Reviewed Preprint
v1 • March 9, 2026
Not revised

Reviewed Preprint
v2 • September 9, 2026
Revised by authors

For correspondence:
ogino-m@g.ecc.u-tokyo.ac.jp
c-oizumi@g.ecc.u-tokyo.ac.jp

Competing interests: No competing interests declared

Funding: See page 25

Reviewing editor: Peter Latham, University College London, United Kingdom

Designing optimal perturbation inputs for system identification in neuroscience

Mikito Ogino[1], **Daiki Sekizawa**[1], **Jun Kitazono**[2], **Masafumi Oizumi**[1]

[1]Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan • [2]School of Data Science, Yokohama City University, Yokohama, Japan

eLife Assessment

This **important** study introduces a method, based on active control, for determining the functional connectivity between neurons or groups of neurons. While the theory is derived for a linear model with a perfectly observed control signal, the authors present **compelling** evidence that their method will work even when the system is nonlinear and the control signal is not fully observed. Furthermore, they provide a practical recipe for doing so.

https://doi.org/10.7554/eLife.110030.2.sa2

Abstract

Investigating the dynamics of neural networks, which are governed by connectivity between neurons, is a fundamental challenge in neuroscience. Because passive (spontaneous) activity provides only limited information for estimating connectivity, perturbation-based approaches are widely applied in neuroscience, as they can evoke underlying hidden dynamics. However, the characteristics of such perturbations have typically been designed based on empirical or biological intuition. To enable more accurate estimation of connectivity, we propose a data-driven and theoretically grounded framework for optimally designing perturbation inputs, based on formulating the neural model as a control system. The core theoretical insight underlying our approach is that neural signals observed in the passive state lack sufficient latent information, which leads to failures in the system identification. Perturbations reveal these hidden dynamics and lead to improved estimation. Guided by these insights, we derive a theoretical basis for optimizing perturbation inputs that minimize estimation errors in neural system identification. Building upon this, we further explore the relationship of this theory with stimulation patterns commonly used in neuroscience, such as frequency, impulse, and step inputs. We demonstrate the effectiveness of this framework for neuroscience through simulations grounded in experimental paradigms such as neural state classification and optimal control of neural states. Our theoretical analysis, together with multiple simulations, consistently shows that perturbations designed according to our framework achieve substantially more accurate system identification compared to the conventional, intuition-based inputs. This study provides a theoretical foundation for designing perturbation inputs to achieve accurate estimation of neural dynamics. This, in turn, enables reliable discrimination of neural states such as levels of consciousness and pathological conditions, and facilitates precise control of their transitions toward recovery from abnormal states.

Introduction

Much recent interest has focused on how interactions between individual neurons and neuronal populations support cognitive functions and behavior, and on how the disruption of these interactions contributes to various neurological and psychiatric disorders [1, 2, 3, 4, 5]. This

interaction—called neuronal connectivity [6, 7]—is often represented as a network structure or a connectivity matrix. In particular, a commonly studied aspect is functional connectivity, which captures the temporal dependencies between neural activities [8, 9, 10, 11]. The functional connectivity is estimated through recording techniques such as neural spike recording techniques [12, 13], electrocorticography (ECoG) [14, 15], functional magnetic resonance imaging (fMRI) [16, 17] and electroencephalography (EEG) [18, 19]. While various methods have been proposed and used to estimate these connections (see for example [2, 20] for a comprehensive review), one typical method is based on dynamic modeling, among which is a simple but widely used linear auto-regressive model (Fig. 1a) [21, 22, 23, 24, 25]. Connectivity is statistically estimated from the time-series data of neural activity (Fig. 1b) by fitting the model parameters (Fig. 1c). A neural connectivity matrix helps visualize the connections between different brain regions, and provides a detailed map of functional interactions within the brain. By comparing connectivity matrices across individuals or groups, researchers can deepen their understanding of neuronal architectures and communication across various disciplines [8, 26, 27, 28, 29, 30].

A fundamental problem with passive observation is that it fails to reveal hidden dynamics, which leads to an invalid estimate of the corresponding parts of the model. This limitation stems from the attenuation of latent dynamical modes, such as transient or damped components. When neural activity is modeled as a linear dynamical system, as shown in Fig. 1a, the true connectivity matrix governs the generation of time-series signals (Fig. 1b). Passive observation records these spontaneous neural activities within an arbitrarily predefined temporal window and attempts to estimate the connectivity matrix using parameter estimation techniques. However, the resulting matrix $\boldsymbol{A}_{\text{passive}}$ deviates significantly from the true model (Fig. 1c), because some modes quickly decay and vanish from the observable data (see also Fig. 2 for an intuitive explanation). Consequently, relying solely on passive recordings leads to misinterpretations of the neural system—such as overlooking fast transient dynamics or underestimating connectivity strength—which, in turn, may result in inaccurate conclusions in both basic neuroscience and clinical contexts.

