Hierarchical Timescales in the Neocortex: Mathematical Mechanism and Biological Insights

Songting Li[1,†], Xiao-Jing Wang[2,†]

[1] *School of Mathematical Sciences, MOE-LSC, and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China*
[2] *Center for Neural Science, New York University, New York 10003, United States*

[†] *Correspondence: songting@sjtu.edu.cn (S.L.) and xjwang@nyu.edu (X.-J. W.)*

Abstract

A cardinal feature of the neocortex is the progressive increase of the spatial receptive fields along the cortical hierarchy. Recently, theoretical and experimental findings have shown that the temporal response windows also gradually enlarge, so that early sensory neural circuits operate on short-time scales whereas higher association areas are capable of integrating information over a long period of time. While an increased receptive field is accounted for by spatial summation of inputs from neurons in an upstream area, the emergence of timescale hierarchy cannot be readily explained, especially given the dense inter-areal cortical connectivity known in modern connectome. To uncover the required neurobiological properties, we carried out a rigorous analysis of an anatomically-based large-scale cortex model of macaque monkeys. Using a perturbation method, we show that the segregation of disparate timescales is defined in terms of the localization of eigenvectors of the connectivity matrix, which depends on three circuit properties: (1) a macroscopic gradient of synaptic excitation, (2) distinct electrophysiological properties between excitatory and inhibitory neuronal populations, and (3) a detailed balance between long-range excitatory inputs and local inhibitory inputs for each area-to-area pathway. Our work thus provides a quantitative understanding of the mechanism underlying the emergence of timescale hierarchy in large-scale primate cortical networks.

Keywords

large-scale cortical network | timescale hierarchy | eigenvector localization | inter-areal heterogeneity | detailed excitation-inhibition balance of long-range cortical connections

Significance Statement

In the neocortex, while early sensory areas encode and process external inputs rapidly, higher association areas are endowed with slow dynamics suitable for accumulating information over time. Such a hierarchy of temporal response windows along the cortical hierarchy naturally emerges in a model of multi-areal primate cortex. This finding raises the question of why diverse temporal modes are not mixed in roughly the same way across the whole cortex, despite high connection density and an abundance of feedback loops. We investigate this question by mathematically analyzing the anatomically-based network model of macaque cortex, and show that three general principles of synaptic excitation and inhibition are crucial for timescale segregation in a hierarchy, a functionally important characteristic of the cortex.

THE brain is organized with a delicate structure to integrate and process both spatial and temporal information received from the external world. For spatial information processing, neurons along cortical visual pathways possess increasingly large spatial receptive fields, and its underlying mechanism has been understood as neurons in higher level visual areas receive input from many neurons with smaller receptive fields in lower level visual areas, thereby aggregating information across space (Hubel, 1995). More recently, a computational model (Chaudhuri et al., 2015) revealed that the timescale over which neural integration occurs also gradually increases from area to area along the cortical hierarchy. The model was based on the anatomically measured directed- and weighted- inter-areal connectivity of the macaque cortex (Markov et al., 2014a) and incorporated heterogeneity of synaptic excitation calibrated by spine count per pyramidal neuron (Elston, 2007). It has been observed that the decay times increased progressively along the cortical hierarchy when signals propagate in the network, and the temporal hierarchy could change dynamically in response to different types of sensory inputs (e.g., different hierarchy of timescales for somatosensory input versus visual input) (Chaudhuri et al., 2015). By manipulating parameters of the model, simulation results further demonstrate that both within and between regions of anatomical properties could affect the hierarchy of timescales in neuronal population activity (Chaudhuri et al., 2015). A hierarchy of temporal receptive windows is functionally desirable, so that the circuit dynamics operate on short time scales in early sensory areas to encode and process rapidly changing external stimuli, whereas parietal and frontal areas can accumulate information over a relatively long period of time window in decision-making and other cognitive processes (Gold and Shadlen, 2007; Wang, 2008).

Despite the accumulating evidence in support of timescale hierarchy across cortical areas in mice (Runyan et al., 2017; Siegle et al., 2021), monkeys (Ogawa and Komatsu, 2010; Murray et al., 2014; Cavanagh et al., 2016; Fascianelli et al., 2019; Spitmaan et al., 2020; Maisson et al., 2021; Manea et al., 2021), and humans (Hasson et al., 2008; Honey et al., 2012; Lerner et al., 2011; Stephens et al., 2013; Yeshurun et al., 2017; Raut et al., 2020; Shafiei et al., 2020; Gao et al., 2020), its underlying mechanism remains unclear. In particular, since inter-areal connections are dense, with roughly 65% of all possible connections present in the macaque cortex (Markov et al., 2014a) and even higher connection density in the mouse cortex (Gamanut et al., 2018), what circuit properties are required to ensure that dynamical modes with disparate time constants are spatially localized? How intra-areal anatomical properties determine the intrinsic timescale of each area, and how these intrinsic timescales remain to be segregated rather than mixed up in the presence of dense inter-areal connections? In this work, we addressed these questions by a mathematical analysis of the model (Chaudhuri et al., 2015). Using a perturbation method, we identified key required conditions, in particular a detailed excitation-inhibition balance for long-distance inter-areal connections which is experimentally testable.

The multi-areal model and hierarchical timescales phenomenon

We first review the mathematical form of the multi-areal model of the macaque cortex, and the hierarchical timescales phenomenon captured by this model (Chaudhuri et al., 2015). The macaque cortical network model contains a subnet of 29 areas widely distributed from sensory to association areas in the macaque cortex, and each area includes both excitatory and inhibitory neuronal populations. The neuronal population dynamics in the ith area is described as

$$\tau_E \frac{d}{dt} v_E^i = -v_E^i + \beta_E \left[I_{syn,E}^i + I_{ext,E}^i \right]_+ \tag{1}$$

$$\tau_I \frac{d}{dt} v_I^i = -v_I^i + \beta_I \left[I_{syn,I}^i + I_{ext,I}^i \right]_+, \tag{2}$$

where v_E^i and v_I^i are the firing rate of the excitatory and inhibitory populations in the ith area, respectively, τ_E and τ_I are their time constants, respectively, and β_E and β_I are the slope of the f-I curve for the excitatory and inhibitory populations, respectively. The f-I curve takes the form of a rectified linear function with $[I]_+ = max(I, 0)$. In addition, $I_{ext,E}^i$ and $I_{ext,I}^i$ are the external currents, and $I_{syn,E}^i$ and $I_{syn,I}^i$ are the synaptic currents that follow

$$I^i_{syn,E} = (1+\eta h_i)\Big(w_{EE}v^i_E + \mu_{EE}\sum_{j=1}^{N} FLN_{ij}v^j_E\Big) - w_{EI}v^i_I,$$

$$I^i_{syn,I} = (1+\eta h_i)\Big(w_{IE}v^i_E + \mu_{IE}\sum_{j=1}^{N} FLN_{ij}v^j_E\Big) - w_{II}v^i_I,$$

where w_{pq}, $p,q \in \{E,I\}$ is the local coupling strength from the q population to the p population within each area. FLN_{ij} is the fraction of labeled neurons (FLN) from area j to area i reflecting the strengths of long-range input (Markov et al., 2014a), and μ_{EE} and μ_{IE} are scaling parameters that control the strengths of long-range input to the excitatory and inhibitory populations, respectively. Both local and long-range excitatory inputs to an area are scaled by its position in the hierarchy quantified by h_i (a value normalized between 0 and 1), based on the observation that the hierarchical position of an area highly correlates with the number of spines on pyramidal neurons in that area (Elston, 2007; Chaudhuri et al., 2015). A constant η maps the hierarchy h_i into excitatory connection strengths. Note that both local and long-range projections are scaled by hierarchy, rather than just local projections, following the observation that the proportion of local to long-range connections is approximately conserved across areas (Markov et al., 2011). The values of all the model parameters are specified in Sec. *Materials and Methods*.

By simulating the model, it has been observed in Ref. (Chaudhuri et al., 2015) that the decay time of neuronal response in each area increases progressively along the visual cortical hierarchy when an pulse input is given to area V1, as shown here in Fig. 1A. Early visual areas show fast and transient responses while prefrontal areas show slower responses and longer integration times with traces lasting for several seconds after the stimulation. In addition, white-noise input to V1 is also integrated with a hierarchy of timescales by computing the autocorrelation of neuronal activity at each area (Chaudhuri et al., 2015). As shown in Fig. 1B, the activity of early sensory areas shows rapid decay of autocorrelation with time lag while that of association areas shows slow decay. In Fig. 1C, by fitting single or double exponentials to the decay of the autocorrelation curves (Chaudhuri et al., 2015), the dominant timescale of each area tends to increase along the hierarchy approximately, thus a hierarchy of widely disparate timescales emerges from this model. It is worth noting, however, that the timescale does not change monotonically with the anatomically defined hierarchy (x-axis), the precise pattern is sculpted by the measured inter-areal wiring properties.

