol{\xi}(t)\|_{\boldsymbol{\alpha}} \leq \|\boldsymbol{\xi}(0)\|_{\boldsymbol{\alpha}} e^{\rho(|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|)t}\\
&\Rightarrow \|\mathbf{x}(t) - \mathbf{x}^*\|_{\boldsymbol{\alpha}} \leq \|\mathbf{x}(0) - \mathbf{x}^*\|_{\boldsymbol{\alpha}} e^{-(1-\rho(|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|))t},
\end{aligned}
$$

establishing GES by the equivalence of norms on $\mathbb{R}^n$. ∎

[9] If $|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3|$ is not irreducible, it can be "upper-bounded" by the irreducible matrix $|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3| + \mu\mathbf{1}_n\mathbf{1}_n^T$, with $\mu > 0$ sufficiently small such that $\rho(|\mathbf{W}_1| + |\mathbf{W}_2|\bar{\mathbf{F}}|\mathbf{W}_3| + \mu\mathbf{1}_n\mathbf{1}_n^T) < 1$. The same argument can then be employed for this upper bound.

Note that Theorem IV.4 applies to each layer of (3) separately. When put together, Theorem IV.3*(ii)* is satisfied if

$$\begin{aligned}
&\rho\big(|\mathbf{W}^{\mathbf{11}}_{2,2}| + |\mathbf{W}^{\mathbf{11}}_{2,3}|\bar{\mathbf{F}}^{\mathbf{1}}_{3}|\mathbf{W}^{\mathbf{11}}_{3,2}|\big) < 1,\\
&\qquad\vdots\\
&\rho\big(|\mathbf{W}^{\mathbf{11}}_{N-1,N-1}| + |\mathbf{W}^{\mathbf{11}}_{N-1,N}|\bar{\mathbf{F}}^{\mathbf{1}}_{N}|\mathbf{W}^{\mathbf{11}}_{N,N-1}|\big) < 1,\\
&\rho\big(|\mathbf{W}^{\mathbf{11}}_{N,N}|\big) < 1,
\end{aligned} \tag{22}$$

where $\bar{\mathbf{F}}^{\mathbf{1}}_{i}, i = 2,\dots,N$ is the matrix described in Theorem IV.4 corresponding to $h^{\mathbf{1}}_{i}$, and the affine form (20) of $h^{\mathbf{1}}_{i}$ is computed recursively using Lemma IV.1. Moreover, if $\mathbf{m}^{\mathbf{1}}_{1} = \infty\mathbf{1}_{r_1}$, then $\rho\big(|\mathbf{W}^{\mathbf{11}}_{1,1}| + |\mathbf{W}^{\mathbf{11}}_{1,2}|\bar{\mathbf{F}}^{\mathbf{1}}_{2}|\mathbf{W}^{\mathbf{11}}_{2,1}|\big) < 1$ serves as a sufficient condition for Theorem IV.3*(i)* (which is trivial if $\mathbf{m}^{\mathbf{1}}_{1} < \infty\mathbf{1}_{r_1}$).

V. Case Study: Selective Listening in Rodents

We present an application of our framework to a specific real-world example of goal-driven selective attention using measurements of single-neuron activity in the brain. Beyond the conceptual illustration of our results in Example III.4 above, we argue that the cross-validation of theoretical results with real data performed here is a necessary step to make a credible case for neuroscience research and significantly enhances the relevance of the developed analysis. We have been fortunate to have access to data from a novel and carefully designed experimental paradigm [13], [66] that involves goal-driven selective listening in rodents and displays the key features of hierarchical selective recruitment noted here.

A. Description of Experiment and Data

A long standing question in neuroscience involves our capability to selectively listen to specific sounds in a crowded environment [2], [67]. To understand the neuronal basis of this phenomena, the work [13] has rats simultaneously presented with two sounds and trains them to selectively respond to one sound while actively suppressing the distraction from the other. In each trial, the animal simultaneously hears a white noise burst and a narrow-band warble. The noise burst may come from the left or the right while the warble may have low or high pitch, both chosen at random. Which of the two sounds (noise burst or warble) is relevant and which is a distraction depends on the "rule" of the trial: in "localization" (LC) and "pitch discrimination" (PD) trials, the animal has to make a motor choice based on the location of the noise burst (left/right) or the pitch of the warble (low/high), resp., to receive a reward. Each rat performs several blocks of LC and PD trials during each session (with each block switching randomly between the 4 possible stimulus pairs), requiring it to quickly switch its response following the rule changes.

While the rats perform the task, spiking activity of single neurons is recorded in two brain areas: the primary auditory cortex (A1) and the medial prefrontal cortex (PFC). A1 is the first region in the cortex that receives auditory information (from subcortical areas and ears), thus forming a (relatively) low level of the hierarchy. PFC is composed of multiple regions that form the top of the hierarchy, and serve functions such as imagination, planning, decision-making, and attention [68]. Spike times of 211 well-isolated and reliable neurons are recorded in 5 rats, 112 in PFC and 99 in A1, see [66].

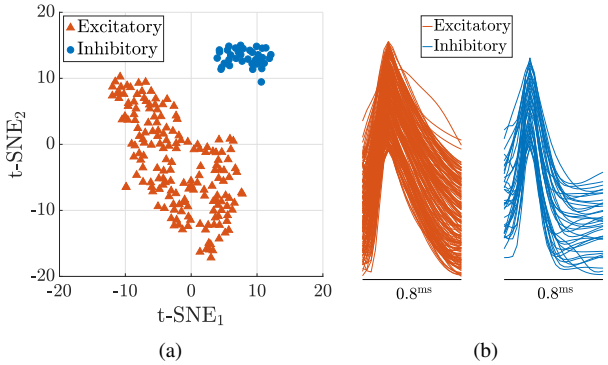

Fig. 3: Excitatory/inhibitory classification of neurons. (a) Clusters of spike waveforms. For illustration, clusters are shown in the two-dimensional space arising from t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction. (b) The spike waveforms of clustered neurons. As expected, the inhibitory neurons have faster and narrower spikes.

Using statistical analysis, it was shown in [13] that (i) the rule of the trial and the stimulus sounds are more strongly encoded by PFC and A1 neurons, resp., (ii) electrical disruption of PFC significantly impairs task performance, and (iii) PFC activity temporally precedes A1 activity. These findings are all consistent with a model where PFC controls the activity of A1 based on the trial rule in order to achieve GDSA. We next build on these observations to define an appropriate network structure and rigorously analyze it using HSR.

B. Choice of Neuronal Populations

To form meaningful populations among the recorded neurons, we perform three classifications of them:

(i) first, we classify the neurons into excitatory and inhibitory. The procedure for this classification is based on the neuron's spike waveform: excitatory neurons have slower and wider spikes while inhibitory neurons have faster and narrower ones [69]. Using standard k-means clustering on the 24-dimensional spike waveform time-series, we identify 174 excitatory and 36 inhibitory neurons [10] (Figure 3(a)). These clusters conform with spike width difference of excitatory and inhibitory neurons (Figure 3(b)) and the common expectation that about 80% of mammalian cortical neurons are excitatory.

(ii) Second, we classify the PFC neurons based on their rule-encoding (RE) property. This classification was also done in [13], so we briefly review the method for completeness. A neuron is said to have a RE property if its firing rate is significantly different during the LC and PD trials *before the stimulus onset*. In the absence of stimulus, any such difference is attributable to the animal's knowledge of the task rule (i.e., which upcoming stimulus it has to pay attention to in order to get the reward). Thus, it is standard to assess neurons' RE property during the *hold period*, namely, the time interval

[10]The type of one neuron could not be identified with confidence and was discarded from further analysis.

between the initiation of each trial and the stimulus onset of that trial. Therefore for each PFC neuron, we calculate its mean firing rate during the hold period of each trial and then statistically compare the results for LC and PD trials ($p < 0.05$, randomization test). Among the 112 neurons in PFC, 40 encoded for LC while 44 encoded for PD (the remaining PFC neurons with no RE property are discarded).

(iii) Finally, we classify the A1 neurons based on their evoked response (ER) property. In contrast to RE, a neuron has an ER property if its firing rate is significantly different in response to the white noise (LC stimulus) and warble (PD stimulus) *after the stimulus onset*. Since the white noise and warble are always presented simultaneously, it is not possible to make such a distinction based on normal trials. However, before each LC or PD block, the animal is only presented with the respective stimulus for a few *cue trials* (which is how the animal realizes the rule change). Thus, for each A1 neuron, we compare its mean firing rate during the *listening period* of each cue trial (namely, the interval between the stimulus onset and the time that the animal commits to a decision) and statistically compare the distribution of the results for LC and PD cue trials ($p < 0.05$, randomization test). Among the 99 A1 neurons, 21 had an ER for LC while another 21 had an ER for PD (the remaining A1 neurons with no ER property are discarded from further analysis).