To overcome these limitations, the application of perturbations has emerged as a powerful approach for improving the estimation of connectivity in neuroscience [31, 32, 29, 33]. Its effectiveness lies in its ability to actively manipulate the neural system, rather than merely observing it, and to thereby uncover dynamic and causal interactions that are otherwise hidden [32]. Recent studies using stimulation techniques such as optogenetics [34, 35] and transcranial magnetic stimulation (TMS) [36, 29, 37] have demonstrated that these interventions can significantly improve the detection of causal relationships within the neural system. Such approaches have been applied to the study of cognitive processes and states of consciousness [31, 29]. These studies underscore the value of perturbation inputs in driving state transitions in the neural system, thereby enabling more accurate and reliable connectivity estimation.

However, the parameters of perturbation protocols, such as the location, intensity, and shape of stimulation, have often been determined based on anatomical and physiological insights and on empirically established techniques [38, 39, 29, 40, 41]. While these approaches have provided practical utility, they may not optimally exploit the underlying neural dynamics or maximize the informativeness of the perturbation. To overcome these limitations and enhance the reliability of connectivity inference, there is a pressing need to design data-driven and theoretically grounded perturbations.

In this paper, we propose a framework for designing the optimal perturbation input through control theory in neuroscience. We interpret neural dynamics as a control system [42, 43, 44, 45, 46, 47], and treat external perturbations as control inputs to design properties of neural stimulation (Fig. 1d). If the optimal perturbation input can be systematically designed, it becomes possible to steer the neural system toward states that are maximally informative (Fig. 1e), thereby enhancing the accuracy of the inferred connectivity (Fig. 1f). While recent studies have begun to develop algorithmic approaches to active stimulation design for specific experimental platforms [48, 49], a general theoretical framework for why and which perturbation inputs are effective has yet to be established. We first describe how to formulate neural dynamics

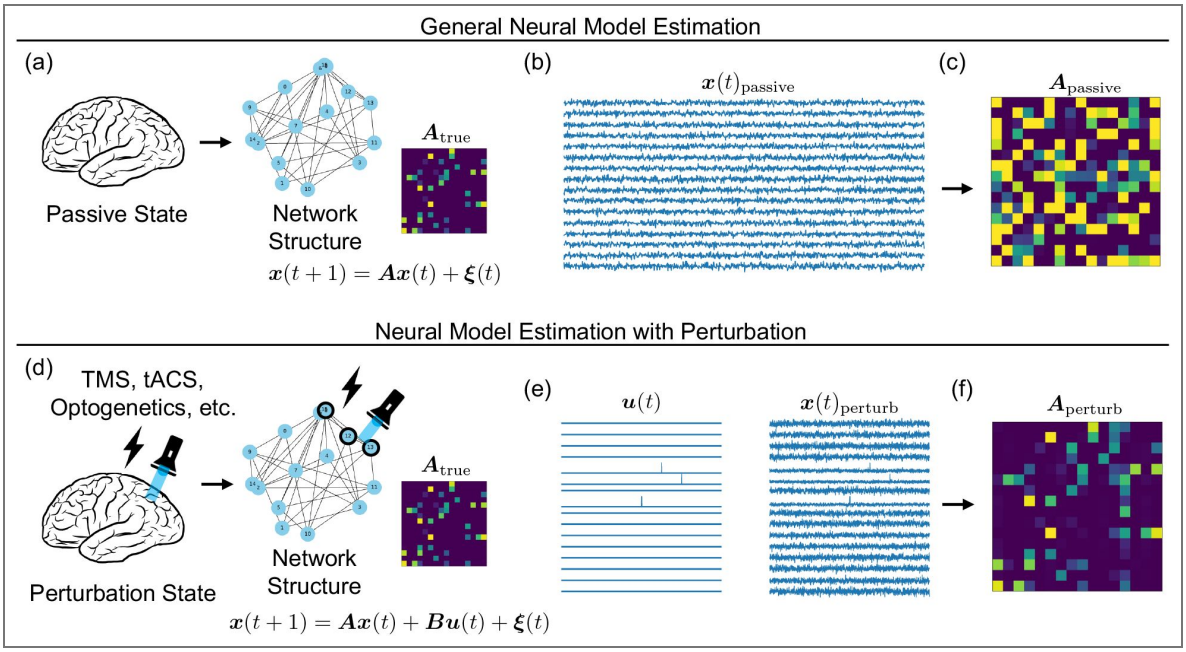


Fig. 1. Passive and perturbation states and the resulting neural dynamics models.