Note that, although the multi-areal model (Eqs. 1-2) is nonlinear by taking into account a rectified linear f-I curve, the stimuli in our simulations drive all neuronal population activities above the firing threshold with positive input currents to all areas. Therefore, the stimuli essentially drive the network dynamics into the linear regime. Before we perform mathematical analysis to understand the mechanism underlying the emergence of hierarchical timescales in the simulations, to simplify the notation, we rewrite the network dynamics Eqs. 1-2 in the linear regime in the following form

$$\frac{d}{dt}\boldsymbol{v} = W\boldsymbol{v} + \boldsymbol{I}_{ext}, \tag{3}$$

where

$$\boldsymbol{v} = \Big[v^1_E,\ldots,v^n_E,v^1_I,\ldots,v^n_I\Big]^T, \quad W = \begin{bmatrix} D_{EE}+F_{EE} & D_{EI} \\ D_{IE}+F_{IE} & D_{II}\end{bmatrix},$$

with $n = 29$, and D_{EE}, D_{EI}, D_{IE}, D_{II} being four diagonal matrices whose ith element on their diagonal line are

$$d^i_{EE} = \frac{\beta_E}{\tau_E}\big[(1+\eta h_i)w_{EE} - \frac{1}{\beta_E}\big], \quad d^i_{EI} = -\frac{\beta_E}{\tau_E}\big[w_{EI}\big],$$

$$d^i_{IE} = \frac{\beta_I}{\tau_I}\big[(1+\eta h_i)w_{IE}\big], \quad d^i_{II} = -\frac{\beta_I}{\tau_I}\big[w_{II} + \frac{1}{\beta_I}\big],$$

respectively, and matrices F_{EE} and F_{IE} being two non-diagonal matrices whose ith-row-jth-column element is

$$f^{ij}_{EE} = \frac{\beta_E}{\tau_E}\big[(1+\eta h_i)\mu_{EE}FLN_{ij}\big], \quad f^{ij}_{IE} = \frac{\beta_I}{\tau_I}\big[(1+\eta h_i)\mu_{IE}FLN_{ij}\big],$$

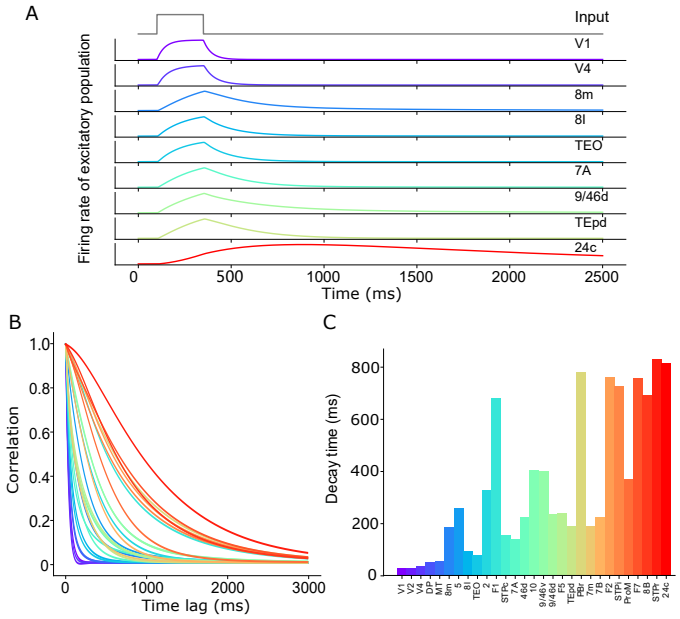

Figure 1: The hierarchical timescales phenomenon simulated in the macaque multi-areal model. (A), a pulse of input to area V1 is propagated along the hierarchy, displaying increasing decay times as it proceeds. (B), autocorrelation of area activity in response to white-noise input to V1. (C), the dominant time constants in all areas, extracted by fitting single or double exponentials to the autocorrelation curves (Chaudhuri et al., 2015). In (A)-(C), areas are arranged and colored by position in the anatomical hierarchy.

respectively. Note that matrices D_{EE}, D_{EI}, D_{IE}, D_{II} reflect local intra-areal interactions, while matrices F_{EE} and F_{IE} reflect long-range inter-areal interactions. In addition, elements in D_{EE}, D_{IE}, F_{EE}, and F_{IE} depend on area hierarchy h_i while elements in D_{EI} and D_{II} are constant. Finally, the external input vector is

$$\boldsymbol{I}_{ext} = \left[\frac{\beta_E}{\tau_E} I^1_{ext,E}, \ldots, \frac{\beta_E}{\tau_E} I^n_{ext,E}, \frac{\beta_I}{\tau_I} I^1_{ext,I}, \ldots, \frac{\beta_I}{\tau_I} I^n_{ext,I}\right]^T.$$

Denoting the eigenvalues and eigenvectors of the connectivity matrix W as λ_i and $\boldsymbol{V}_i$ $(i = 1, 2, ..., 2n)$, respectively, i.e., $W\boldsymbol{V}_i = \lambda_i \boldsymbol{V}_i$, the analytical solution of Eq. 3 can be obtained as

$$\boldsymbol{v}^i(t) = \sum_{j=1}^{2n} \left(\tilde{a}_j e^{\lambda_j t} + \int_0^t e^{\lambda_j (t-t')} \tilde{I}_j(t') dt'\right) \boldsymbol{V}^i_j, \tag{4}$$

where $\boldsymbol{v}^i$ and $\boldsymbol{V}^i_j$ are the ith element in $\boldsymbol{v}$ and $\boldsymbol{V}_j$, respectively, $\tilde{a}_j$ and $\tilde{I}_j$ are the coefficients for the initial condition and the external input, respectively, represented in the coordinate system of the eigenvectors $\{\boldsymbol{V}_j\}$. Note that, from Eq. 4, each area integrates input current with the same set of time constants $\{\tau_i\}$ determined by the real part of the eigenvalues, i.e., $\tau_i = -1/Re\{\lambda_i\}$. Therefore, the characteristic timescale of each area across the network is expected to be similar in general case. To obtain distinct timescales at each area, it requires (1) the localization of eigenvectors $\boldsymbol{V}_j$, i.e., most of the elements in $\boldsymbol{V}_j$ are close to zero, (2) the orthogonality of all pairs of eigenvectors, i.e., the non-zero elements are nearly non-overlap for different $\boldsymbol{V}_j$.

By computing the eigenvalues and eigenvectors of matrix W, as shown in Fig. 2A, the timescale pool $\{\tau_i\}$ derived from eigenvalues can be classified into two groups, one group shows quite fast timescale about 2 milliseconds, and the other group includes relatively slow timescales ranging from tens to hundreds of milliseconds. In addition, we are particularly interested in the excitatory population because the majority of neurons in the cortex are excitatory neurons. We observe that the magnitude of the eigenvectors corresponding to the fast timescale is nearly zero for the excitatory population at each area, while that corresponding to the slow timescale is weakly localized and weakly orthogonal, i.e., each eigenvector

has a few nonzero elements that almost do not overlap with other eigenvectors' nonzero elements. According to Eq. 4, the pattern of eigenvectors gives rise to the disparate timescales for the excitatory neuronal population at each cortical area. We next perform mathematical analysis to investigate the sufficient conditions for (1) vanishing magnitude of the excitatory component of fast-eigenmode eigenvectors, and (2) weak localization and orthogonality of the excitatory component of slow-eigenmode eigenvectors in this network system.

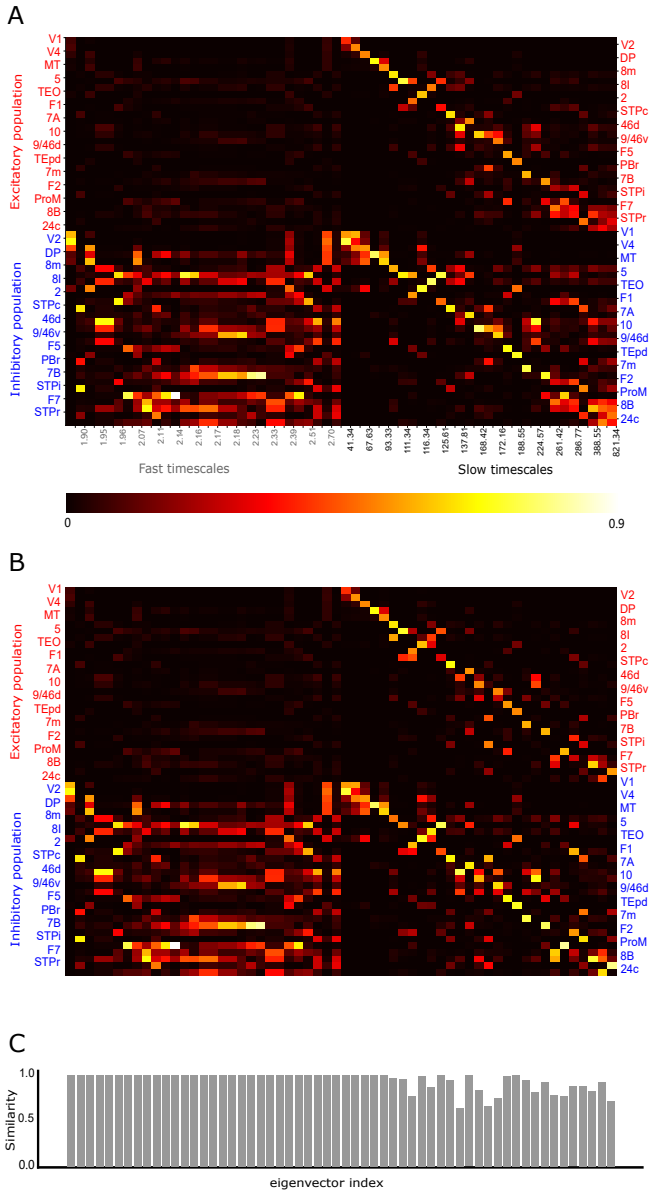

Figure 2: Eigenvectors of the network connectivity matrix and their approximations from the perturbation analysis. (A), eigenvectors of the network connectivity matrix W. Each column shows the amplitude of an eigenvector at the 29 areas, with corresponding timescale labeled below. (B), eigenvectors of W calculated from the first-order perturbation analysis. (C), similarity measure defined as the inner product of the corresponding eigenvectors in (A) and (B).