Remark V.1. ***(RE vs. ER detection).*** It is noteworthy that a smaller fraction of PFC and A1 neurons also have ER and RE properties, resp. However, it is expected from systems neuroscience that these properties arise from the PFC-A1 interaction, as auditory and attention/decision making information disseminate from A1 and PFC, resp. This motivates our classification of A1 and PFC neurons based on ER and RE, resp., and their reciprocal connection in the proposed network structure below. Further, we note that our ER detection has a difference with respect to [13]. In [13], the difference between the post-stimulus and pre-stimulus firing rates (the latter being RE) is used for ER detection, with the motivation of removing the portion of post-stimulus firing rate that is due to RE (and thus independent of stimulus). However, this relies on the strong assumption that the RE and ER responses superimpose linearly, which we found likely not to be true based on the statistical analysis of the present dataset, perhaps as RE may have driven neurons close to their maximum firing rate, leaving little room for *additional* ER. We thus use the complete post-stimulus firing rate for ER detection, as above. □

As a result of the classifications described above, we group the neurons into 8 populations based on the PFC/A1, excitatory/inhibitory, and LC/PD classifications. The firing rate of each population (as a function of time) is then calculated as follows. For each neuron and each trial, the interval $[-10, 10]$ (with time 0 corresponding to stimulus onset) is decomposed into 100^{ms}-wide bins and the firing rate of each bin (spike count divided by bin width) is assigned to the bin's center time. This time series is then averaged over all trials with the same stimulus pair and all the neurons within each population, and finally smoothed with a Gaussian kernel with 1^{s} standard deviation. This results in one firing rate time series for each neuron and each stimulus pair.

We limit our choice of stimulus pairs as follows. Recall that each of LC and PD blocks contains 4 stimulus pairs (left-low, left-high, right-low, right-high). In each block, these 4 pairs are divided into two *go* and two *no-go* pairs. When the animal hears a go stimulus pair, his correct response is to go to a nearby food port to receive his reward. In no-go trials, the correct response is simply inaction (action is punished by a delay before the animal can do the next trial). Due to strong motor and reward-consumption artifacts in go trials (cf. [13, Fig. S4]), we limit our analysis here to no-go trials. Further, we also discard the no-go stimulus pair that is shared between LC and PD blocks, since the correct decision (no-go) is independent of the block and thus does not require selective attention. Hence, our analysis only involves one firing rate time series for each neuronal population in each block.

C. Network Binary Structure

We next describe our proposed network binary structure[11]. In each of the two regions (PFC and A1), the 4 populations are connected to each other according to the following physiological properties (see [70]–[72] and [72]–[74] for evidence of these properties in PFC and A1, resp.):

(i) each excitatory population projects to (i.e., makes synapses on) the inhibitory population with the same LC/PD preference (RE in PFC or ER in A1);
(ii) neurons in each excitatory population project to each other (captured by the excitatory self-loops in Figure 4).
(iii) each inhibitory population projects to the populations (both excitatory and inhibitory) with *opposite* LC/PD preference (the so-called *lateral inhibition* property);

While within-region connections are both excitatory and inhibitory, between-region connections in the cortex (including PFC and A1) are almost entirely excitatory, completing the binary structure shown in Figure 4.

Hierarchical Structure: To apply the HSR framework to the network of Figure 4, we still need to assign the nodes to hierarchical layers. This assignment is in general arbitrary except for two critical requirements, (i) the existence of timescale separation between layers and (ii) the existence of both excitatory and inhibitory projections from any layer to the layer below (to allow for simultaneous inhibition and recruitment). The trivial choice here is to consider each region as a layer, which also satisfies (i) (since PFC has slower dynamics than A1) but not (ii) (since there would be no inhibitory connection between regions). We thus propose an alternative 3-layer choice, as shown in Figure 4.[12] This choice clearly satisfies (ii), and we next show that it also satisfies (i).

Computation of Timescales: To assess the intrinsic timescales of each population, we employ the common method in neuroscience based on the decay rate of the correlation coefficient [20], [21]. In brief, for each neuron ℓ, we partition

[11] We here make a distinction between the binary structure of the network, composed of only the connectivity pattern among nodes, and its full structure, that also includes the connection weights.

[12] The bottom-most layer $\mathcal{N}_4$ represents "external" inputs from sub-cortical areas. Since we have no recordings from these areas, we do not consider any dynamics for $\mathcal{N}_4$ and accordingly do not include it in HSR analysis.

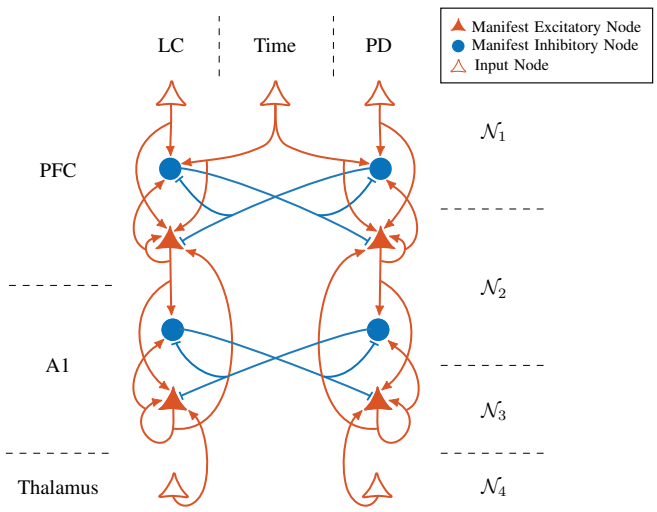

Fig. 4: Proposed network binary structure. The physiological region, hierarchical layer, and encoding properties of nodes are indicated on the left, right, and above the figure, resp.

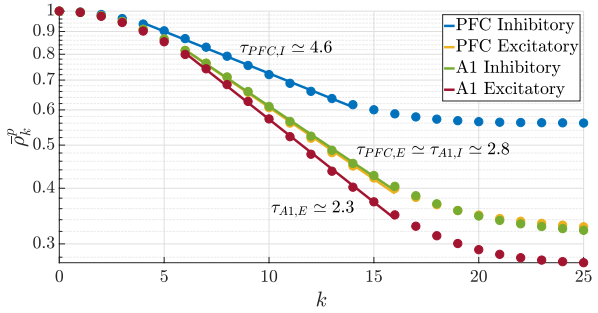

Fig. 5: Timescale separation among the layers $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_3$ in Figure 4. The circles illustrate the values of the average auto-correlation coefficient $\bar{\rho}^p_k$ as a function of time lag k, whereas the lines represent the best exponential fit over the range of time lags where each $\bar{\rho}^p_k$ decays exponentially (note the logarithmic scale on the y-axis).

the time window *before* the stimulus onset[13] into small bins (200[ms]-wide here) and compute the smoothed mean firing rate of this neuron during each bin and each trial. This yields a set $\{r^\ell_{i,k}\}_{i,k,\ell}$, where $r^\ell_{i,k}$ denotes the mean firing rate of neuron ℓ in the k'th time bin of trial i. The Pearson correlation coefficient between two time bins k_1 and k_2 is estimated as

$$\rho^\ell_{k_1,k_2} = \frac{\sum_i (r^\ell_{i,k_1} - \bar{r}^\ell_{k_1})(r^\ell_{i,k_2} - \bar{r}^\ell_{k_2})}{\sqrt{\sum_i (r^\ell_{i,k_1} - \bar{r}^\ell_{k_1})^2 \sum_i (r^\ell_{i,k_2} - \bar{r}^\ell_{k_2})^2}} \in [-1,1],$$

where $\bar{r}^\ell_k$ is the average of $r^\ell_{i,k}$ across all the trials for neuron ℓ. Let ρ^ℓ_k be the average of $\rho^\ell_{k_1,k_2}$ over all k_1, k_2 such that $|k_1 - k_2| = k$ and $\bar{\rho}^p_k$, for any population p, be the average of ρ^ℓ_k for all the neurons ℓ in the population p. Figure 5 shows this function for populations of excitatory and inhibitory neurons in PFC and A1 (we do not split the neurons based on their LC/PD preference because it is not relevant for timescale separation). Fitting $\bar{\rho}^p_k$ by an exponential function of the form $Ae^{-k/\tau}$ gives an estimate of the intrinsic timescale τ of this population, which becomes exact for spikes generated by a Poisson point process under certain regularity conditions [20]. Here, we use the range of k values for which the decay of $\bar{\rho}^p_k$ is approximately exponential for calculating the fit. As seen in Figure 5, there is a clear timescale separation between the layer of A1 excitatory neurons, the layer of A1 inhibitory and PFC excitatory neurons, and the layer of PFC inhibitory neurons, satisfying the requirement (i) above.[14]

Exogenous Inputs and Latent Nodes: The last step in specifying the binary structure of the network involves the exogenous inputs to the prescribed neuronal populations (nodes). Clearly, nodes at the bottom layer (layer 3) receive auditory inputs from subcortical areas which we represent as two input signals x^4_1 and x^4_2 coming from layer 4 and corresponding to the white noise and warble, resp. Both these signals are constructed by smoothing a square pulse that equals 1 during stimulus presentation and 0 otherwise with the same Gaussian window used for smoothing the firing rate time-series.

The choice of the inputs to the PFC populations is more intricate. PFC is itself composed of a complex network of several regions, each involved in some aspects of high-level cognitive functions. The RE properties of the recorded PFC populations is only one outcome of such complex PFC dynamics that also host the animal's overall understanding of how the task works, his perception of time, etc. In order to capture the effects of such unrecorded PFC dynamics, we consider 3 additional excitatory PFC populations, as follows. Two input populations x^1_3 and x^1_4 simply encode the rule of each block[15]:

$$x^1_3 \equiv \begin{cases} 1, & \text{if in LC block,} \\ 0, & \text{if in PD block,} \end{cases} \qquad x^1_4 \equiv \begin{cases} 0, & \text{if in LC block,} \\ 1, & \text{if in PD block.} \end{cases}$$

Populations with such a sustained constant activity only as a function of task parameters are indeed observed during GDSA in PFC [75]. The third additional PFC population encodes the time relative to the stimulus onset, which is critical for the functioning of the recorded PFC populations. Among the various forms of encoding time, we consider a population x^1_5 with firing rate

$$x^1_5(t) = \begin{cases} |t_0| - t & t \in [t_0, 0), \\ 0 & t \in (0, t_f], \end{cases}$$

where $[t_0, t_f] = [-7, 7]$ is the duration of each trial, since populations with such activity patterns have been observed in PFC [76].[16] Since these three populations have very slow dynamics but are excitatory, following the same logic as before, we position them in the layer 1 together with the recorded inhibitory PFC populations x^1_1, x^1_2.