(a) Passive state of the neural network and the corresponding ground-truth model. (b) Recorded neural activities without perturbation. (c) Estimated model obtained without perturbation. (d) Perturbation state of the neural network and corresponding ground-truth model. (e) Recorded neural activity with perturbation. (f) Estimated model obtained with perturbation

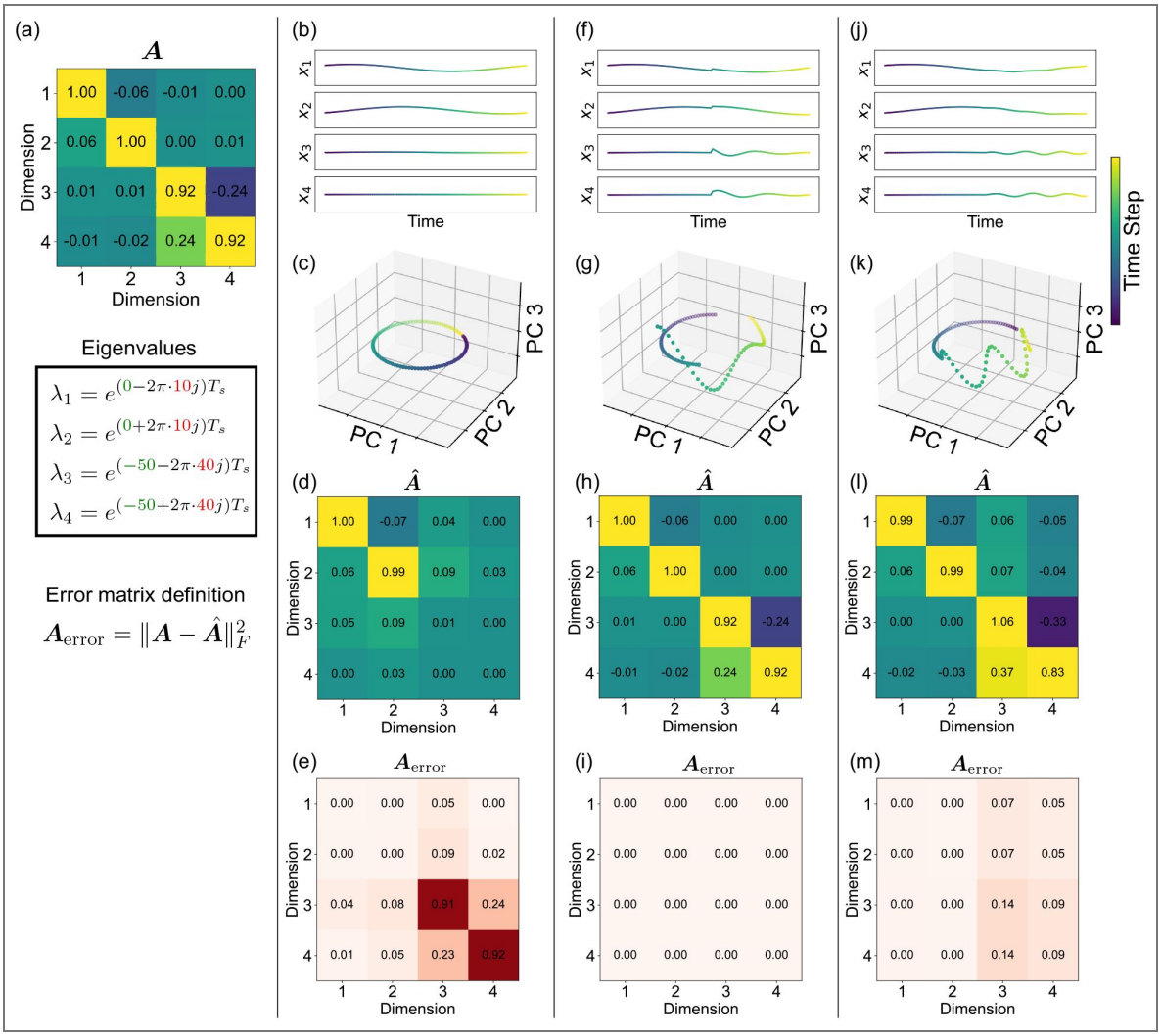


Fig. 2. Visualization of continuous-time system dynamics by perturbations.

(a) System matrix ***A*** representing the dynamics and its eigenvalues. (b) Temporal changes in the state variables of the passive state without perturbation. (c) Time-varying trajectories of each PCA component. The color gradient indicates temporal progression. (d) Estimated connectivity matrix in the passive condition. (e) Error matrix for the passive-state estimation. (f-i) Temporal changes and estimation results in response to an impulse input. (j-m) Temporal changes and estimation results in response to a sinusoidal input.

as a control system and how to estimate the model parameters from observed data. Building upon this formulation, we derive a theoretical basis that enables us to design the optimal perturbation inputs for the neural system identification. We demonstrate the validity and utility of this theoretical basis by exploring its implications for optimizing parameters of common neurostimulation techniques and by applying it to practical examples, including neural state classification [31, 36, 50, 47] and control of neural states [42, 51, 44, 45]. In these demonstrations, we define concrete problems and apply the theory to validate its practical utility.

Our research offers a comprehensive framework for system identification with perturbation in neuroscience, and paves the way for more precise and effective analysis of neural dynamics. By providing clear guidelines on the design and application of perturbations, this framework serves as a practical reference for experimental settings, helping researchers determine the most effective stimulation parameters for their studies.