Perturbation analysis of the model

We note that the parameters of the model (specified in Sec. *Materials and Methods*) gives

$$\varepsilon = \frac{\beta_E}{\tau_E}/\frac{\beta_I}{\tau_I} \approx 0.094, \quad \delta = \frac{\mu_{EE}}{\mu_{IE}} - \frac{w_{EI}}{w_{II}+1/\beta_I} \approx 0.038,$$

which can be viewed as two small parameters in order to allow us to perform perturbation analysis below.

We first study the network in the absence of the long-range interactions among areas. In this scenario, we study the 2 by 2 block matrix $D = \begin{bmatrix} D_{EE} & D_{EI} \\ D_{IE} & D_{II} \end{bmatrix}$ in which each block is a diagonal matrix defined above. By viewing $\varepsilon = \frac{\beta_E}{\tau_E}/\frac{\beta_I}{\tau_I}$ as a small parameter, we have

$$D_{II}, D_{IE} \sim \mathscr{O}(1);\ D_{EI}, D_{EE} \sim \mathscr{O}(\varepsilon) \tag{5}$$

from their definitions. Accordingly, we can prove the following proposition:

Proposition 1 *If $D_{II} \sim \mathscr{O}(1)$, $D_{IE} \sim \mathscr{O}(1)$, $D_{EI} \sim \mathscr{O}(\varepsilon)$, $D_{EE} \sim \mathscr{O}(\varepsilon)$, then D has n eigenvalues being $\mathscr{O}(\varepsilon)$, and n eigenvalues being $\mathscr{O}(1)$.*

Proof: It is straightforward to prove that Matrix D can be diagonalized by matrix P, i.e.,

$$\Lambda = P^{-1}DP = \begin{bmatrix} \Lambda_U & O \\ O & \Lambda_L \end{bmatrix} = \begin{bmatrix} D_{EE}+AD_{IE} & O \\ O & -D_{IE}A+D_{II} \end{bmatrix},$$

where

$$P = \begin{bmatrix} I+AB & -A \\ -B & I \end{bmatrix},$$

I is the identity matrix, and diagonal matrix A satisfies $-(D_{EE}+AD_{IE})A + D_{EI} + AD_{II} = 0$, diagonal matrix B satisfies $D_{IE} + B(D_{EE}+AD_{IE}) + (D_{IE}A - D_{II})B = 0$.

We solve the equation of A and choose one of the two solutions of A as

$$A = \frac{1}{2}D_{IE}^{-1}\left[D_{II} - D_{EE} + \sqrt{(D_{II}-D_{EE})^2 + 4D_{IE}D_{EI}}\right],$$

where the square root of a diagonal matrix is defined as taking the square root of its elements. Due to the fact that $D_{II} \sim \mathscr{O}(1)$, $D_{IE} \sim \mathscr{O}(1)$, $D_{EI} \sim \mathscr{O}(\varepsilon)$, $D_{EE} \sim \mathscr{O}(\varepsilon)$, we have $A = -D_{EI}D_{II}^{-1} + \mathscr{O}(\varepsilon^2) \sim \mathscr{O}(\varepsilon)$, and $B = D_{IE}D_{II}^{-1} + \mathscr{O}(\varepsilon) \sim \mathscr{O}(1)$. Accordingly, we have

$$\begin{aligned} \Lambda_U &= D_{EE} - D_{EI}D_{II}^{-1}D_{IE} + \mathscr{O}(\varepsilon^2) \sim \mathscr{O}(\varepsilon), \\ \Lambda_L &= D_{II} + D_{EI}D_{II}^{-1}D_{IE} + \mathscr{O}(\varepsilon^2) \sim \mathscr{O}(1). \end{aligned}$$

■

From Proposition 1, the eigenvalues of D have two separated scales belonging to Λ_U and Λ_L, respectively. As the timescales of the network system is given by $\tau_i = -1/Re\{\lambda_i\}$ (λ_i is the ith diagonal element in matrix Λ), the separation of scales for eigenvalues in Λ_U and Λ_L explains that the intrinsic timescale pool can be classified into two group with a separation of scales, which is mainly determined by the distinct electrophysiological properties between excitatory and inhibitory neuronal populations within each area described by ε. In addition, from the analysis, the eigenvalues in Λ_L with large magnitude (fast timescale) are less sensitive to the hierarchy level because $\Lambda_L \approx D_{II}$, and the elements in D_{II} does not depend on h_i. Therefore, the gradient of h_i across areas barely affects fast timescale pool. In contrast, the eigenvalues

in Λ_U with small magnitude (slow timescale) are more sensitive to the hierarchy level because $\Lambda_U \approx D_{EE} - D_{EI}D_{II}^{-1}D_{IE}$, and both the elements in D_{EE} and D_{IE} depend on h_i. Therefore, the gradient of h_i across areas increases the range of slow timescale pool. Further, the slow timescales of each area in this disconnected network are segregated and follows the hierarchical order h_i as the corresponding eigenvectors are perfectly localized and orthogonal to each other.

Now we consider the multi-areal network in the presence of long-range interactions. Adding long-range connectivity to local connectivity matrix D changes the eigenvalues and eigenvectors of matrix D, which can be analyzed in the following.

By multiplying P and P^{-1} (given in the proof of Proposition 1) on both sides of W, we have

$$\Gamma = P^{-1}WP = P^{-1}\left(\begin{bmatrix} D_{EE} & D_{EI} \\ D_{IE} & D_{II} \end{bmatrix} + \begin{bmatrix} F_{EE} & O \\ F_{IE} & O \end{bmatrix}\right)P = \Lambda + \Sigma,$$

where

$$\Sigma = P^{-1}\begin{bmatrix} F_{EE} & O \\ F_{IE} & O \end{bmatrix}P = \begin{bmatrix} \Sigma_{UL} & \Sigma_{UR} \\ \Sigma_{LL} & \Sigma_{LR} \end{bmatrix},$$

with $\Sigma_{UL} = (F_{EE} + AF_{IE})(I + AB)$, $\Sigma_{UR} = -(F_{EE} + AF_{IE})A$, $\Sigma_{LL} = \left[BF_{EE} + (I + BA)F_{IE}\right]\left(I + AB\right)$, $\Sigma_{LR} = -\left[BF_{EE} + (I + BA)F_{IE}\right]A$, and with Λ, A, and B defined in the proof of Proposition 1.

Denoting one of the eigenvectors of matrix Γ as $[\boldsymbol{u}, \boldsymbol{v}]^T$, and the corresponding eigenvalue as λ, we have

$$\left(\begin{bmatrix} \Lambda_U & 0 \\ 0 & \Lambda_L \end{bmatrix} + \begin{bmatrix} \Sigma_{UL} & \Sigma_{UR} \\ \Sigma_{LL} & \Sigma_{LR} \end{bmatrix}\right)\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \tag{6}$$

According to the definition of $\varepsilon = \frac{\beta_E}{\tau_E}/\frac{\beta_I}{\tau_I}$, and $\delta = \frac{\mu_{EE}}{\mu_{IE}} - \frac{w_{EI}}{w_{II}+1/\beta_I}$, $A \sim \mathscr{O}(\varepsilon)$, $B \sim \mathscr{O}(1)$, and $F_{EE} \sim \mathscr{O}(\varepsilon)$, $F_{IE} \sim \mathscr{O}(1)$, we have $F_{EE} + AF_{IE} \sim \mathscr{O}(\varepsilon\delta)$, $I + AB \sim \mathscr{O}(1)$, $BF_{EE} + (I + BA)F_{IE} \sim \mathscr{O}(1)$, and accordingly,

$$\Sigma_{UL} \sim \mathscr{O}(\varepsilon\delta), \quad \Sigma_{UR} \sim \mathscr{O}(\varepsilon^2\delta), \quad \Sigma_{LL} \sim \mathscr{O}(1), \quad \Sigma_{LR} \sim \mathscr{O}(\varepsilon).$$

As $\Sigma_{UR} \sim \mathscr{O}(\varepsilon^2\delta)$ is a higher order term compared with Λ_U, Λ_L, Σ_{UL}, Σ_{LL}, and Σ_{LR}, it can be dropped out in Eq. 6 and the error of eigenvalue and eigenvector is at most $\mathscr{O}(\varepsilon^2\delta)$ (see Proposition 4 in *Appendix* for a detailed proof). Accordingly, we can obtain two equations from Eq. 6 in the vector form,

$$(\Lambda_U + \Sigma_{UL})\boldsymbol{u} = \lambda\boldsymbol{u}, \tag{7}$$

$$\Sigma_{LL}\boldsymbol{u} + (\Lambda_L + \Sigma_{LR})\boldsymbol{v} = \lambda\boldsymbol{v}. \tag{8}$$

To describe the eigenvector property of Eqs. 7-8, we first introduce the definitions of weak localization and weak orthogonality as follows.