Finally, to capture the effects of the large populations of neurons whose activity is not recorded, we consider one *latent*

[13] In general, the time interval used for timescale estimation should not include stimulus presentation in order to reduce the effects of external factors on the internal neuronal dynamics.

[14] Note that this method inherently underestimates the timescale separation between layers due to the mutual dynamical interactions between them.

[15] Note that this static response is different from, and much simpler than, the RE of the recorded PFC neurons, which is greatly dynamic.

[16] Even though both [75] and [76] involve primates, populations with similar activity patterns are expected to exist in rodents.

node for each of the 8 *manifest* nodes in the network[17] with the same in- and out-neighbors as their respective manifest node (latent nodes are not plotted in Figure 4 to avoid cluttering the network structure). We let $\{x_{1,j}\}_{j=6,7}$, $\{x_{2,j}\}_{j=5}^{8}$, and $\{x_{3,j}\}_{j=3,4}$ denote these nodes in $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_3$, resp.

D. Identification of Network Parameters

Having established the binary structure of the network, we next seek to determine its unknown parameters $\mathbf{W}^{i,j}$. While there are physiological methods for measuring the synaptic weight between a pair of neurons in vitro, they are not applicable in vivo and thus not available for our dataset. Also, our nodes consist of several neurons, making their aggregate synaptic weight an abstract quantity. Therefore, we resort to system identification/machine learning techniques to "learn" the structure of the network given its input-output signals. For this purpose, the choice of objective function is crucial, for which we propose

$$f(z) = f_{\text{SSE}}(z) + \gamma_1 f_{\text{corr}}(z) + \gamma_2 f_{\text{var}}(z), \tag{23}$$

$$f_{\text{SSE}}(z) = \sum_{\ell=1}^{2}\sum_{i=1}^{3}\sum_{j=1}^{n_{m,i}}\sum_{k}(\hat{x}_{i,j}(kT;\ell) - x_{i,j}(kT;\ell))^2,$$

$$f_{\text{corr}}(z) = 1 - \frac{1}{2n_m}\sum_{\ell=1}^{2}\frac{1}{n_m}\sum_{i=1}^{3}\sum_{j=1}^{n_{m,i}}\frac{1}{K-1} \times \sum_{k=1}^{K}\frac{(\hat{x}_{i,j}(kT;\ell) - \hat{\mu}_{i,j,\ell})(x_{i,j}(kT;\ell) - \mu_{i,j,\ell})}{\hat{\sigma}_{i,j,\ell}\sigma_{i,j,\ell}},$$

$$f_{\text{var}}(z) = \Big(\sum_{\ell=1}^{2}\sum_{i=1}^{3}\sum_{j=1}^{n_{m,i}}(\hat{\sigma}_{i,j,\ell} - \sigma_{i,j,\ell})^4\Big)^{1/4},$$

where,

– z is the vector of all unknown network parameters consisting of not only the synaptic weights but also the time constants τ_i, the background inputs $\mathbf{c}_i$, and the initial states $\mathbf{x}_i(0), i = 1, 2, 3$;

– $n_{m,i}$ is the number of manifest nodes in layer i (so $n_{m,1} = 2, n_{m,2} = 4, n_{m,3} = 2$) and $n_m = 8$ is the total number of manifest nodes;

– $x_{i,j}(t;\ell)$ is the measured state of j'th node in the i'th layer in response to the ℓ'th stimulus at time t (where $\ell = 1$ indicates the LC block and $\ell = 2$ the PD block) and $\hat{x}_{i,j}(t;\ell)$ is its model estimate;

– $T = 0.1$ is the sampling time and K is the total number of samples of each signal; and

– $\mu_{i,j,\ell}, \sigma_{i,j,\ell}, \hat{\mu}_{i,j,\ell}, \hat{\sigma}_{i,j,\ell}$ are the means and standard deviations of $x_{i,j}(\cdot;\ell)$ and $\hat{x}_{i,j}(\cdot;\ell)$, resp.

The rationale behind (23) is as follows. $f_{\text{SSE}}(z)$ is the standard sum of squared error (SSE). In HSR, an important property of nodal state trajectories is the sign of their derivatives, which *transiently* indicate recruitment (positive derivative) or inhibition (negative derivative). This is captured by the average correlation coefficient $f_{\text{corr}}(z)$, which is added to $f_{\text{SSE}}(z)$ to enforce similar recruitment and inhibition patterns between measured states and their estimates. Nevertheless, correlation coefficient between a pair of signals is invariant to the amount of variation in them, requiring us to add the third term $f_{\text{var}}(z)$. The use of 4-norm in $f_{\text{var}}(z)$ particularly weights the nodes with large standard deviation mismatches. Appropriate weights $\gamma_1 = 250$ and $\gamma_2 = 150$ were found via trial and error.

[17]A node is *manifest* if its activity is recorded during the experiment and *latent* otherwise.

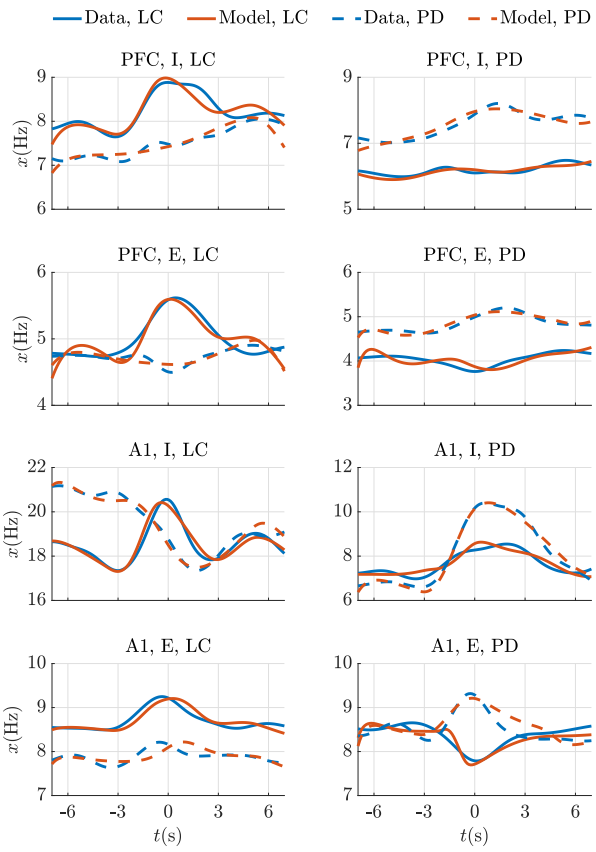

Fig. 6: State trajectories of manifest nodes in the network of Figure 4 (blue: measured, red: model estimate). $t = 0$ indicates stimulus onset. Solid and dashed lines correspond to LC and PD blocks, resp. The description of each node is indicated above its corresponding panel. The LC/PD in the legend refers to the trial rule, while the LC/PD above each panel refers to the preference of that particular node.

The objective function f is highly nonconvex and we thus use the GlobalSearch algorithm from the MATLAB Optimization Toolbox to minimize it. Figure 6 shows the manifest nodal states as well as their best model estimates. In order to quantify the similarity between these states and their estimates, we use the standard R^2 measure given by

$$R^2 = 1 - \frac{\sum_{\ell,i,j,k}(x_{i,j}(kT;\ell) - \hat{x}_{i,j}(kT;\ell))^2}{\sum_{\ell,i,j,k}(x_{i,j}(kT;\ell) - \mu_{i,j,\ell})^2} \simeq 93.6\%.$$

This high value is indeed remarkable, especially given the small network size and the limited availability of measurements in the experiment (2240 data points, 175 parameters).

E. Concurrence of the Identified Network with Analysis

To conclude, we verify here whether the identified network structure satisfies the requirements of the HSR framework

in terms of timescale separation and stability. Regarding the former, the identified time constants are given by

$$\tau_1 = 3.36, \qquad \tau_2 = 1.68, \qquad \tau_3 = 0.70,$$

yielding an almost twofold separation of timescales conforming to Figure 5. Regarding stability, we have to consider the LC and PD blocks separately (as the definition of task-relevant (^{1}) and task-irrelevant ($^\circ$) nodes changes according to the block).