Background

To evaluate how external perturbations enhance the accuracy of system identification, this section reviews the estimation formulations for linear dynamical systems, comparing the cases with and without perturbation inputs.

System Identification with External Perturbation

This section reviews a well-established framework for system identification of linear dynamical systems with external perturbations. Researchers have modeled brain dynamics as a discrete-time linear dynamics with external inputs and stochastic system noise [52, 42, 53, 54, 43]. We assume that the brain state evolves according to the following linear dynamics:

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{\xi}(t), \tag{1}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ denotes the neural state vector at time t, such as the activity of neurons, populations, or brain regions, and n represents its dimension. The connectivity matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ characterizes intrinsic interactions reflecting functional connectivity. The input matrix $\boldsymbol{B} \in \mathbb{R}^{n\times m}$ specifies how external inputs $\boldsymbol{u}(t) \in \mathbb{R}^m$—including sensory stimuli or interventions such as TMS—affect the system, where m represents the number of perturbation channels. The input $\boldsymbol{u}(t)$ is freely designed and known to the experimenter. Finally, $\boldsymbol{\xi}(t)$ $\mathbb{R}^n$ represents stochastic fluctuations (e.g., synaptic noise or unobserved inputs), modeled as i.i.d. zero-mean Gaussian noise across time with covariance matrix $\boldsymbol{\Sigma}_{\xi} = \mathbb{E}[\boldsymbol{\xi}(t)\boldsymbol{\xi}(t)^{\top}] > 0$. We remark that neural activities measured at the macroscopic level, such as functional magnetic resonance imaging (fMRI) or intracranial electroencephalography (iEEG) have been experimentally and theoretically validated to follow approximately linear dynamical systems in Refs. [55, 56].

Here, we assume the input matrix $\boldsymbol{B}$ is known, and we are only estimating the connectivity matrix $\boldsymbol{A}$ from the time series data of $\boldsymbol{x}$ and $\boldsymbol{u}$. We construct the data matrices as

$$\begin{aligned} \boldsymbol{X} &= [\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(T-1)], \quad \boldsymbol{Y} \\ &= [\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)], \quad \boldsymbol{U} = [\boldsymbol{u}(0), \ldots, \boldsymbol{u}(T-1)], \end{aligned} \tag{2}$$

where T represents the length of the time series data. Using these matrices, the parameter matrix $\boldsymbol{A}$ can be estimated by minimizing the reconstruction error in the sense of ordinary least squares (OLS):

$$\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \|\boldsymbol{Y} - (\boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}\boldsymbol{U})\|_F^2 \tag{3}$$

$$= (\boldsymbol{Y} - \boldsymbol{BU})\boldsymbol{X}^{\dagger}, \tag{4}$$

where $\|\cdot\|_F$ represents the Frobenius norm and the symbol † denotes the pseudoinverse of a matrix. This manuscript derives a theoretical framework based on this formulation, which assumes linear dynamics with the input matrix $\boldsymbol{B}$— known. The relaxation of these assumptions—nonlinear state dynamics and the case of unknown $\boldsymbol{B}$—is addressed in Supplementary Material B.2 and B.3.

System Identification without External Perturbation

To compare the quality of system identification with and without perturbations, we also consider the passive case without external input. This corresponds to setting $\boldsymbol{u}(t) = \boldsymbol{0}$ in the formulation in the perturbed case, yielding the following dynamics:

$$\boldsymbol{x}_{\text{passive}}(t+1) = \boldsymbol{A}\boldsymbol{x}_{\text{passive}}(t) + \boldsymbol{\xi}(t). \tag{5}$$

Here, $\boldsymbol{A}$ and $\boldsymbol{\xi}(t)$ are identical to those defined in Eq. 1. Because this system evolves differently from the perturbed case, we denote its state trajectory as $\boldsymbol{x}_{\text{passive}}(t)$ to explicitly distinguish it from the dynamics under perturbation.

Similarly, the parameter matrix $\boldsymbol{A}$ in the passive condition can be estimated by applying the ordinary least squares (OLS) method:

$$\hat{\boldsymbol{A}}_{\text{passive}} = \boldsymbol{Y}_{\text{passive}}\boldsymbol{X}_{\text{passive}}^{\dagger}, \tag{6}$$

where

$$\begin{aligned}\boldsymbol{X}_{\text{passive}} &= [\boldsymbol{x}_{\text{passive}}(0), \boldsymbol{x}_{\text{passive}}(1), \ldots, \boldsymbol{x}_{\text{passive}}(T-1)], \quad \boldsymbol{Y}_{\text{passive}} \\ &= [\boldsymbol{x}_{\text{passive}}(1), \ldots, \boldsymbol{x}_{\text{passive}}(T)].\end{aligned} \tag{7}$$

The subscript “passive” is used again to explicitly distinguish variables associated with the unperturbed dynamics from those obtained under external perturbations.