Definition 1 *A vector $\boldsymbol{u}(\delta)$ is weakly localized if it can be represented as $\boldsymbol{u} = a\boldsymbol{e_k} + \delta\boldsymbol{b} + \mathscr{O}(\delta^2)$ for some k, where $a \sim \mathscr{O}(1)$ is a constant number, $\boldsymbol{b} \sim \mathscr{O}(1)$ is a constant vector, δ is a small parameter, and $\boldsymbol{e}_k$ represents the natural basis with only the kth element being 1 and others being zero, i.e., $\boldsymbol{e}_k = [0, ..., 1(kth), ..., 0]$.*

Definition 2 *Two vectors $\boldsymbol{u}(\delta)$ and $\boldsymbol{v}(\delta)$ are weakly orthogonal to each other if their inner product $< \boldsymbol{u}, \boldsymbol{v} > \sim \mathscr{O}(\delta)$, where δ is a small parameter.*

With the concept of weak localization and weak orthogonality defined above, we introduce the following proposition that describes the property of $\boldsymbol{u}$ in the system of Eqs. 7-8.

Proposition 2 *In the system described by Eqs. 7-8, if all matrices are analytic with respect to ε and δ, and if $\Sigma_{UL} \sim \mathscr{O}(\varepsilon\delta)$, $\Sigma_{LL} \sim \mathscr{O}(1)$, $\Sigma_{LR} \sim \mathscr{O}(\varepsilon)$, $\Lambda_U \sim \mathscr{O}(\varepsilon)$, $\Lambda_L \sim \mathscr{O}(1)$, and if Λ_U has n simple eigenvalues, then*
(1) there exists n eigenvectors $[\boldsymbol{u},\boldsymbol{v}]^T$ in which $\boldsymbol{u}=0$, with $\lambda \sim \mathscr{O}(1)$ correspondingly.
(2) there exists n eigenvectors $[\boldsymbol{u},\boldsymbol{v}]^T$ in which $\boldsymbol{u}$ is weakly localized and weakly orthogonal to each other, with $\lambda \sim \mathscr{O}(\varepsilon)$ correspondingly.

Proof: (1) It is noted that $\boldsymbol{u}=0$ is a trivial solution of Eq. 7. By defining $\bar{\Sigma}_{LR} = \Sigma_{LR}/\varepsilon \sim \mathscr{O}(1)$, Eq. 8 becomes

$$(\Lambda_L + \varepsilon\bar{\Sigma}_{LR})\boldsymbol{v} = \lambda\boldsymbol{v},$$

in which $\boldsymbol{v}$ is the eigenvector of matrix $\Lambda_L + \varepsilon\bar{\Sigma}_{LR}$. By viewing $\varepsilon\bar{\Sigma}_{LR}$ as a perturbation matrix to Λ_L, then the leading order of λ shall be the same as that of n elements in the diagonal line of Λ_L, which takes the order of $\mathscr{O}(1)$.

(2) In Eq. 7, if $\boldsymbol{u} \neq 0$, by defining $\bar{\Lambda}_U = \Lambda_U/\varepsilon$, $\bar{\Sigma}_{UL} = \Sigma_{UL}/\varepsilon\delta$, and $\bar{\lambda} = \lambda/\varepsilon$, we have

$$(\bar{\Lambda}_U + \delta\bar{\Sigma}_{UL})\boldsymbol{u} = \bar{\lambda}\boldsymbol{u}. \tag{9}$$

Therefore, $\boldsymbol{u}$ is also the eigenvector of matrix $\bar{\Lambda}_U + \delta\bar{\Sigma}_{UL}$, and $\bar{\lambda}$ is the corresponding eigenvalue. As Λ_U has n simple eigenvalues, so does $\bar{\Lambda}_U$, then $\boldsymbol{u}$ and $\bar{\lambda}$ are analytic with respect to the perturbation parameter δ (Kato, 1966), i.e., $\boldsymbol{u} = \sum_{i=0}^{\infty}\delta^i\boldsymbol{u}_i$, and $\bar{\lambda} = \sum_{j=0}^{\infty}\delta^j\mu_i$ for δ near zero. Therefore, to the leading order, we have

$$\bar{\Lambda}_U\boldsymbol{u}_0 = \mu_0\boldsymbol{u}_0,$$

in which μ_0 is the eigenvalue of the diagonal matrix $\bar{\Lambda}_U$, and $\boldsymbol{u}_0$ is the corresponding eigenvector. Accordingly, $\boldsymbol{u}_0 \in \{\boldsymbol{e}_k\}$, thereafter

$$\boldsymbol{u} = \boldsymbol{e}_k + \delta\boldsymbol{u}_1 + \mathscr{O}(\delta^2), \qquad k = 1,2,...,n$$

where $\boldsymbol{e}_k$ represents the kth natural basis, and the leading order of $\lambda = \varepsilon\bar{\lambda}$ is $\varepsilon\mu_0 \sim \mathscr{O}(\varepsilon)$. It's straightforward to verify that $\boldsymbol{u}$ are weakly localized and weakly orthogonal to each other. ■

If we denote the unit-length eigenvector of the connectivity matrix W as $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$, and denote the corresponding eigenvalue as λ (the same as that of matrix Γ after the similarity transform), then from Propositions 1 and 2, we have

Proposition 3 *Under the same conditions in Propositions 1-2, the unit-length eigenvector $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$ of the connectivity matrix W has the following properties,*
(1) for eigenvalue $\lambda \sim \mathscr{O}(\varepsilon)$, the corresponding $\boldsymbol{r}_E$ is weakly localized and weakly orthogonal to each other.
(2) for eigenvalue $\lambda \sim \mathscr{O}(1)$, the corresponding $\boldsymbol{r}_E \sim \mathscr{O}(\varepsilon)$.

Proof: We first consider $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$ with non-unit length. According to the similarity transform, we have the following linear relation between $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$ and $[\boldsymbol{u},\boldsymbol{v}]^T$,

$$\begin{bmatrix}\boldsymbol{r}_E\\ \boldsymbol{r}_I\end{bmatrix} = P\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{bmatrix} = \begin{bmatrix}I+AB & -A\\ -B & I\end{bmatrix}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{bmatrix},$$

i.e., $\boldsymbol{r}_E = (I+AB)\boldsymbol{u} - A\boldsymbol{v}$, and $\boldsymbol{r}_I = -B\boldsymbol{u} + \boldsymbol{v}$. From Proposition 1, we have $A \sim \mathscr{O}(\varepsilon)$, and $B \sim \mathscr{O}(1)$.

(1) From Proposition 2, we have $\boldsymbol{u} = \boldsymbol{e}_k + \delta\boldsymbol{u}_1 + \mathscr{O}(\delta^2)$ for $\lambda \sim \mathscr{O}(\varepsilon)$ $(k = 1,2,...,n)$. Accordingly, $\boldsymbol{v}$ can be solved as $\boldsymbol{v} = (\lambda I - \Lambda_L - \Sigma_{LR})^{-1}\Sigma_{LL}\boldsymbol{u} \sim \mathscr{O}(1)$, which gives $\boldsymbol{r}_E \sim \mathscr{O}(1)$, and $\boldsymbol{r}_I \sim \mathscr{O}(1)$. Therefore, the length of $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$ denoted by c is order $\mathscr{O}(1)$. By normalizing the length of $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$ to be unity, we have $\boldsymbol{r}_E = c^{-1}(I+AB)\boldsymbol{e}_k + \mathscr{O}(\varepsilon) + \mathscr{O}(\delta)$ being weakly

localized and weakly orthogonal to each other.

(2) From Proposition 2, we have $\boldsymbol{u}=0$ for $\lambda \sim \mathscr{O}(1)$. Accordingly, $\boldsymbol{v}$ can be solved as the eigenvector of matrix $(\Lambda_L+\Sigma_{LR})$ with unit-length. Therefore, $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T=[-A\boldsymbol{v},\boldsymbol{v}]^T$, and the length of $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$ denoted by c is order $\mathscr{O}(1)$. By normalizing the length of $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$ to be unity, we have $\boldsymbol{r}_E=-c^{-1}A\boldsymbol{v}\sim\mathscr{O}(\varepsilon)$. ■

Note that, Propositions 2-3 hold for sufficiently small ε and δ near zero. However, the convergence radius of the power series of $\boldsymbol{u}$ in Proposition 2 is not specified yet. Although difficult to calculate the convergence radius, we can compute the analytical expression of $\boldsymbol{u}_1$ in the power series $\boldsymbol{u}=\sum_{i=0}^{\infty}\delta^i\boldsymbol{u}_i$ in Proposition 2 to obtain the first-order perturbation solution of $\boldsymbol{u}$ and thereby $\boldsymbol{r}_E$, which could help us gain insight about when weak localization and orthogonality of $\boldsymbol{u}$ and $\boldsymbol{r}_E$ will break down approximately.