In the LC block, the (manifest) LC nodes are task-relevant and the (manifest) PD nodes are task-irrelevant. Therefore, under this condition,

$$W^{11}_{3,3} = 0.01, \qquad \mathbf{W}^{11}_{3,2} = \begin{bmatrix} 0.01 & 0 \end{bmatrix},$$
$$\mathbf{W}^{11}_{2,2} = \begin{bmatrix} 0.83 & 0 \\ 0.76 & 0 \end{bmatrix}, \qquad \mathbf{W}^{11}_{2,3} = \begin{bmatrix} 0.04 \\ 0.58 \end{bmatrix}.$$

It is then straightforward to see that

$$h^1_3(c^1_3) = \begin{cases} 0 & ; \quad c^1_3 \leq 0 \\ c^1_3/(1 - W^{11}_{3,3}) & ; \quad c^1_3 \geq 0 \end{cases} \Rightarrow \bar{F}^1_3 = \frac{1}{1 - W^{11}_{3,3}}.$$

Therefore,

$$\rho(|W^{11}_{3,3}|) = 0.01 < 1,$$
$$\rho\big(|\mathbf{W}^{11}_{2,2}| + |\mathbf{W}^{11}_{2,3}|\bar{F}^1_3|\mathbf{W}^{11}_{3,2}|\big) = \rho\Big(\begin{bmatrix} 0.83 & 0 \\ 0.77 & 0 \end{bmatrix} \Big) = 0.83 < 1,$$

satisfying the sufficient conditions for GES in (22). Similarly, in the PD block, we have

$$W^{11}_{3,3} = 0.01 < 1, \qquad \mathbf{W}^{11}_{3,2} = \begin{bmatrix} 4.7 \times 10^{-3} & 0 \end{bmatrix},$$
$$\mathbf{W}^{11}_{2,2} = \begin{bmatrix} 0.12 & 0 \\ 0.56 & 0 \end{bmatrix}, \qquad \mathbf{W}^{11}_{2,3} = \begin{bmatrix} 0.39 \\ 0.02 \end{bmatrix},$$
$$\rho\big(|\mathbf{W}^{11}_{2,2}| + |\mathbf{W}^{11}_{2,3}|\bar{F}^1_3|\mathbf{W}^{11}_{3,2}|\big) = \rho\Big(\begin{bmatrix} 0.12 & 0 \\ 0.56 & 0 \end{bmatrix} \Big) = 0.12 < 1,$$

also satisfying the GES conditions of (22).

While this concurrence is promising, its robustness to the choice of dataset and data processing steps is critical. A comprehensive robustness analysis requires access to multiple datasets and experimental re-design, which is beyond the scope of this case study. However, we repeated our entire analysis with Mann-Whitney-Wilcoxon rank-sum test (used originally in [13]) and also with varying significance thresholds $0.001 \leq \alpha \leq 0.05$ and observed that, despite the change in the neural populations, our theoretical conditions remained satisfied.

Given the concurrence between the identified network structure and the hypotheses of our results, Theorems III.2 and IV.3 provide strong analytical support to explain the conclusions drawn in [13], [66] from experimental data and statistical analysis. We believe HSR constitutes a rigorous framework for the analysis of the multiple-timescale network interactions underlying GDSA, complementing the conventional statistical and computational analyses in neuroscience.

VI. Conclusions and Future Work

We have proposed hierarchical selective recruitment as a framework to explain several fundamental components of goal-driven selective attention. HSR consists of an arbitrary number of neuronal subnetworks that operate at different timescales and are arranged in a hierarchy according to their intrinsic timescales. In this paper, we have resorted to control-theoretic tools to focus on the top-down recruitment of the task-relevant nodes. We have derived conditions on the structure of multi-layer networks guaranteeing the convergence of the state of the task-relevant nodes in each layer towards their reference trajectory determined by the layer above in the limit of maximal timescale separation between the layers. In doing so, we have characterized the piecewise affinity and global Lipschitzness properties of the equilibrium maps and unveiled their key role in the multiple-timescale dynamics of the network. Combined with the results of Part I, these contributions provide conditions for the simultaneous GES of the state of task-irrelevant nodes of all layers to the origin (inhibition) as well as the GES of the state of task-relevant nodes towards an equilibrium that moves at a slower timescale as a function of the state of the subnetwork at the layer above (recruitment). To demonstrate that applicability to brain networks, we have presented a detailed case study of GDSA in rodents and showed that a network with a binary structure based on HSR and parameters learned using a carefully designed optimization procedure can achieve remarkable accuracy in explaining the data while conforming to the theoretical requirements of HSR. Our technical treatment has also established a novel converse Lyapunov theorem for continuous GES switched affine systems with state-dependent switching. Future work will include the extension of this framework to selective inhibition using output feedback and cases where the recruited subnetworks are asymptotically stable towards more complex attractors such as limit cycles. Also of paramount importance is the study of the robustness of network trajectories as well as the theoretical conditions of HSR to network parameters, disturbances, and experimental variations (inter-subject variability, different tasks, measurement noise, etc.). Other topics of relevance to the understanding of GDSA that we plan to explore are the analysis of the information transfer along the hierarchy, the controllability and observability of linear-threshold networks, and the optimal sensor and actuator placement in hierarchical interconnections of these networks.

Acknowledgments

We would like to thank Dr. Erik J. Peterson for piquing our interest with his questions on dimensionality control in brain networks and for introducing us to linear-threshold modeling in neuroscience. We are also indebted to Drs. Michael R. DeWeese and Chris C. Rodgers for the public release of their dataset [66] and their subsequent discussion of its details. This work was supported by NSF Award CMMI-1826065 (EN and JC) and ARO Award W911NF-18-1-0213 (JC).

References

[1] E. Nozari and J. Cortés, "Selective recruitment in hierarchical complex dynamical networks with linear-threshold rate dynamics," in *IEEE Conf. on Decision and Control*, Miami Beach, FL, Dec. 2018, pp. 5227–5232.

[2] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[3] A. M. Treisman, "Strategies and models of selective attention." *Psychological review*, vol. 76, no. 3, p. 282, 1969.

[4] J. Moran and R. Desimone, "Selective attention gates visual processing in the extrastriate cortex," *Science*, vol. 229, no. 4715, pp. 782–784, 1985.

[5] B. C. Motter, "Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli," *Journal of Neurophysiology*, vol. 70, no. 3, pp. 909–919, 1993.

[6] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual Review of Neuroscience*, vol. 18, no. 1, pp. 193–222, 1995.

[7] S. Kastner, P. DeWeerd, R. Desimone, and L. G. Ungerleider, "Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI," *Science*, vol. 282, no. 5386, pp. 108–111, 1998.

[8] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, p. 194, 2001.

[9] M. A. Pinsk, G. M. Doniger, and S. Kastner, "Push-pull mechanism of selective attention in human extrastriate cortex," *Journal of Neurophysiology*, vol. 92, no. 1, pp. 622–629, 2004.

[10] N. Lavie, "Distracted and confused?: Selective attention under load," *Trends in Cognitive Sciences*, vol. 9, no. 2, pp. 75–82, 2005.

[11] J. J. Foxe and A. C. Snyder, "The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention," *Frontiers in Psychology*, vol. 2, p. 154, 2011.

[12] A. Gazzaley and A. C. Nobre, "Top-down modulation: bridging selective attention and working memory," *Trends in Cognitive Sciences*, vol. 16, no. 2, pp. 129–135, 2012.

[13] C. C. Rodgers and M. R. DeWeese, "Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents," *Neuron*, vol. 82, no. 5, pp. 1157–1170, 2014.

[14] M. Gomez-Ramirez, K. Hysaj, and E. Niebur, "Neural mechanisms of selective attention in the somatosensory system," *Journal of neurophysiology*, vol. 116, no. 3, pp. 1218–1231, 2016.

[15] U. Hasson, E. Yang, I. Vallines, D. J. Heeger, and N. Rubin, "A hierarchy of temporal receptive windows in human cortex," *Journal of Neuroscience*, vol. 28, no. 10, pp. 2539–2550, 2008.

[16] C. J. Honey, T. Thesen, T. H. Donner, L. J. Silbert, C. E. Carlson, O. Devinsky, W. K. Doyle, N. Rubin, D. J. Heeger, and U. Hasson, "Slow cortical dynamics and the accumulation of information over long timescales," *Neuron*, vol. 76, no. 2, pp. 423–434, 2012.

[17] B. Gauthier, E. Eger, G. Hesselmann, A. Giraud, and A. Kleinschmidt, "Temporal tuning properties along the human ventral visual stream," *Journal of Neuroscience*, vol. 32, no. 41, pp. 14 433–14 441, 2012.

[18] U. Hasson, J. Chen, and C. J. Honey, "Hierarchical process memory: memory as an integral component of information processing," *Trends in Cognitive Sciences*, vol. 19, no. 6, pp. 304–313, 2015.

[19] M. G. Mattar, D. A. Kahn, S. L. Thompson-Schill, and G. K. Aguirre, "Varying timescales of stimulus integration unite neural adaptation and prototype formation," *Current Biology*, vol. 26, no. 13, pp. 1669–1676, 2016.

[20] J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, and X. Wang, "A hierarchy of intrinsic timescales across primate cortex," *Nature Neuroscience*, vol. 17, no. 12, p. 1661, 2014.

[21] R. Chaudhuri, K. Knoblauch, M. Gariel, H. Kennedy, and X. Wang, "A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex," *Neuron*, vol. 88, no. 2, pp. 419–431, 2015.

[22] N. Tinbergen, "The hierarchical organization of nervous mechanisms underlying instinctive behaviour," in *Symposium for the Society for Experimental Biology*, vol. 4, no. 305-312, 1950.

[23] A. R. Luria, "The functional organization of the brain," *Scientific American*, vol. 222, no. 3, pp. 66–79, 1970.

[24] D. J. Felleman and D. C. V. Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[25] A. Krumnack, A. T. Reid, E. Wanke, G. Bezgin, and R. Kötter, "Criteria for optimizing cortical hierarchies with continuous ranges," *Frontiers in Neuroinformatics*, vol. 4, p. 7, 2010.

[26] G. Zamora-López, C. Zhou, and J. Kurths, "Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks," *Frontiers in Neuroinformatics*, vol. 4, p. 1, 2010.