Neural Stimulation as Control Inputs

This section describes how commonly used neural stimulation techniques can be related to input signals in control theory. Their adjustable parameters vary depending on how the stimulation inputs are modulated.

Three non-invasive electrical stimulation methods illustrate how stimulation paradigms map onto basic control inputs. Transcranial magnetic stimulation (TMS) induces brief and transient perturbations via electromagnetic pulses [57], which are naturally represented as a sequence of impulse-like inputs, where the timing and intensity of each pulse are the primary controllable parameters. Transcranial direct current stimulation (tDCS) primarily modulates neural activity through approximately constant inputs [58], which can be viewed as a step-like signal whose main controllable parameter is the amplitude of the applied current. Transcranial alternating current stimulation (tACS) delivers oscillatory inputs [59], corresponding to sinusoidal signals characterized by amplitude, frequency, and phase. In control theory, impulse, step, and sinusoidal inputs are the basic components used to characterize system responses and dynamics [60, 61].

The control input framework extends beyond non-invasive techniques to invasive and optogenetic stimulation. Invasive electrical stimulation, including intracranial microstimulation and deep brain stimulation (DBS), enables direct delivery of electrical inputs to neural tissue [62], providing flexible control over amplitude and timing through pulse trains or temporally structured waveforms. Optogenetic stimulation allows genetically targeted activation or inhibition of specific

neurons using light [63], providing fine-grained control over multiple input dimensions, including amplitude (light intensity), temporal pattern, and cell-type specificity. In particular, recent developments enable stimulation at the level of individual neurons with high temporal precision [64, 65], allowing flexible construction of spatiotemporal input patterns.

These stimulation examples demonstrate that the theoretical framework developed in this paper connects to practical experimental settings. While a substantial gap remains between idealized control inputs in theory and experimentally realizable stimulation, the core principles established in the following sections provide a foundation that naturally extends to these practical stimulation paradigms.

Results

We present a theoretical framework for determining the optimal perturbation inputs to enhance neural system identification. We begin by presenting an intuitive example in which the failure of passive estimation is demonstrated, and the benefit of perturbation inputs in re-exciting weakly observable dynamics becomes evident. Following this, we provide a guiding theoretical principle that estimation error of $\boldsymbol{A}$ is reduced by excitation of the state $\boldsymbol{x}$ by perturbation input $\boldsymbol{u}$. We then validate this theoretical foundation by applying it to canonical perturbation signals used in neuroscience—sinusoidal inputs and impulse as typically observed in tACS, tDCS, and TMS—and derive analytical relationships between input parameters and estimation errors. Furthermore, we demonstrate the applicability of the proposed framework to practical problems in neuroscience, such as neural state classification and optimal control for neural state transitions. Finally, we outline a practical framework for designing optimal perturbation inputs to estimate neural dynamics.

Typical Failure of Passive-State Model Estimation

We demonstrate why passive state observation, a common experimental condition in system neuroscience, fails to estimate the system model of neural dynamics. Although passive observation is widely employed due to its convenience and non-invasiveness, this practice has fundamental limitations: it inevitably overlooks causal and dynamical information that vanishes under spontaneous conditions but can be recovered through external perturbations, leading to degraded connectivity estimates. Such limitations lead to misinterpretations of the underlying neural functions when neural dynamics are modeled as a control system.

We estimate the connectivity matrix from simulated time series data generated by the true model. The true model, characterized by a connectivity matrix $\boldsymbol{A}$, is illustrated in Fig. 2a. Neural data are generated based on the connectivity matrix, yielding the time series shown in Fig. 2b. These data are not influenced by external input (passive state), and thus the oscillations of x_3 and x_4 gradually attenuate over time. This attenuation is further confirmed by principal component analysis (PCA), which reveals that the first and second principal components capture the 10Hz dynamics, while the third principal component, associated with the 40Hz mode, contributes negligibly to the total variance (Fig. 2c). Consequently, the matrix estimated by OLS deviates substantially from the true connectivity matrix (Fig. 2d), particularly showing errors in the lower-right components, as highlighted in Fig. 2e. This failure arises because the true matrix $\boldsymbol{A}$ contains two distinct dynamical modes: one is a continuous oscillatory mode at 10Hz, and the other is a damped mode at 40Hz. The damped mode becomes unobservable in the time series once its contribution has attenuated and vanished. Passive recordings capture only superficial aspects of the connectivity and fail to estimate the true underlying neural dynamics.