To the order of δ in Eq. 9, we have

$$\bar{\Lambda}_U\boldsymbol{u}_1+\bar{\Sigma}_{UL}\boldsymbol{u}_0=\mu_0\boldsymbol{u}_1+\mu_1\boldsymbol{u}_0. \tag{10}$$

Without loss of generality, we assume $\boldsymbol{u}_0=\boldsymbol{e}_k$, and accordingly, $\mu_0=\bar{\lambda}_k$ is the kth element in the diagonal line of matrix $\bar{\Lambda}_U$. In addition, we normalize $\boldsymbol{u}$ to make the kth element in $\boldsymbol{u}$ denoted by $\boldsymbol{u}^k$ to be unity, correspondingly $\boldsymbol{u}_i^k=0$ for $i\geq 1$. By left multiplying $\boldsymbol{e}_j^T$ $(j\neq k)$ to Eq. 10, we have

$$\boldsymbol{u}_1^j=\frac{\bar{\Sigma}_{UL}^{jk}}{\bar{\lambda}_k-\bar{\lambda}_j}, \tag{11}$$

where $\bar{\Sigma}_{UL}^{jk}$ is the element in the jth row and kth column in matrix $\bar{\Sigma}_{UL}$. To make the first-order perturbation solution valid, we expect that $\boldsymbol{u}_1^j$ is small compared with δ^{-1}, otherwise the separation of orders will no longer hold in the power series (i.e., first-order term $\delta\boldsymbol{u}_1$ becomes larger than the zeroth-order term $\boldsymbol{e}_k$). In such a case, the spectral gap $\bar{\lambda}_k-\bar{\lambda}_j$ shall be large enough compared with elements in $\bar{\Sigma}_{UL}$. The spectral gap of matrix $\bar{\Lambda}_U$ attributes to the gradient of excitation across areas, or simply h_i. In Fig. 2B-2C, we show that the eigenvector $[\boldsymbol{r}_E,\boldsymbol{r}_I]^T$, which is solved using the perturbation theory in Proportions 2-3 to the first-order accuracy (Eq. 11), well agrees with the original eigenvector in most cases. However, some eigenvectors show less similarity to the original eigenvectors when the first-order perturbation theory breaks down for the reason discussed above.

Biological interpretations of the three requirements

From the above analysis, three conditions are required to obtain weakly localized and orthogonal eigenvectors in order to maintain the hierarchy of timescales: (1) small ε, (2) small δ, (3) the gradient of h_i across areas. We briefly summarize the important roles of the three conditions on proving eigenvector localization and orthogonality in the perturbation analysis illustrated in Fig. 3. As shown in Fig. 3A-3B, to remove intra-areal interactions between the excitatory and inhibitory populations within each area, we first change the coordinate system from $(\boldsymbol{r}_E,\boldsymbol{r}_I)$ to $(\boldsymbol{u},\boldsymbol{v})$ with a transform matrix P given in Proposition 1. In the new coordinate system, there is no local interaction between the dynamical variables $\boldsymbol{u}$ and $\boldsymbol{v}$. Furthermore, considering a directed long-range projection from area j to area i in Fig. 3C, it has been shown that small δ leads to weak interaction from u^j to u^i, and small ε additionally leads to even weaker interaction from v^j to u^i that can be removed with an ignorable error (see Proposition 4 in Appendix). And a gradient of h_i leads to a nonzero spectral gap between area i and area j. All the three conditions result in the weak localization and orthogonality of the $\boldsymbol{u}$ component in the $(\boldsymbol{u},\boldsymbol{v})$ coordinate system, as proved by Proposition 2 and Eq. 11. Finally, as shown in Fig. 3D, small ε gives rise to the fact that the leading order of $\boldsymbol{u}$ and $\boldsymbol{r}_E$ are the same, thus $\boldsymbol{r}_E$ is also weakly localized and orthogonalized similar to $\boldsymbol{u}$, as proved by Proposition 3.

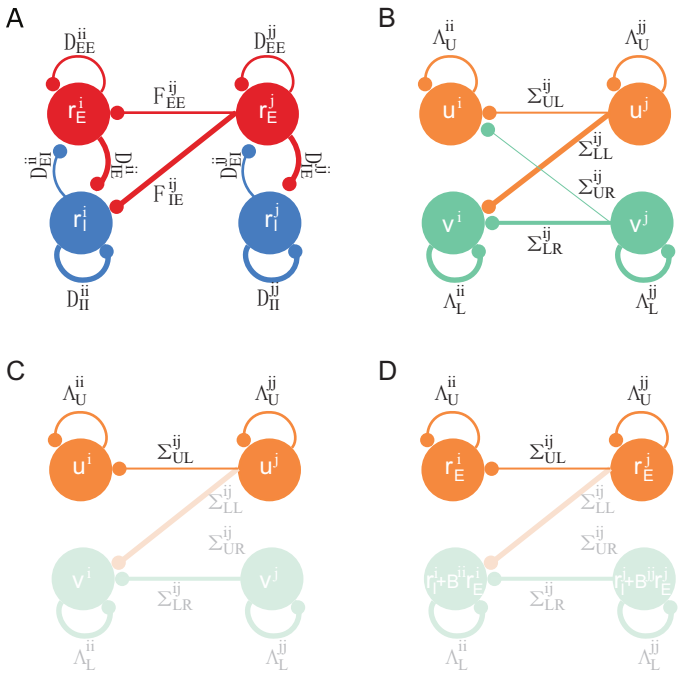

Figure 3: Schematics for the proof of weakly localized and orthogonal eigenvectors of the connectivity matrix W. (A), directed interaction from area j to area i in the original model (Eqs. 1-2). (B), one-way interaction from area j to area i after changing the coordinate system from $(\boldsymbol{r}_E, \boldsymbol{r}_I)$ to $(\boldsymbol{u}, \boldsymbol{v})$. (C), small δ leads to weak interaction from u^j to u^i, small ε additionally leads to even weaker interaction from v^j to u^i that is ignorable (proved by Proposition 4 in Appendix), and a gradient of h_i leads to a nonzero spectral gap between area i and area j. Accordingly, they together lead to the weak localization and orthogonality of the $\boldsymbol{u}$ component in the $(\boldsymbol{u}, \boldsymbol{v})$ coordinate system (proved by Proposition 2 and Eq. 11). (D), one-way interaction from area j to area i after changing the coordinate system from $(\boldsymbol{u}, \boldsymbol{v})$ back to $(\boldsymbol{r}_E, \boldsymbol{r}_I)$. To the leading order, one has $u^i \approx r_E^i$, $v \approx r_I^i + B^{ii} r_E^i$. In this step, small ε ensures that the leading order of $\boldsymbol{u}$ and $\boldsymbol{r}_E$ are identical, so are their localization and orthogonality properties (proved by Proposition 3). In (A)-(D), the width of lines codes the interaction strength, and light colored lines and nodes are not important in the proofs.

We next discuss the biological interpretation of the three conditions. First, according to the definition of $\varepsilon = \frac{\beta_E}{\tau_E} / \frac{\beta_I}{\tau_I}$, small ε indicates that the electrophysiological properties of excitatory and inhibitory neurons are different, in particular, their membrane time constant and the slope of the gain function. The substantial difference of electrophysiological properties between the excitatory and inhibitory neurons has been supported by experimental evidence, i.e., inhibitory neurons have larger slope of the gain function and smaller membrane time constant (Ahmed et al., 1998; Nowak et al., 2003; Povysheva et al., 2008; Zaitsev et al., 2012).

Second, small $\delta = \frac{\mu_{EE}}{\mu_{IE}} - \frac{w_{EI}}{w_{II}+1/\beta_I}$ indicates the balanced condition between the inter-areal excitatory and intra-areal inhibitory inputs. When the presynaptic excitatory input from the jth area is increased by Δr_E^j, its influence on the excitatory population activity in the ith area in the steady state can be calculated in a straightforward way as $\Delta r_E^i = C^{ij} \Delta r_E^j$, in which $C^{ij} \sim \mathscr{O}(\delta)$. This indicates that the signal from the jth area have a small influence on the activity of the excitatory population in the ith area, because the global long-range excitatory input is balanced with and canceled by the local inhibitory synaptic input, leading to small net inputs in each signal pathway, as shown in Fig. 4. This condition corresponds to a detailed balance of excitation and inhibition that may benefit signal control and gating, as proposed in previous studies (Vogels and Abbott, 2009). The importance of excitation-inhibition balance on timescale hierarchy is supported by a recent study showing that the imbalance of excitation and inhibition could have a substantial effect on the change of intrinsic timescales across brain areas, which is a manifestation of psychosis such as hallucination and delusion (Wengler et al., 2020).