[27] N. T. Markov, J. Vezoli, P. Chameau, A. Falchier, R. Quilodran, C. Huissoud, C. Lamy, P. Misery, P. Giroud, S. Ullman, P. Barone, C. Dehay, K. Knoblauch, and H. Kennedy, "Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex," *Journal of Comparative Neurology*, vol. 522, no. 1, pp. 225–259, 2014.

[28] P. Lennie, "Single units and visual cortical organization," *Perception*, vol. 27, no. 8, pp. 889–935, 1998.

[29] D. Badre and M. D'Esposito, "Is the rostro-caudal axis of the frontal lobe hierarchical?" *Nature Reviews Neuroscience*, vol. 10, no. 9, pp. 659–669, 2009.

[30] J. P. Gilman, M. Medalla, and J. I. Luebke, "Area-specific features of pyramidal neurons-a comparative study in mouse and rhesus monkey," *Cerebral Cortex*, vol. 27, no. 3, pp. 2078–2094, 2016.

[31] C. Cioli, H. Abdi, D. Beaton, Y. Burnod, and S. Mesmoudi, "Differences in human cortical gene expression match the temporal properties of large-scale functional networks," *PLOS One*, vol. 9, no. 12, p. e115913, 2014.

[32] C. A. Runyan, E. Piasini, S. Panzeri, and C. D. Harvey, "Distinct timescales of population coding across cortex," *Nature*, vol. 548, no. 7665, p. 92, 2017.

[33] S. J. Kiebel, J. Daunizeau, and K. J. Friston, "A hierarchy of time-scales and the brain," *PLOS Computational Biology*, vol. 4, no. 11, p. e1000209, 2008.

[34] Y. Yamashita and J. Tani, "Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment," *PLOS Computational Biology*, vol. 4, no. 11, p. e1000220, 2008.

[35] D. S. Bassett, E. T. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger, and A. Meyer-Lindenberg, "Hierarchical organization of human cortical networks in health and schizophrenia," *Journal of Neuroscience*, vol. 28, no. 37, pp. 9239–9248, 2008.

[36] D. Meunier, R. Lambiotte, A. Fornito, K. Ersche, and E. T. Bullmore, "Hierarchical modularity in human brain functional networks," *Frontiers in Neuroinformatics*, vol. 3, p. 37, 2009.

[37] D. Meunier, R. Lambiotte, and E. T. Bullmore, "Modular and hierarchically modular organization of brain networks," *Frontiers in Neuroscience*, vol. 4, p. 200, 2010.

[38] Z. Zhen, H. Fang, and J. Liu, "The hierarchical brain network for face recognition," *PLOS One*, vol. 8, no. 3, p. e59886, 2013.

[39] P. Lakatos, A. S. Shah, K. H. Knuth, I. Ulbert, G. Karmos, and C. E. Schroeder, "An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex," *Journal of Neurophysiology*, vol. 94, no. 3, pp. 1904–1911, 2005.

[40] E. Nozari and J. Cortés, "Hierarchical selective recruitment in linear-threshold brain networks. Part I: Intra-layer dynamics and selective inhibition," *IEEE Transactions on Automatic Control*, 2018, submitted. Available at https://arxiv.org/abs/1809.01674.

[41] Y. Yeshurun and M. Carrasco, "Attention improves or impairs visual performance by enhancing spatial resolution," *Nature*, vol. 396, no. 6706, p. 72, 1998.

[42] J. B. Fritz, M. Elhilali, S. V. David, and S. A. Shamma, "Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in a1?" *Hearing research*, vol. 229, no. 1-2, pp. 186–203, 2007.

[43] K. Anton-Erxleben, V. M. Stephan, and S. Treue, "Attention reshapes center-surround receptive field structure in macaque cortical area mt," *Cerebral Cortex*, vol. 19, no. 10, pp. 2466–2478, 2009.

[44] A. N. Tikhonov, "Systems of differential equations containing small parameters in the derivatives," *Matematicheskii Sbornik*, vol. 73, no. 3, pp. 575–586, 1952.

[45] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.

[46] A. B. Vasilieva, "On the development of singular perturbation theory at Moscow State University and elsewhere," *SIAM Review*, vol. 36, no. 3, pp. 440–452, 1994.

[47] D. Naidu, "Singular perturbations and time scales in control theory and applications: an overview," *Dynamics of Continuous Discrete and Impulsive Systems Series B*, vol. 9, pp. 233–278, 2002.

[48] J. K. Kevorkian and J. D. Cole, *Multiple scale and singular perturbation methods*. Springer Science & Business Media, 2012, vol. 114.

[49] R. E. O'Malley, *Singular perturbation methods for ordinary differential equations*. Springer Science & Business Media, 2012, vol. 89.

[50] A. L. Dontchev and V. M. Veliov, "Singular perturbation in mayer's problem for linear systems," *SIAM Journal on Control and Optimization*, vol. 21, no. 4, pp. 566–581, 1983.

[51] A. L. Dontchev and I. I. Slavov, "Upper semicontinuity of solutions of singularly perturbed differential inclusions," in *System Modelling and Optimization*, H. J. Sebastian and K. Tammer, Eds. Springer Berlin Heidelberg, 1990, pp. 273–280.

[52] M. Quincampoix, "Singular perturbations for systems of differential inclusions," *Banach Center Publications*, vol. 32, no. 1, pp. 341–348, 1995.

[53] A. Dontchev, T. Donchev, and I. I. Slavov, "A Tikhonov-type theorem for singularly perturbed differential inclusions," *Nonlinear Analysis, Theory, Methods & Applications*, vol. 26, no. 9, pp. 1547–1554, 1996.

[54] V. Veliov, "A generalization of the Tikhonov theorem for singularly

perturbed differential inclusions," *Journal of Dynamical & Control Systems*, vol. 3, no. 3, pp. 291–319, 1997.
[55] F. Watbled, "On singular perturbations for differential inclusions on the infinite interval," *Journal of Mathematical Analysis and Applications*, vol. 310, no. 2, pp. 362–378, 2005.
[56] G. Grammel, "Exponential stability of nonlinear singularly perturbed differential equations," *SIAM Journal on Control and Optimization*, vol. 44, no. 5, pp. 1712–1724, 2005.
[57] P. Fries, J. H. Reynolds, A. E. Rorie, and R. Desimone, "Modulation of oscillatory neuronal synchronization by selective visual attention," *Science*, vol. 291, no. 5508, pp. 1560–1563, 2001.
[58] A. Sokolov, M. Pavlova, W. Lutzenberger, and N. Birbaumer, "Reciprocal modulation of neuromagnetic induced gamma activity by attention in the human visual and auditory cortex," *NeuroImage*, vol. 22, no. 2, pp. 521–529, 2004.
[59] S. Ray, E. Niebur, S. S. Hsiao, A. Sinai, and N. E. Crone, "High-frequency gamma activity (80-150 hz) is increased in human cortex during selective attention," *Clinical Neurophysiology*, vol. 119, no. 1, pp. 116–133, 2008.
[60] N. Kahlbrock, M. Butz, E. S. May, and A. Schnitzler, "Sustained gamma band synchronization in early visual areas reflects the level of selective attention," *NeuroImage*, vol. 59, no. 1, pp. 673–681, 2012.
[61] J. A. Cardin, M. Carlén, K. Meletis, U. Knoblich, F. Zhang, K. Deisseroth, L. Tsai, and C. I. Moore, "Driving fast-spiking cells induces gamma rhythm and controls sensory responses," *Nature*, vol. 459, no. 7247, p. 663, 2009.
[62] J. Feng and K. P. Hadeler, "Qualitative behaviour of some simple networks," *Journal of Physics A: Mathematical and General*, vol. 29, no. 16, pp. 5019–5033, 1996.
[63] D. S. Bernstein, *Matrix Mathematics*, 2nd ed. Princeton University Press, 2009.
[64] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.
[65] K. P. Hadeler and D. Kuhn, "Stationary states of the Hartline-Ratliff model," *Biological Cybernetics*, vol. 56, no. 5-6, pp. 411–417, 1987.
[66] C. C. Rodgers and M. R. DeWeese, "Spiking responses of neurons in rodent prefrontal cortex and auditory cortex during a novel stimulus selection task," CRCNS.org, 2014. [Online]. Available: http://dx.doi.org/10.6080/K0W66HPJ
[67] A. W. Bronkhorst, "The cocktail-party problem revisited: early processing and selection of multi-talker speech," *Attention, Perception, & Psychophysics*, vol. 77, no. 5, pp. 1465–1487, 2015.
[68] J. Fuster, *The Prefrontal Cortex*. Elsevier Science, 2015.
[69] R. M. Bruno and D. J. Simons, "Feedforward mechanisms of excitatory and inhibitory cortical receptive fields," *Journal of Neuroscience*, vol. 22, no. 24, pp. 10 966–10 975, 2002.
[70] P. S. Goldman-Rakic, "Cellular basis of working memory," *Neuron*, vol. 14, no. 3, pp. 477–485, 1995.
[71] A. F. T. Arnsten, M. J. Wang, and C. D. Paspalas, "Neuromodulation of thought: flexibilities and vulnerabilities in prefrontal cortical network synapses," *Neuron*, vol. 76, no. 1, pp. 223–239, 2012.
[72] P. Somogyi, G. Tamasab, R. Lujan, and E. H. Buhl, "Salient features of synaptic organisation in the cerebral cortex," *Brain Research Reviews*, vol. 26, no. 2, pp. 113–135, 1998.
[73] G. K. Wu, R. Arbuckle, B. Liu, H. W. Tao, and L. I. Zhang, "Lateral sharpening of cortical frequency tuning by approximately balanced inhibition," *Neuron*, vol. 58, no. 1, pp. 132–143, 2008.
[74] H. K. Kato, S. K. Asinof, and J. S. Isaacson, "Network-level control of frequency tuning in auditory cortex," *Neuron*, vol. 95, no. 2, pp. 412–423, 2017.
[75] N. P. Bichot, M. T. Heard, E. M. DeGennaro, and R. Desimone, "A source for feature-based attention in the prefrontal cortex," *Neuron*, vol. 88, no. 4, pp. 832–844, 2015.
[76] A. Mita, H. Mushiake, K. Shima, Y. Matsuzaka, and J. Tanji, "Interval time coding by neurons in the presupplementary and supplementary motor areas," *Nature Neuroscience*, vol. 12, no. 4, p. 502, 2009.
[77] A. P. Molchanov and Y. S. Pyatnitskiy, "Criteria of asymptotic stability of differential and difference inclusions encountered in control theory," *Systems & Control Letters*, vol. 13, no. 1, pp. 59–64, 1989.
[78] W. P. Dayawansa and C. F. Martin, "A converse Lyapunov theorem for a class of dynamical systems which undergo switching," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 751–760, 1999.
[79] F. M. Hante and M. Sigalotti, "Converse Lyapunov theorems for switched systems in Banach and Hilbert spaces," *SIAM Journal on Control and Optimization*, vol. 49, no. 2, pp. 752–770, 2011.
[80] F. Wirth, "A converse Lyapunov theorem for linear parameter-varying and linear switching systems," *SIAM Journal on Control and Optimization*, vol. 44, no. 1, pp. 210–239, 2005.
[81] S. G. Krantz and H. R. Parks, *The Implicit Function Theorem: History, Theory, and Applications*. Birkhäuser, 2002.
[82] J. Kurzweil, "On the inversion of Lyapunov's second theorem on stability of motion," *American Mathematical Society Translations*, vol. 24, no. 2, pp. 19–77, 1963.
[83] P. Hartman, *Ordinary Differential Equations*, 2nd ed., ser. Classics in Applied Mathematics. SIAM, 1982.