The limitation of the passive state can be resolved by applying perturbation inputs. Figures 2f–2i show the application of an impulse input $\boldsymbol{u}$ to the system. This perturbation input primarily affects the nodes corresponding to x_3 and x_4, reintroducing variations in the third principal component direction. Estimating matrix $\boldsymbol{A}$ from this perturbed time-series data yields a more accurate estimate than in the passive case. Similarly, an improvement is observed when different types of input—a sinusoidal input for example—are used instead of the impulse input, as shown in Figs. 2j–2m. These results illustrate that appropriate perturbation inputs can effectively re-excite weak

dynamics, thereby enhancing system identification accuracy. This can be theoretically supported by the concept of persistent excitation [66, 67, 68, 69]. According to this theory, a perturbation input is said to be persistently exciting if it causes the system's state to sufficiently explore the state space over time. These insights underscore the importance of incorporating controlled stimulation in experimental design, particularly when accurate system identification is desired. A similar trend under nonlinear dynamics is shown in Supplementary Material B.2.

Intuitive Perturbation Design Informed by Covariance Eigenvalues

In this section, we show how perturbations reduce the estimation error of the system matrix $\boldsymbol{A}$ by exciting the state dynamics, thereby providing a theoretical basis for designing effective perturbation inputs. When the data length T is sufficiently large ($T \to \infty$), the estimation error of $\boldsymbol{A}$ asymptotically converges to the following expression [70]:

$$\mathbb{E}\left[\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_F^2\right] \approx \frac{1}{T}\mathrm{tr}\left(\boldsymbol{\Sigma}_\xi\right)\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right). \tag{8}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \frac{1}{T}(\boldsymbol{X} - \overline{\boldsymbol{X}})(\boldsymbol{X} - \overline{\boldsymbol{X}})$ with $\overline{\boldsymbol{X}} = \frac{1}{T}\sum_{t=0}^{T-1}\boldsymbol{x}(t)$ represents the covariance matrix of the state vector $\boldsymbol{x}$, and tr denotes the trace of a matrix. The detailed derivation of this equation is provided in the Supplementary Material A.1. This equation indicates that the estimation error of $\boldsymbol{A}$ can be reduced either by increasing the measurement duration T or by minimizing the trace of the inverse covariance matrix $\Sigma_{\boldsymbol{X}}^{-1}$, which depends on the perturbation input $\boldsymbol{u}(t)$. Intuitively, this relationship resembles a signal-to-noise ratio, where increasing the covariance of the state dynamics relative to the noise covariance leads to improved estimation accuracy.

The asymptotic expression in Eq. 8 reveals that the estimation error of $\boldsymbol{A}$ depends inversely on the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$. Expressing this relationship explicitly in terms of the eigenvalues yields

$$\mathbb{E}\left[\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_F^2\right] \approx \frac{1}{T}\mathrm{tr}\left(\boldsymbol{\Sigma}_\xi\right)\sum_{i=1}^{n}\frac{1}{\mu_i}. \tag{9}$$

where μ_i denotes the i-th eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$. This relationship indicates that small eigenvalues dominate the estimation error through their inverse contributions. Therefore, the perturbation input should be designed not only to increase the overall variance of the state $\boldsymbol{x}$, but to enlarge the eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ more uniformly across all directions of the state space. In other words, exciting all dynamical modes of the system–rather than amplifying a limited subset–is essential for improving estimation accuracy.

This theoretical insight can be illustrated intuitively in Fig. 3. The figure visualizes ellipsoids constructed from the eigenvalues μ_i and their reciprocals $1/\mu_i$ of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$. The covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ are calculated from the time series of the state vector $\boldsymbol{x}$ obtained under passive and perturbation conditions (Fig. 3a). The length of each axis of the ellipsoid corresponds to the variance of the state along that direction, which is associated with the eigenvalue μ_i. Its reciprocal counterpart $1/\mu_i$ represents the contribution to the estimation error (Fig. 3b). As shown in Fig. 3c, when the neural dynamics is dominated by a single eigenvalue μ_1, the other eigenvalues μ_2 and μ_3 remain small, resulting in an elongated reciprocal ellipsoid (Fig. 3d) and a large estimation error. When perturbation inputs that excite the directions associated with μ_2 and μ_3 are applied, the ellipsoid expands and becomes closer to a sphere (Fig. 3e). As a result, the estimation error decreases in all directions, as illustrated in Fig. 3f. This uniform enlargement of the eigenvalues provides a clear guideline: perturbations should be designed so that the variance becomes large in all directions, rather than being confined to specific modes.