Third, the gradient of h_i parameterizes the gradient of synaptic excitation across areas in the model, supported by the fact that h_i is proportional to the spine count per pyramidal neuron across areas (Chaudhuri et al., 2015; Elston, 2007) in the form of a macroscopic gradient (Wang, 2020). The gradient of synaptic excitation leads to two consequences, (1) it gives rise to the hierarchy of intrinsic timescale for each area while being disconnected to other parts of the cortex, (2) it stabilizes the localization of intrinsic timescale for each area in the presence of long-range connections. From the perturbation analysis and Eq. 11, the degree of eigenvector localization is determined by the competition between the strength of long-range connections encoded in matrix $\bar{\Sigma}_{UL}$ and the spectral gap of matrix $\bar{\Lambda}_U$. Therefore, the long-range connections tend to delocalize eigenvectors thus break the timescale hierarchy, but the heterogeneity of local recurrent excitation level weakens its effect on eigenvector delocalization in a divisive fashion. In fact, the heterogeneity or randomness in local node properties has been shown to give localized eigenvectors in models of physical system, for instance, a phenomenon known as Anderson localization (Anderson, 1958) that describes the transition from a conducting medium (corresponding to delocalized eigenvectors) to an insulating medium (corresponding to localized eigenvectors). Similar mechanism has been identified in studying the eigenvector localization of a idealized neural network with simple nodes in each cortical area (Chaudhuri et al., 2014).

Discussion

In this work, we investigated the requirements for the emergence of a hierarchy of temporal response windows in a multi-areal model of the macaque cortex (Chaudhuri et al., 2015). The original model is a nonlinear dynamical system by including a rectified linear f-I curve, and it becomes essentially linear when neural population activities are all above the firing threshold, as happened in our simulations of hierarchical timescale phenomenon. This fact enabled us to define the time constants precisely from the eigenmodes of the connectivity matrix, and carried out a detailed mathematical analysis to identify biologically interpretable conditions. (Rectified) linear models have been broadly used in theoretical and experimental neuroscience studies (Roxin et al., 2005; Binzegger et al., 2009; Jercog et al., 2017). Although microscopic neural activity is nonlinear in general, it has been shown in a recent study (Nozari et al., 2020) that linear models can capture macroscopic cortical dynamics in the resting state more accurately than nonlinear model families, including neural field models for describing the spatiotemporal average of individual neuronal activities. Nonlinear models are more general

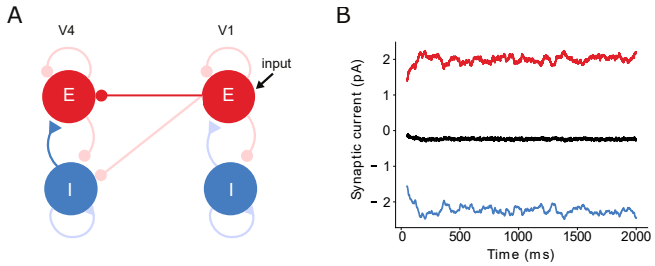

Figure 4: The illustration of detailed balance between inter-areal excitation and intra-areal inhibition. The projection from V1 to V4 is shown as an example. (A), one-way interaction from V1 to V4. V1 receives external Gaussian input. The excitatory population in V4 receives balanced excitatory inter-areal inputs from V1 (dark red) and intra-areal inhibitory inputs from the inhibitory population in V4 (dark blue). Other excitatory and inhibitory interactions in this circuit are colored by light red and blue, respectively. (B), simulation of the synaptic currents received by V4 excitatory population induced by V1 activity. The inter-areal excitatory inputs (red) is balanced with the intra-areal inhibitory inputs (blue), leading to small net inputs (black).

for capturing neural circuits. However, for a nonlinear model, the time constants of the system are not uniquely defined. A linear model can be understood as a linearization of a nonlinear dynamical system around an internal state such as the resting state of the brain.

In contrast to previous computational models studying the emergence of timescales (Chaudhuri et al., 2014), the model we studied is anatomically more realistic as it incorporates (1) experimental measurements of directed and weighted anatomical connectivity, (2) a gradient of synaptic excitation reflected by spine counts in pyramidal neurons across areas, and (3) both excitatory and inhibitory neuronal populations. By performing rigorous perturbation analysis, we show that the segregation of timescales is attributable to the localization of eigenvectors of the connectivity matrix, and the parameter regime that makes this happen has three crucial properties: (1) a macroscopic gradient of synaptic excitation, (2) distinct electrophysiological properties between excitatory and inhibitory neuronal populations, and (3) a detailed balance between long-range excitatory inputs and local inhibitory inputs for each area-to-area pathway.

The theoretically identified biological conditions for the segregation of timescales enable us to make experimentally testable predictions. First, the condition of the macroscopic gradient of synaptic excitation suggests that shallower gradient of synaptic excitation shall lead to less localized eigenvectors. Consequently, the difference of time constants for a pair of areas shall be larger if their synaptic excitation or hierarchical level are less similar, which can be directly tested in experiments. Second, the condition of distinct electrophysiological properties between excitatory and inhibitory neuronal populations suggests that the change of neuronal physiology will affect the segregation of timescales. This condition can be tested experimentally by using genetic tools to knockdown or knockout specific genes in order to change the firing properties of neurons (Gingras et al., 2014). Third, the condition of detailed balance of excitation and inhibition suggests that areas with unbalanced excitation and inhibition could also alter hierarchical time constants. With growing evidence for E/I imbalance in schizophrenia (Foss-Feig et al., 2017; Jardri et al., 2016), this condition is supported in recent experiments that the intrinsic time constants of schizophrenia patients have been substantially changed (Wengler et al., 2020). And this condition can be further tested in animal models with genetic tools to disturb the excitation-inhibition balance (Gatto and Broadie, 2010).

Beyond neuroscience, eigenmode localization is an important phenomenon in nature that has aroused broad interests. A prominent example is known as Anderson localization in physics (Anderson, 1958), finding that the eigenmodes of the Schrödinger equation that govern the wave function of electrons can be spatially localized with heterogeneous potentials. Anderson localization successfully explains the transition from metal to insulator given impurities or defects in materials. In addition to that, light (Wiersma et al., 1997), acoustic waves (Hu et al., 2008), seismic

waves (Larose et al., 2004) and others (Billy et al., 2008; Choi et al., 2017; White et al., 2020) in heterogeneous medias have also been found to exhibit localized modes. The interactions in these systems are short-range, in contrast to brain circuits with an abundance of long-range connections. Here we found that eigenmodes can also be localized in the cortical network with similar but different mechanisms from the aforementioned systems. The localization in the cortex also requires heterogeneous local properties (i.e., recurrent excitation level), and it requires additional distinct properties of the excitatory and inhibitory neurons and their balanced interactions. Our work thus provided another example of eigenmode localization in natural systems and demonstrated that the intrinsic properties of "particles" in the system could be important to eigenmode localization.

It is worth mentioning that, although the specific pattern of inter-areal connectivity does not affect the eigenvector localization substantially based on the perturbation analysis, it shapes the timescale hierarchy qualitatively. In particular, the timescale hierarchy does not exactly follow monotonically the areal anatomical hierarchy in the presence of long-range connections, as shown in Fig. 1C. Furthermore, within a brain region time constants are heterogeneous across individual neurons (Bernacchia et al., 2011; Cavanagh et al., 2020). To better relate the model with experimentally observed timescales in various specific cortical areas, the roles of long-range connections, cell types and other circuit properties require further elucidation.

It has been noticed that the neuronal activity propagates along the hierarchy with significant attenuation in the model of Ref. (Chaudhuri et al., 2015). The attenuation can be alleviated by tuning the model parameters to the regime of strong global balanced amplification (GBA) (Joglekar et al., 2018) (see parameters in Sec. *Materials and Methods*). Balanced amplification was originally introduced for a local neural network, associated with strong non-normality of the system where eigenmodes are far from being orthogonal with each other (Murphy and Miller, 2009). A quantity called κ measures the degree of non-normality of a matrix (Trefethen and Embree, 2020) ($\kappa = 1$ for a normal matrix; the larger the κ value, the more non-normal the system). We have $\kappa = 4.35$ for the original model (Chaudhuri et al., 2015), which is thus only slightly non-normal. By contrast, $\kappa = 96.58$ for the model in the strong GBA regime. Therefore, the enhancement of signal propagation in the model correlates with the increase of the non-orthogonality of the eigenvectors, or the non-normality of the connectivity matrix. In the strong GBA regime, $\delta \approx 0.38$, which is ten times larger than its original value, suggesting that the detailed balance condition is less well satisfied. In such a case, the localization of time scales may no longer exist in this linear model. The situation is different in nonlinear models (Chaudhuri et al., 2015; Mejias and Wang, 2020; Chien and Honey, 2020), where inputs may be amplified by strongly recurrent circuit dynamics to enhance signal propagation or routing of information is selectively gated (for a subset of connection pathways in a goal-directed manner) (Vogels and Abbott, 2009; Wang and Yang, 2018), while the conditions for a timescale hierarchy are satisfied. For a nonlinear system, however, eigenmodes can be defined only with respect to a particular network state. Consequently, the time constants observed in single neurons are no longer unique and may differ, for instance, when the brain is at rest or during a cognitive process. It remains to be seen to what extent the conditions identified here hold in various brain's internal states, while the precise pattern of timescales can be flexibly varied to meet behavioral demands.

Acknowledgements

We thank Y. Xiao and H. Wang for helpful discussions. This work was supported by National Key R&D Program of China 2019YFA0709503, Shanghai Municipal Science and Technology Major Project 2021SHZDZX0102, National Natural Science Foundation of China Grant 11901388, and Shanghai Sailing Program 19YF1421400 (S.L.), James Simons foundation grant 543057SPI (X.-J.W.)

Author Contributions

S.L. and X.-J.W. designed the research, S.L. performed the work, S.L. and X.-J.W. wrote the paper.