APPENDIX A. A CONVERSE LYAPUNOV THEOREM FOR GES SWITCHED-AFFINE SYSTEMS

The existence of a converse Lyapunov function for asymptotically/exponentially stable switched linear systems has been extensively studied for time-dependent switching. Early works [77], [78] showed that if a switched linear system is asymptotically (or, equivalently, exponentially) stable under *arbitrary switching*, then it admits a common Lyapunov function. This was later extended to infinite-dimensional spaces in [79]. The limitations of these works, however, is the strong requirement of stability under arbitrary switching. [80] proved the existence of a Lyapunov function under the weaker condition of exponential stability with minimum dwell-time. Nevertheless, similar results are still missing for state-dependent switching. In this appendix, we prove a converse Lyapunov theorem for continuous GES switched affine systems with state-dependent switching that is used in both Parts I and II of this work via [40, Lemma A.2]. The considered dynamics are general and subsume the linear-threshold dynamics.

Theorem A.1. ***(Converse Lyapunov theorem for GES switched affine systems).*** *Consider the state-dependent switched affine system*

$$\begin{aligned} \tau\dot{\mathbf{x}} &= f(\mathbf{x}), \qquad && \mathbf{x}(0) = \mathbf{x}_0, \\ f(\mathbf{x}) &= \mathbf{A}_\lambda \mathbf{x} + \mathbf{b}_\lambda, \quad && \forall \mathbf{x} \in \Omega_\lambda = \{\mathbf{x} \in D \mid \mathbf{N}_\lambda \mathbf{x} + \mathbf{p}_\lambda \leq \mathbf{0}\}, \\ & && \forall \lambda \in \Lambda, \end{aligned} \tag{24}$$

where Λ is a finite index set, $\mathbf{A}_\lambda$ is nonsingular for all $\lambda \in \Lambda$, $D = \bigcup_{\lambda\in\Lambda} \Omega_\lambda \subseteq \mathbb{R}^n$ is an (open) domain, and $\{\Omega_\lambda\}_{\lambda\in\Lambda}$ have mutually disjoint interiors. Assume that f is continuous. If (24) *is GES towards a unique equilibrium $\mathbf{x}^*$, then there exists a C^∞-function $V : \mathbb{R}^n_{\geq 0} \to \mathbb{R}$ and positive constants c_1, c_2, c_3, c_4 such that for all $\mathbf{x} \in D$,*

$$c_1\|\mathbf{x} - \mathbf{x}^*\|^2 \leq V(\mathbf{x}) \leq c_2\|\mathbf{x} - \mathbf{x}^*\|^2, \tag{25a}$$

$$\frac{\partial V}{\partial \mathbf{x}} f \leq -c_3\|\mathbf{x} - \mathbf{x}^*\|^2, \tag{25b}$$

$$\left\| \frac{\partial V}{\partial \mathbf{x}} \right\| \leq c_4\|\mathbf{x} - \mathbf{x}^*\|. \tag{25c}$$

Proof: We structure the proof in three steps: (i) showing that the solutions of (24) are continuously differentiable with respect to $\mathbf{x}_0$ *along its trajectories*, (ii) construction of a (not necessarily smooth) Lyapunov-like function that satisfies (25) along the trajectories of (24), and (iii) construction of V from this Lyapunov-like function (smoothening). We only prove the result for $\mathbf{x}^* = \mathbf{0}$ as the general case can be reduced to it with the change of variables $\mathbf{x} \leftarrow \mathbf{x} - \mathbf{x}^*$.

(i) Let $\psi(t; \mathbf{x}_0)$ denote the unique solution of (24) at time $t \in \mathbb{R}$ (note that we let $t < 0$). In this step, we prove that ψ

is continuously differentiable with respect to $\mathbf{x}_0$ on D *if* $\mathbf{x}_0$ moves along ψ. Precisely, that

$$\frac{\partial}{\partial \tau}\psi(t;\psi(\tau;\mathbf{x}_0)) \text{ exists and is continuous at } \tau = 0, \quad (26)$$

for all $\mathbf{x}_0 \in D$. First, assume that $\mathbf{x}_0 \notin H$, where $H \subset D$ is the union of all the switching hyperplanes.[18] Thus, $\mathbf{x}_0$ belongs to the interior of a switching region, say Ω_{λ_1}. Let $\{\lambda_j\}_{j=1}^J$, with $J = J(t) \geq 1$, be the indices of the regions visited by $\psi(\tau;\mathbf{x}_0)$ during $\tau \in [0,t]$. With a slight abuse of notation, let $\mathbf{A}_j \triangleq \mathbf{A}_{\lambda_j}$ and $\mathbf{b}_j \triangleq \mathbf{b}_{\lambda_j}$, for $j = 1,\ldots,J$. Then,

$$\psi(\tau;\mathbf{x}_0) = \quad (27)$$
$$\begin{cases} e^{\mathbf{A}_1\tau}(\mathbf{x}_0 + \mathbf{A}_1^{-1}\mathbf{b}_1) - \mathbf{A}_1^{-1}\mathbf{b}_1; & \tau \in [0,t_1], \\ e^{\mathbf{A}_2(\tau-t_1)}(\psi(t_1;\mathbf{x}_0) + \mathbf{A}_2^{-1}\mathbf{b}_2) - \mathbf{A}_2^{-1}\mathbf{b}_2; & \tau \in [t_1,t_2], \\ \quad\vdots & \\ e^{\mathbf{A}_J(\tau-t_{J-1})}(\psi(t_{J-1};\mathbf{x}_0)+\mathbf{A}_J^{-1}\mathbf{b}_J)-\mathbf{A}_J^{-1}\mathbf{b}_J; & \tau \in [t_{J-1},t], \end{cases}$$

where $t_j = t_j(\mathbf{x}_0)$ is the time at which $\psi(\tau;\mathbf{x}_0)$ crosses the boundary between Ω_{λ_j} and $\Omega_{\lambda_{j+1}}$. This expression for ψ is valid for all $\mathbf{x}$ near $\mathbf{x}_0$ that undergo the same sequence of switches. To be precise, let $S \subset D$ be the set of points lying at the intersection of two or more switching hyperplanes and

$$S_{(-\infty,0]} = \{\mathbf{x} \in D \mid \exists t \in [0,\infty) \quad \text{s.t.} \quad \psi(t;\mathbf{x}) \in S\}.$$

In words, $S_{(-\infty,0]}$ is the set of all points that, when evolving according to (24), will pass through S at some point in time. Since S is composed of a finite number of affine manifolds of dimensions $n-2$ or smaller, $S_{(-\infty,0]}$ is in turn the union of a finite number of manifolds of dimensions $n-1$ or smaller, and thus has Lebesgue measure zero.