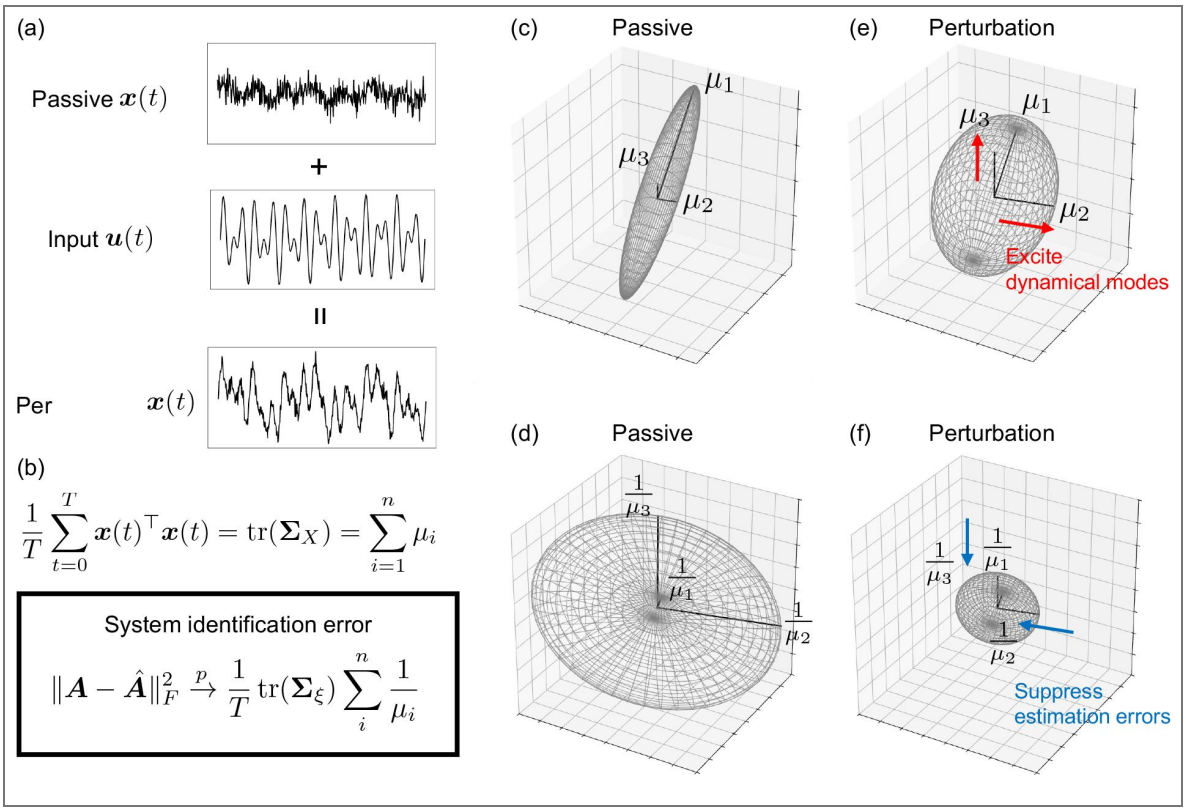


Fig. 3. Effect of perturbation on the eigenvalues of the state covariance matrix.

Ellipsoidal plots show the eigenvalues and reciprocal eigenvalues of the covariance matrix of the state trajectories ***X***, comparing passive dynamics and the case with perturbation. All ellipsoidal plots (c–f) are drawn along the first three eigenvector directions of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$; panels (c, e) scale the axes by the eigenvalues μ_i, while panels (d, f) scale them by the reciprocals $1/\mu_i$. (a) Examples of a passive signal, a perturbation input, and a perturbed signal. (b) Analytical relationship linking the eigenvalues of the state covariance matrix to system identification error. (c) Passive: The covariance matrix is dominated by a single eigenvalue direction. (d) Passive (reciprocal): A single dominant eigenvalue produces a widely spread reciprocal-space ellipsoid (e) Perturbation: Additional dynamical modes are excited, increasing the smaller eigenvalues. (f) Perturbation (reciprocal): Perturbations yield a more uniform reciprocal-space ellipsoid, reducing estimation error.

Theoretical Formulation of Sinusoidal Inputs for an Overall Increase in Eigenvalues

Based on the guideline presented in the previous section, we derive analytical expressions to investigate which frequency of perturbation input will efficiently excite the dynamical modes of a system. In neuroscience, such a perturbation is analogous to tACS. Since neural dynamics possess modal frequencies, there should exist perturbation input frequencies that resonate with these modes. Such resonance leads to an amplification of $\boldsymbol{x}(t)$, which generally results in larger eigenvalues μ_i of the covariance matrix $\boldsymbol{\Sigma_X}$, reflecting increased variability along the corresponding modes. To clarify this relationship, we derive an analytical expression for $\boldsymbol{x}(t)$ to investigate how the input frequency affects its amplitude. The solution $\boldsymbol{x}(t)$ to the model in Eq. 1 can be obtained by iterating the state transition matrix. Specifically,

$$\boldsymbol{x}(t) = \boldsymbol{A}^t \boldsymbol{x}(0) + \sum_{k=0}^{t-1} \boldsymbol{A}^{t-1-k} \boldsymbol{B}\boldsymbol{u}(k) + \sum_{k=0}^{t-1} \boldsymbol{A}^{t-1-k} \boldsymbol{\xi}(k) \tag{10}$$

We assume a cosine input $\boldsymbol{u}(k) = \sum_{l=1}^{L} \cos(\omega_l k)\, \boldsymbol{u}_0$, where ω_l are the angular frequencies. By substituting this into the second term, we can obtain the following equation by performing diagonalization of $\boldsymbol{A}$ and expressing the eigenvalues in polar form.