Materials and Methods

Model parameters

In the macaque cortical network model, we set $\tau_E = 20$ ms, $\tau_I = 10$ ms, $\beta_E = 0.066$ Hz/pA, $\beta_I = 0.351$ Hz/pA, $w_{EE} = 24.4$ pA/Hz, $w_{IE} = 12.2$ pA/Hz, $w_{EI} = 19.7$ pA/Hz, $w_{II} = 12.5$ pA/Hz, $\mu_{EE} = 33.7$ pA/Hz, $\mu_{IE} = 25.5$ pA/Hz and $\eta = 0.68$. We set $w_{EI} = 25.2$ pA/Hz and $\mu_{EE} = 51.5$ pA/Hz for the strong balanced amplification regime (Joglekar et al., 2018) introduced in Sec. Discussion. Some of the parameters are derived from experimental measurements of primary visual cortex (Binzegger et al., 2009). The FLN values are obtained from the experimental measurements of macaque cortical connectivity (Markov et al., 2014a). The hierarchy values h_i of each cortical area are obtained by fitting a generalized linear model that assigns hierarchical values to areas (Chaudhuri et al., 2015) such that the differences in hierarchical values predict the supragranular layer neurons (SLNs) measured in experiment (Markov et al., 2014b).

Code Availability

Upon publication, the python code for model simulation and perturbation analysis will be made publicly available on Github (https://github.com/songting858).

References

Ahmed B., Anderson J., Douglas R., Martin K.A., and Whitteridge D. (1998). Estimates of the net excitatory currents evoked by visual stimulation of identified neurons in cat visual cortex. Cerebral Cortex *8*, 462–476.

Anderson P. (1958). Absence of diffusion in certain random lattices. Physical Review *109*, 1492–1505.

Bernacchia A., Seo H., Lee D., and Wang X.J. (2011). A reservoir of time constants for memory traces in cortical neurons. Nature Neuroscience *14*, 366–372.

Billy J., Josse V., Zuo Z., Bernard A., Hambrecht B., Lugan P., Clement D., Sanchez-Palencia L., Bouyer P., and Aspect A. (2008). Direct observation of anderson localization of matter waves in a controlled disorder. Nature *453*, 891–894.

Binzegger T., Douglas R.J., and Martin K.A. (2009). Topology and dynamics of the canonical circuit of cat v1. Neural Networks *22*, 1071–1078.

Cavanagh S.E., Hunt L.T., and Kennerley S.W. (2020). A diversity of intrinsic timescales underlie neural computations. Frontiers in Neural Circuits *14*, 615626.

Cavanagh S.E., Wallis J.D., Kennerley S.W., and Hunt L.T. (2016). Autocorrelation structure at rest predicts value correlates of single neurons during reward-guided choice. eLife *5*, e18937.

Chaudhuri R., Bernacchia A., and Wang X.J. (2014). A diversity of localized timescales in network activity. eLife *3*, e01239.

Chaudhuri R., Knoblauch K., Gariel M.A., Kennedy H., and Wang X.J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. Neuron *88*, 419–431.

Chien H.Y.S., and Honey C.J. (2020). Constructing and forgetting temporal context in the human cerebral cortex. Neuron *106*, 675–686.

Choi S.H., Kim S.W., Ku Z., Visbal-Onufrak M.A., Kim S.R., Choi K.H., Ko H., Choi W., Urbas A., Goo T., and Kim Y. (2017). Anderson light localization in biological nanostructures of native silk. Nature Communications *9:452*.

Elston G. (2007). Specialization of the neocortical pyramidal cell during primate evolution. In Evolution of Nervous Systems: A Comprehensive Reference, Volume 4 (Amsterdam: Elsevier).

Fascianelli V., Tsujimoto S., Marcos E., and Genovesio A. (2019). Autocorrelation structure in the macaque dorsolateral, but not orbital or polar, prefrontal cortex predicts response-coding strength in a visually cued strategy task. Cerebral cortex *29*, 230–241.

Foss-Feig J., Adkinson B., Ji J.L., Yang G., Srihari V., McPartland J., Krystal J., Murray J., and Anticevic A. (2017). Searching for cross-diagnostic convergence: Neural mechanisms governing excitation and inhibition balance in schizophrenia and autism spectrum disorders. Biological Psychiatry *81*, 848–861.

Gamanut R., Kennedy H., Toroczkai Z., Ercsey-Ravasz M., Van Essen D.C., Knoblauch K., and Burkhalter A. (2018). The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. Neuron *97*, 698–715.

Gao R., van den Brink R.L., Pfeffer T., and Voytek B. (2020). Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. eLife *9*, e61277.

Gatto C., and Broadie K. (2010). Genetic controls balancing excitatory and inhibitory synaptogenesis in neurodevelopmental disorder models. Frontiers in Synaptic Neuroscience *2:4 doi: 10.3389/fnsyn.2010.00004*.

Gingras J., Smith S.J., Matson D., Johnson D., Nye K., Couture L., Ferić E., Yin R., Moyer B., Peterson M., Rottman J., Beiler R., Malmberg A., and McDonough S. (2014). Global nav1.7 knockout mice recapitulate the phenotype of human congenital indifference to pain. PLoS ONE *9(9): e105895*.

Gold J.I., and Shadlen M.N. (2007). The neural basis of decision making. Annual Review of Neuroscience *30*, 535–574.

Hasson U., Yang E., Vallines I., Heeger D.J., and Rubin N. (2008). A hierarchy of temporal receptive windows in human cortex. Journal of Neuroscience *28*, 2539–2550.

Honey C.J., Thesen T., Donner T.H., Silbert L.J., Carlson C.E., Devinsky O., Doyle W.K., Rubin N., Heeger D.J., and Hasson U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. Neuron *76*, 423–434.

Hu H., Strybulevych A., Page J., Skipetrov S., and Tiggelen B.A. (2008). Localization of ultrasound in a three-dimensional elastic network. Nature Physics *4*, 945–948.

Hubel D.H. (1995). Eye, brain, and vision. (Scientific American Library/Scientific American Books).

Jardri R., Hugdahl K., Hughes M., Brunelin J., Waters F., Alderson-Day B., Smailes D., Sterzer P., Corlett P., Leptourgos P., Debbané M., Cachia A., and Denéve S. (2016). Are hallucinations due to an imbalance between excitatory and inhibitory influences on the brain? Schizophrenia bulletin *42(5)*, 1124–1134.

Jercog D., Roxin A., Barthó P., Luczak A., Compte A., and de la Rocha J. (2017). Up-down cortical dynamics reflect state transitions in a bistable network. eLife *6:e22425*.

Joglekar M.R., Mejías J., Yang G.R., and Wang X.J. (2018). Inter-areal balanced amplification enhances signal propagation in a large-scale circuit model of the primate cortex. Neuron *98*, 222–234.

Kato T. (1966). Perturbation theory for linear operators (Berlin: Springer-Verlag).

Larose E., Margerin L., van Tiggelen B.V., and Campillo M. (2004). Weak localization of seismic waves. Physical review letters *93 4*, 048501.

Lerner Y., Honey C.J., Silbert L.J., and Hasson U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. Journal of Neuroscience *31*, 2906–2915.

Maisson D.J.N., Cash-Padgett T.V., Wang M.Z., Hayden B.Y., Heilbronner S.R., and Zimmermann J. (2021). Choice-relevant information transformation along a ventrodorsal axis in the medial prefrontal cortex. Nature communications *12*, 4830.

Manea A.M.G., Zilverstand A., Uğurbil K., Heilbronner S.R., and Zimmermann J. (2021). Intrinsic timescales as an organizational principle of neural processing across the whole rhesus macaque brain. bioRxiv *463277*.

Markov N., Ercsey-Ravasz M., Gomes A.R.R., Lamy C., Magrou L., Vezoli J., Misery P., Falchier A., Quilodran R., Gariel M., Sallet J., Gamanut R., Huissoud C., Clavagnier S., Giroud P., Sappey-Marinier D., Barone P., Dehay C., Toroczkai Z., Knoblauch K., Essen D.V.V., and Kennedy H. (2014a). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cerebral Cortex *24*, 17–36.

Markov N., Misery P., Falchier A., Lamy C., Vezoli J., Quilodran R., Gariel M., Giroud P., Ercsey-Ravasz M., Pilaz L., et al. (2011). Weight consistency specifies regularities of macaque cortical networks. Cerebral Cortex *21*, 1254–1272.

Markov N., Vezoli J., Chameau P., Falchier A., Quilodran R., Huissoud C., Lamy C., Misery P., Giroud P., Ullman S., Barone P., Dehay C., Knoblauch K., and Kennedy H. (2014b). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. The Journal of Comparative Neurology *522*, 225–259.

Mejias J.F., and Wang X.J. (2020). Mechanisms of distributed working memory in a large-scale model of the macaque neocortex. BioRxiv p. 760231.

Murphy B.K., and Miller K. (2009). Balanced amplification: A new mechanism of selective amplification of neural activity patterns. Neuron *61*, 635–648.

Murray J.D., Bernacchia A., Freedman D.J., Romo R., Wallis J.D., Cai X., Padoa-Schioppa C., Pasternak T., Seo H., Lee D., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. Nature Neuroscience *17*, 1661–1663.