If $\mathbf{x}_0 \notin S_{(-\infty,0]}$, then it follows from the continuity of ψ with respect to $\mathbf{x}_0$ on D, see e.g., [45, Thm 3.5], that (27) is valid over a sufficiently small neighborhood of $\mathbf{x}_0$. Clearly, $\frac{\partial \psi}{\partial \mathbf{x}_0}$ then exists and is continuous if and only if t_j's are continuously differentiable with respect to $\mathbf{x}_0$. Consider t_1 and let $\mathbf{n}^T\mathbf{x}+p=0$ be the corresponding switching surface, where $\mathbf{n}^T$ is equal to some row of $\mathbf{N}_{\lambda_1}$ and equal to minus some row of $\mathbf{N}_{\lambda_2}$. t_1 is the (smallest) solution to

$$\mathbf{n}^T\big(e^{\mathbf{A}_1\tau}(\mathbf{x}_0 + \mathbf{A}_1^{-1}\mathbf{b}_1) - \mathbf{A}_1^{-1}\mathbf{b}_1\big) + p = 0, \quad \tau \geq 0. \quad (28)$$

The derivative of the lefthand side of (28) with respect to τ equals $\mathbf{n}^T f(\psi(t_1;\mathbf{x}_0))$, which is nonzero if and only if the curve of ψ is not tangent to $\mathbf{n}^T\mathbf{x}+p=0$. If so, then the continuous differentiability of t_1 with respect to $\mathbf{x}_0$ follows from the implicit function theorem [81]. Otherwise, it is not difficult to show that $\psi(t;\mathbf{x}_0)$ remains in Ω_{λ_1} after t_1[19], contradicting the fact that t_1 is a switching time. The same argument guarantees that $t_j, j = 2,\ldots,J$ are also continuously differentiable with respect to $\mathbf{x}_0$, and so is $\psi(t;\mathbf{x}_0)$.

Before moving on to the case when $\mathbf{x}_0 \in S_{(-\infty,0]}$, we analyze the case where still $\mathbf{x}_0 \notin S_{(-\infty,0]}$ but $\mathbf{x}_0 \in H$, i.e., $\mathbf{x}_0$ belongs to a switching hyperplane, say $\mathbf{n}^T\mathbf{x}+p=0$ between Ω_{λ_1} from Ω_{λ_2}, as above. For simplicity, assume t is small enough such that $\psi(\tau;\mathbf{x}_0)$ remains within Ω_{λ_2} for all $\tau \in [0,t]$.[20] Let $\mathbf{x}$ belong to a sufficiently small neighborhood of $\mathbf{x}_0$ such that for $\tau \in [0,t]$,

$$\psi(\tau;\mathbf{x}) = \quad (29)$$
$$\begin{cases} e^{\mathbf{A}_2\tau}(\mathbf{x} + \mathbf{A}_2^{-1}\mathbf{b}_2) - \mathbf{A}_2^{-1}\mathbf{b}_2; & \mathbf{x} \in \Omega_{\lambda_2}, \\ e^{\mathbf{A}_1\tau}(\mathbf{x} + \mathbf{A}_1^{-1}\mathbf{b}_1) - \mathbf{A}_1^{-1}\mathbf{b}_1; & \mathbf{x} \in \Omega_{\lambda_1}, \tau \leq t_1, \\ e^{\mathbf{A}_2(\tau-t_1)}(\psi(t_1;\mathbf{x}) + \mathbf{A}_2^{-1}\mathbf{b}_2) - \mathbf{A}_2^{-1}\mathbf{b}_2; & \mathbf{x} \in \Omega_{\lambda_1}, \tau \geq t_1, \end{cases}$$

where $t_1 = t_1(\mathbf{x})$ is now the solution to $\mathbf{n}^T\psi(t_1;\mathbf{x}) + p = 0$. It is not difficult to show that for $\mathbf{x} \in \Omega_{\lambda_1}$,

$$\begin{aligned} \frac{\partial \psi(t;\mathbf{x})}{\partial x_i} = e^{\mathbf{A}_2 t}\Big[e^{-\mathbf{A}_2 t_1}e^{\mathbf{A}_1 t_1}e_i + \frac{\partial t_1}{\partial x_i} \\ \times \Big(-\mathbf{A}_2 e^{-\mathbf{A}_2 t_1}e^{\mathbf{A}_1 t_1}(\mathbf{x} + \mathbf{A}_1^{-1}\mathbf{b}_1) + e^{-\mathbf{A}_2 t_1}\mathbf{A}_1 e^{\mathbf{A}_1 t_1} \\ \times (\mathbf{x} + \mathbf{A}_1^{-1}\mathbf{b}_1) + \mathbf{A}_2 e^{-\mathbf{A}_2 t_1}(\mathbf{A}_2^{-1}\mathbf{b}_2 - \mathbf{A}_1^{-1}\mathbf{b}_1)\Big)\Big], \end{aligned}$$

where e_i is the i'th column of $\mathbf{I}_n$. Taking the limit of this expression as $\mathbf{x} \to \mathbf{x}_0$ and using the facts that $\lim_{\mathbf{x}\to\mathbf{x}_0} t_1 = 0$ and $\mathbf{A}_1\mathbf{x}_0 + \mathbf{b}_1 = \mathbf{A}_2\mathbf{x}_0 + \mathbf{b}_2$, we get

$$\begin{aligned} &\lim_{\substack{\Omega_{\lambda_1} \\ \mathbf{x} \to \mathbf{x}_0}} \frac{\partial \psi(t;\mathbf{x})}{\partial x_i} = e^{\mathbf{A}_2 t}e_i, \qquad \forall i \in \{1,\ldots,n\}, \\ \Rightarrow &\lim_{\substack{\Omega_{\lambda_1} \\ \mathbf{x} \to \mathbf{x}_0}} \frac{\partial \psi(t;\mathbf{x})}{\partial \mathbf{x}} = e^{\mathbf{A}_2 t} = \lim_{\substack{\Omega_{\lambda_2} \\ \mathbf{x} \to \mathbf{x}_0}} \frac{\partial \psi(t;\mathbf{x})}{\partial \mathbf{x}}, \end{aligned}$$

where the second equality follows directly from (29). Therefore, $\psi(t;\mathbf{x}_0)$ is continuously differentiable with respect to $\mathbf{x}_0$ on the entire $D \setminus S_{(-\infty,0]}$.

Finally, if $\mathbf{x}_0 \in S_{(-\infty,0]}$, the same expression as (27) or (29) (depending on whether $\mathbf{x}_0 \in H$ or not) holds for $\mathbf{x}_0$ and also for all $\mathbf{x}$ within a sufficiently small neighborhood of it *that lie on the same system trajectory as* $\mathbf{x}_0$. This curve can be parameterized in many ways, one of which is given by $\psi(\tau;\mathbf{x}_0)$. Together with the analysis of the case $\mathbf{x}_0 \notin S_{(-\infty,0]}$ above, this proves that (26) exists and is continuous at τ_0[21].

(ii) In this step we introduce a function $\hat{V}$ that may not be smooth but satisfies properties similar to (25). Let

$$\hat{V}(\mathbf{x}) \triangleq \int_0^\delta \|\psi(t;\mathbf{x})\|^2 dt, \qquad \forall \mathbf{x} \in D,$$

where δ is a constant to be chosen. It is straightforward to show that f is globally Lipschitz. Using this and the GES of (24), the same argument as in [45, Thm 4.14] shows that

$$2c_1\|\mathbf{x}\|^2 \leq \hat{V}(\mathbf{x}) \leq \frac{2}{3}c_2\|\mathbf{x}\|^2, \quad (30)$$

for some $c_1, c_2 > 0$. Further, let

$$D_{\psi\circ\psi}(t;\tau;\mathbf{x}) \triangleq \frac{\partial}{\partial \tau}\psi(t;\psi(\tau;\mathbf{x})), \qquad t,\tau \in \mathbb{R}, \mathbf{x} \in D.$$

By the definition of ψ, we have the identity $\psi(t;\psi(s-t;\mathbf{x})) = \psi(s,\mathbf{x})$, $t,s \in \mathbb{R}, \mathbf{x} \in D$. Taking $\frac{d}{dt}$ of both sides, we get $\psi_t(t;\psi(s-t;\mathbf{x})) - D_{\psi\circ\psi}(t;s-t;\mathbf{x}) = 0$, where $\psi_t(t;\mathbf{x}) =$

[18] Recall that for each λ, each row of $\mathbf{N}_\lambda\mathbf{x} + \mathbf{p}_\lambda = \mathbf{0}$ defines a switching hyperplane.

[19] This is a general fact about the solutions of linear systems and can be shown using the series expansion of the matrix exponential.

[20] Note that if t is larger, then subsequent switches to $\Omega_{\lambda_j}, j \geq 3$ are similar to the case above (where $\mathbf{x}_0$ was not on a switching hyperplane) and thus do not violate continuous differentiability of ψ with respect to $\mathbf{x}_0$.

[21] We have indeed proved a slightly stronger result than (26) for $\mathbf{x}_0 \notin S_{(-\infty,0]}$, which we use in step (ii) below.

$\frac{\partial \psi(t;\mathbf{x})}{\partial t}$. Setting $s = t+\tau$, $D_{\psi\circ\psi}(t;\tau;\mathbf{x}) = \psi_t(t;\psi(\tau;\mathbf{x}))$. For the parallel of (25b), we then have

$$\begin{aligned}\frac{d}{d\tau}\hat{V}(\psi(\tau;\mathbf{x})) &= \int_0^\delta 2\psi(t;\psi(\tau;\mathbf{x}))^T D_{\psi\circ\psi}(t;\tau;\mathbf{x})dt \\ &= \int_0^\delta 2\psi(t;\psi(\tau;\mathbf{x}))^T \psi_t(t;\psi(\tau;\mathbf{x}))dt \\ &= \int_0^\delta \frac{\partial}{\partial t}\|\psi(t;\psi(\tau;\mathbf{x}))\|^2 dt \\ &= \|\psi(\delta;\psi(\tau;\mathbf{x}))\|^2 - \|\psi(\tau;\mathbf{x})\|^2.\end{aligned}$$

Thus

$$\frac{d}{d\tau}\hat{V}(\psi(\tau;\mathbf{x}))\Big|_{\tau=0} = \|\psi(\delta;\mathbf{x})\|^2 - \|\mathbf{x}\|^2 \le -2c_3\|\mathbf{x}\|^2, \quad (31)$$

where the last inequality holds, as shown in [45, Thm 4.14], for an appropriate choice of δ and $c_3 = \frac{1}{4}$. Finally, for the parallel of (25c), recall from step (i) that $\frac{\partial}{\partial \mathbf{x}}\psi(t;\mathbf{x})$ exists and is continuous on $D \setminus S_{(-\infty,0]}$. Therefore, from (24), we have

$$\frac{\partial}{\partial t}\frac{\partial \psi(t;\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial f}{\partial \mathbf{x}}(\psi(t;\mathbf{x}))\frac{\partial \psi(t;\mathbf{x})}{\partial \mathbf{x}}, \quad \frac{\partial \psi(t;\mathbf{x})}{\partial \mathbf{x}}\Big|_{t=0} = \mathbf{I}_n,$$

on $D \setminus (S_{(-\infty,0]} \cup H)$. Using the global Lipschitzness of f and the fact that $D \setminus S_{(-\infty,0]}$ is invariant under (24), we have $\left\|\frac{\partial \psi(t;\mathbf{x})}{\partial \mathbf{x}}\right\| \le e^{Lt}$, for all $x \in D \setminus S_{(-\infty,0]}$, where L is the Lipschitz constant of f. The same argument as in [45, Thm 4.14] then yields

$$\left\|\frac{\partial \hat{V}}{\partial \mathbf{x}}\right\| \le \frac{2}{3}c_4\|\mathbf{x}\|, \qquad \forall x \in D \setminus S_{(-\infty,0]}, \quad (32)$$

for some $c_4 > 0$.

(iii) In this step, we follow [82, Thm 3 & 4] to construct V as an smooth approximation to $\hat{V}$ and show that it satisfies (25). Since f is globally Lipschitz, $\psi(t;\mathbf{x})$ is Lipschitz in $\mathbf{x}$ (see, e.g., [83, Ch 5]) and so is $\hat{V}$. This, together with (31), satisfies all the assumptions of [82, Thm 4], which in turn guarantees the existence of an infinitely smooth V such that

$$|V(\mathbf{x}) - \hat{V}(\mathbf{x})| < \frac{1}{2}\hat{V}(\mathbf{x}), \qquad \forall \mathbf{x} \in D, \quad (33a)$$

$$\frac{\partial V}{\partial \mathbf{x}} f(\mathbf{x}) < -c_3\|\mathbf{x}\|^2, \quad (33b)$$

for all $\mathbf{x} \in D$. Equation (25a) follows immediately from (33b) and (30). To prove (25c), we note that the same construction of V as in [82, Thm 3 & 4] satisfies

$$\left\|\frac{\partial V}{\partial \mathbf{x}} - \frac{\partial \hat{V}}{\partial \mathbf{x}}\right\| < \frac{1}{2}\left\|\frac{\partial \hat{V}}{\partial \mathbf{x}}\right\|, \qquad \forall \mathbf{x} \in D \setminus S_{(-\infty,0]},$$

if the constants $\xi_{i,k}$ and $\zeta_{i,k}, i,k = \ldots,-2,0,2,\ldots$ (and consequently the corresponding $\bar{r}_{i,k}, i,k = \ldots,-2,0,2,\ldots$) are chosen sufficiently small. This, together with (32), guarantees (25c), completing the proof. ■

Appendix B. Additional Proofs

Proof of Lemma IV.1: Pick $\mathbf{c}' \in \mathbb{R}^{n'}$ and let $\mathbf{x}^*$ be the unique solution of (15). Since $\bigcup_{\lambda\in\Lambda}\Psi_\lambda = \mathbb{R}^n$, let $\lambda \in \Lambda$ with

$$\mathbf{W}_3\mathbf{x}^* + \bar{\mathbf{c}} \in \Psi_\lambda. \quad (34)$$

If $\mathbf{W}_3\mathbf{x}^* + \bar{\mathbf{c}}$ lies on the boundary of more than one Ψ_λ, pick one arbitrarily. Therefore, $\mathbf{x}^*$ satisfies

$$\mathbf{x}^* = [(\mathbf{W}_1 + \mathbf{W}_2\mathbf{F}_\lambda\mathbf{W}_3)\mathbf{x}^* + \mathbf{W}_2(\mathbf{F}_\lambda\bar{\mathbf{c}} + \mathbf{f}_\lambda) + \mathbf{c}']_{\mathbf{0}}^{\mathbf{m}}.$$

From (8), it follows that h' has the form (16) with $\lambda' \triangleq (\lambda, \boldsymbol{\sigma})$ and $\Lambda' = \Lambda \times \{0,\ell,\mathrm{s}\}^{n'}$. The quantities $\mathbf{F}'_{\lambda'}, \mathbf{f}'_{\lambda'}, \mathbf{G}'_{\lambda'}, \mathbf{g}'_{\lambda'}$ also have the same form as in (8) except that here

$$\begin{aligned}&\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2\mathbf{F}_\lambda\mathbf{W}_3, \\ &\mathbf{f}'_{\lambda'} = (\mathbf{I} - \boldsymbol{\Sigma}^\ell\mathbf{W})^{-1}\boldsymbol{\Sigma}^{\mathrm{s}}\mathbf{m} + (\mathbf{I} - \boldsymbol{\Sigma}^\ell\mathbf{W})^{-1}\boldsymbol{\Sigma}^\ell\mathbf{W}_2(\mathbf{F}_\lambda\bar{\mathbf{c}} + \mathbf{f}_\lambda).\end{aligned}$$

The proof is complete noting that $\bigcup_{\lambda'\in\Lambda'}\Psi'_{\lambda'} = \mathbb{R}^{n'}$ since any $\mathbf{c}' \in \mathbb{R}^{n'}$ must be in at least one $\Psi'_{\lambda'}$ by construction. ■

Proof of Lemma IV.2: Pick any $\mathbf{c}, \hat{\mathbf{c}} \in \mathbb{R}^n$. Since all the sets Ψ_λ are convex, the line segment $\gamma \triangleq \{(\theta, (1-\theta)\mathbf{c} + \theta\hat{\mathbf{c}}) \mid \theta \in [0,1]\}$ joining $\mathbf{c}$ and $\hat{\mathbf{c}}$ can be broken into $k \le |\Lambda| < \infty$ pieces such that $\gamma = \bigcup_{i=1}^k \gamma_i, \gamma_i \triangleq \{(\theta, (1-\theta)\mathbf{c} + \theta\hat{\mathbf{c}}) \mid \theta \in [\theta_{i-1}, \theta_i]\}, \theta_0 = 0, \theta_k = 1$ and each $\gamma_i \subset \Psi_{\lambda_i}$ for some $\lambda_i \in \Lambda$. Let $\mathbf{c}_i \triangleq (1-\theta_i)\mathbf{c} + \theta_i\hat{\mathbf{c}}$. Then,

$$\begin{aligned}&\|h(\mathbf{c}) - h(\hat{\mathbf{c}})\| = \Big\|\sum_{i=1}^k \big(h(\mathbf{c}_{i-1}) - h(\mathbf{c}_i)\big)\Big\| \\ &\le \sum_{i=1}^k \|h(\mathbf{c}_{i-1}) - h(\mathbf{c}_i)\| = \sum_{i=1}^k \|\mathbf{F}_{\lambda_i}(\mathbf{c}_{i-1} - \mathbf{c}_i)\| \\ &\le \Big[\max_{\lambda\in\Lambda}\|\mathbf{F}_\lambda\|\Big]\sum_{i=1}^k \|\mathbf{c}_{i-1} - \mathbf{c}_i\| = \Big[\max_{\lambda\in\Lambda}\|\mathbf{F}_\lambda\|\Big]\|\mathbf{c} - \hat{\mathbf{c}}\|.\end{aligned}$$ ■

Erfan Nozari received his B.Sc. degree in Electrical Engineering-Control in 2013 from Isfahan University of Technology, Iran and Ph.D. in Mechanical Engineering and Cognitive Science in 2019 from University of California San Diego. He is currently a postdoctoral researcher at the University of Pennsylvania Department of Electrical and Systems Engineering. He has been the (co)recipient of the 2019 IEEE Transactions on Control of Network Systems Outstanding Paper Award, the Best Student Paper Award from the 57th IEEE Conference on Decision and Control, the Best Student Paper Award from the 2018 American Control Conference, and the Mechanical and Aerospace Engineering Distinguished Fellowship Award from the University of California San Diego. His research interests include dynamical systems and control theory and its applications in computational and theoretical neuroscience and complex network systems.

Jorge Cortés (M'02-SM'06-F'14) received the Licenciatura degree in mathematics from Universidad de Zaragoza, Spain, in 1997, and the Ph.D. degree in engineering mathematics from Universidad Carlos III de Madrid, Spain, in 2001. He held postdoctoral positions with the University of Twente, The Netherlands, and the University of Illinois at Urbana-Champaign, USA. He was an Assistant Professor with the Department of Applied Mathematics and Statistics, University of California, Santa Cruz, USA, from 2004 to 2007. He is currently a Professor in the Department of Mechanical and Aerospace Engineering, University of California, San Diego, USA. He is the author of Geometric, Control and Numerical Aspects of Nonholonomic Systems (Springer-Verlag, 2002) and co-author (together with F. Bullo and S. Martínez) of Distributed Control of Robotic Networks (Princeton University Press, 2009). His current research interests include distributed control and optimization, network science, resource-aware control, decision making under uncertainty, and distributed coordination in power networks, robotics, and transportation.