Nowak L., Azouz R., Sanchez-Vives M.V., Gray C., and McCormick D. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. Journal of Neurophysiology *89*, 1541–1566.

Nozari E., Stiso J., Caciagli L., Cornblath E.J., He X., Bertolero M., Mahadevan A., Pappas G.J., and Bassett D. (2020). Is the brain macroscopically linear? a system identification of resting state dynamics. bioRxiv *423856*.

Ogawa T., and Komatsu H. (2010). Differential temporal storage capacity in the baseline activity of neurons in macaque frontal eye field and area v4. Journal of neurophysiology *103*, 2433–2445.

Povysheva N.V., Zaitsev A., Rotaru D., Gonzalez-Burgos G., Lewis D., and Krimer L.S. (2008). Parvalbumin-positive basket interneurons in monkey and rat prefrontal cortex. Journal of Neurophysiology *100*, 2348–2360.

Raut R.V., Snyder A.Z., and Raichle M.E. (2020). Hierarchical dynamics as a macroscopic organizing principle of the human brain. Proceedings of the National Academy of Sciences *117*, 20890–20897.

Roxin A., Brunel N., and Hansel D. (2005). Role of delays in shaping spatiotemporal dynamics of neuronal activity in large networks. Physical review letters *94 23*, 238103.

Runyan C.A., Piasini E., Panzeri S., and Harvey C.D. (2017). Distinct timescales of population coding across cortex. Nature *548*, 92–96.

Shafiei G., Markello R.D., De Wael R.V., Bernhardt B.C., Fulcher B.D., and Misic B. (2020). Topographic gradients of intrinsic dynamics across neocortex. eLife *9*, e62116.

Siegle J.H., Jia X., Durand S., Gale S., Bennett C., Graddis N., Heller G., Ramirez T.K., Choi H., Luviano J.A., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. Nature *592*, 86–92.

Spitmaan M., Seo H., Lee D., and Soltani A. (2020). Multiple timescales of neural dynamics and integration of task-relevant signals across cortex. Proceedings of the National Academy of Sciences *117*, 22522–22531.

Stephens G.J., Honey C.J., and Hasson U. (2013). A place for time: the spatiotemporal structure of neural dynamics during natural audition. Journal of Neurophysiology *110*, 2019–2026.

Trefethen L.N., and Embree M. (2020). Spectra and pseudospectra (Princeton University Press).

Vogels T.P., and Abbott L. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. Nature Neuroscience *12*, 483–491.

Wang X.J. (2008). Decision making in recurrent neuronal circuits. Neuron *60*, 215–234.

Wang X.J. (2020). Macroscopic gradients of synaptic excitation and inhibition in the neocortex. Nature Reviews Neuroscience *21*, 169–178.

Wang X.J., and Yang G.R. (2018). A disinhibitory circuit motif and flexible information routing in the brain. Current Opinion in Neurobiology *49*, 75–83.

Wengler K., Goldberg A., Chahine G., and Horga G. (2020). Distinct hierarchical alterations of intrinsic neural timescales account for different manifestations of psychosis. eLife *9*, e56151.

White D., Haase T., Brown D.J., Hoogerland M., Najafabadi M.S., Helm J., Gies C., Schumayer D., and Hutchinson D. (2020). Observation of two-dimensional anderson localisation of ultracold atoms. Nature Communications *11:4942*.

Wiersma D., Bartolini P., Lagendijk A., and Righini R. (1997). Localization of light in a disordered medium. Nature *390*, 671–673.

Yeshurun Y., Nguyen M., and Hasson U. (2017). Amplification of local changes along the timescale processing hierarchy. Proceedings of the National Academy of Sciences *114*, 9475–9480.

Zaitsev A., Povysheva N.V., Gonzalez-Burgos G., and Lewis D. (2012). Electrophysiological classes of layer 2/3 pyramidal cells in monkey prefrontal cortex. Journal of Neurophysiology *108*, 595–609.

Appendix

Proposition 4 *Given*

$$\Gamma = \begin{bmatrix} \varepsilon\bar{\Lambda}_U & 0 \\ 0 & \bar{\Lambda}_L \end{bmatrix} + \begin{bmatrix} \varepsilon\delta\bar{\Sigma}_{UL} & \varepsilon^2\delta\bar{\Sigma}_{UR} \\ \bar{\Sigma}_{LL} & \varepsilon\bar{\Sigma}_{LR} \end{bmatrix}$$

and

$$\Gamma' = \Gamma - \begin{bmatrix} 0 & \varepsilon^2\delta\bar{\Sigma}_{UR} \\ 0 & 0 \end{bmatrix}$$

with $\bar{\Lambda}_U, \bar{\Lambda}_L, \bar{\Sigma}_{UL}, \bar{\Sigma}_{UR}, \bar{\Sigma}_{LL}, \bar{\Sigma}_{LR}$ being $\mathscr{O}(1)$, and A_U has n simple eigenvalues. If λ and $\boldsymbol{X}$ are the eigenvalue and eigenvector of Γ respectively, then there exists λ' and $\boldsymbol{X'}$ being the eigenvalue and eigenvector of Γ' respectively, such that $\lambda - \lambda' \sim \mathscr{O}(\varepsilon^2\delta)$ or higher order, $\boldsymbol{X} - \boldsymbol{X'} \sim \mathscr{O}(\varepsilon^2\delta)$ or higher order.

Proof: The eigenvalue of Γ denoted by λ can be solved as the root of the characteristic polynomial

$$f(\lambda) = det(\lambda I - \Gamma) = f_1(\lambda) f_2(\lambda),$$

where

$$f_1(\lambda) = det\Big(\lambda I - \varepsilon\bar{\Lambda}_U - \varepsilon\delta\bar{\Sigma}_{UL} - \varepsilon^2\delta\bar{\Sigma}_{UR}(\lambda I - \bar{\Lambda}_L - \varepsilon\bar{\Sigma}_{LR})^{-1}\bar{\Sigma}_{LL}\Big),$$

and

$$f_2(\lambda) = det(\lambda I - \bar{\Lambda}_L - \varepsilon\bar{\Sigma}_{LR}),$$

here $det(\cdot)$ is the determinant of a matrix, and I is the identity matrix.

For the eigenvalues that solve $f_1(\lambda) = 0$, if we define $\bar{\lambda} = \lambda/\varepsilon$, then $\bar{\lambda}$ solves

$$\bar{f}_1(\bar{\lambda}) = det\Big(\bar{\lambda} I - \bar{\Lambda}_U - \delta\bar{\Sigma}_{UL} - \varepsilon\delta\bar{\Sigma}_{UR}(\varepsilon\bar{\lambda} I - \bar{\Lambda}_L - \varepsilon\bar{\Sigma}_{LR})^{-1}\bar{\Sigma}_{LL}\Big) = 0.$$

If we set $\beta = \varepsilon\delta$, then $\bar{f}_1(\bar{\lambda}, \varepsilon, \delta, \beta)$ is analytic with respect to $\bar{\lambda}$, ε, δ, β. In addition, as $\bar{A}_U$ has n different eigenvalues denoted by $(\bar{A}_U)_i$, $i = 1, 2, ..., n$, we have $\frac{\partial f}{\partial \bar{\lambda}}(\bar{\lambda} = (\bar{A}_U)_i, \varepsilon = 0, \delta = 0, \beta = 0) \neq 0$ for $i = 1, 2, ..., n$. According to the implicit function theorem, $\bar{\lambda}$ is analytic near $(\bar{A}_U)_i$ with respect to ε, δ, β. Therefore, the eigenvalue of Γ can be expanded as $\lambda = \varepsilon[(\bar{A}_U)_i + a_{i0}(\varepsilon, \delta) + \sum_{j=1}^{\infty} a_{ij}(\varepsilon, \delta)\beta^j]$ with coefficients $a_{ij}(\varepsilon, \delta)$ being analytic with respect to ε and δ. And the eigenvalue of Γ' denoted by λ' can thus be obtained by setting $\beta = 0$ as $\lambda' = \varepsilon[(\bar{A}_U)_i + a_{i0}(\varepsilon, \delta)]$. Consequently, we have $\lambda - \lambda' \sim \mathscr{O}(a_{i1}(\varepsilon, \delta)\varepsilon\beta)$ which is of order $\mathscr{O}(\varepsilon^2\delta)$ or higher. For the eigenvalues that solve $f_2(\lambda) = 0$, they do not depend on β as $f_2(\lambda)$ does not include β. Consequently, these n eigenvalues of Γ and Γ' are identical.

For eigenvector X of matrix Γ, if we set $\gamma = \varepsilon^2\delta$, then Γ is a analytic function of ε, δ, and γ. Accordingly, its eigenvector X is analytic with respect to ε, δ, γ (Kato, 1966). In particular, we can expand $X = \sum_{k=0}^{\infty} b_k(\varepsilon, \delta)\gamma^k$ with coefficient $b(\varepsilon, \delta)$ being analytic. And the eigenvector X' of matrix Γ' can thus be obtained by setting $\gamma = 0$ as $X' = b_0(\varepsilon, \delta)$. Therefore, the difference is $X - X' \sim \mathscr{O}(b_1(\varepsilon, \delta)\gamma)$ which is of order $\mathscr{O}(\varepsilon^2\delta)$ or higher. ■