$$\boldsymbol{x}(t) = \boldsymbol{x}_{\text{passive}}(t) + \boldsymbol{x}_{\text{diff}}(t), \tag{11}$$

$$\boldsymbol{x}_{\text{passive}}(t) = \boldsymbol{A}^t \boldsymbol{x}(0) + \sum_{k=0}^{t-1} \boldsymbol{A}^{t-1-k} \boldsymbol{\xi}(k) \tag{12}$$

$$\boldsymbol{x}_{\text{diff}}(t) = \sum_{l=1}^{L} \sum_{d=1}^{n} \text{Re}\left[\frac{r_d^{t+1} e^{i\theta_d(t-1)} - r_d^t e^{i(\theta_d t - \omega_l)} - r_d e^{i(\omega_l t - \theta_d)} + e^{i\omega_l(t-1)}}{r_d^2 - 2r_d \cos(\theta_d - \omega_l) + 1} \boldsymbol{v}_d \boldsymbol{w}_d^\top \boldsymbol{B}\boldsymbol{u}_0 \right] \tag{13}$$

where each eigenvalue of $\boldsymbol{A}$ is written in polar form as $\lambda_d = r_d e^{i\theta_d}$, with $r_d = |\lambda_d|$ denoting the magnitude and $\theta_d = \arg(\lambda_d)$ denoting the argument. $\boldsymbol{x}_{\text{passive}}$ represents the state $\boldsymbol{x}$ that would be realized if no perturbation were applied, i.e., the state that would evolve according to Eq. 5. The vector $\boldsymbol{v}_d$ denotes the right eigenvector of $\boldsymbol{A}$ associated with λ_d, and $\boldsymbol{w}_d^\top$ denotes the corresponding left eigenvector. The detailed derivation is provided in the Supplementary Material A.2.1.

We focus on $\boldsymbol{x}_{\text{diff}}(t)$ simply because it is the component directly driven by the control input. When the input frequencies ω_l coincide with the mode's angular component θ_d, a resonance-like amplification occurs in $\boldsymbol{x}_{\text{diff}}(t)$—that is, the amplitude of $\boldsymbol{x}(t)$ increases markedly. Increasing the magnitude of $\boldsymbol{x}(t)$ through resonance naturally leads to an increase in the overall variance in $\boldsymbol{\Sigma_X}$. This enhancement directly contributes to increasing all eigenvalues μ_i of the covariance matrix rather than leaving some unexcited. Therefore, the analysis of Eq. 13 provides the necessary guidelines to target and amplify specific dynamical modes, suggesting that oscillatory inputs, such as tACS, should be designed with frequencies that correspond to the true dynamical mode frequencies.

Theoretical Formulation of Impulse and Step Inputs

The eigenvalues of the covariance matrix can also be increased by manipulating the intensity of the perturbation input. In neuroscience, external perturbations such as TMS and tDCS can be modeled as impulse-like or step-like inputs to neural systems. The state vector $\boldsymbol{x}(t)$ for impulse inputs can be written as:

$$\boldsymbol{x}_{\text{passive}}(t) = \boldsymbol{A}^t \boldsymbol{x}(0) + \sum_{k=0}^{t-1} \boldsymbol{A}^{t-1-k} \boldsymbol{\xi}(k), \quad \boldsymbol{x}_{\text{diff}}(t) = \alpha \boldsymbol{A}^{t-1} \boldsymbol{B} \boldsymbol{u}_0. \tag{14}$$

For step input, the state vector is described by,

$$\boldsymbol{x}_{\text{passive}}(t) = \boldsymbol{A}^t \boldsymbol{x}(0) + \sum_{k=0}^{t-1} \boldsymbol{A}^{t-1-k} \boldsymbol{\xi}(k), \quad \boldsymbol{x}_{\text{diff}}(t) = \beta \sum_{k=0}^{t-1} \boldsymbol{A}^{t-1-k} \boldsymbol{B} \boldsymbol{u}_0. \tag{15}$$

where α and β are intensity of the impulse and step inputs. The detailed derivation is provided in the Supplementary Material A.2.2. From these equations, it is clear that the contribution of the perturbation to the state covariance increases proportionally to the squares of the input strengths, α^2 and β^2. However, estimation accuracy is not determined by input intensity alone. Eqs.14 and 15 show that the resulting dynamics $\boldsymbol{x}_{\text{diff}}(t)$ are critically dependent on the term $\boldsymbol{B}\boldsymbol{u}_0$. This term represents how the input vector $\boldsymbol{u}_0$ (which defines the spatial pattern of the stimulation, i.e., which nodes are targeted) interacts with the system's input matrix $\boldsymbol{B}$. Therefore, while increasing intensity (α, β) within experimental constraints is beneficial, the spatial pattern $\boldsymbol{u}_0$ is the key design parameter that determines which dynamical modes are excited. An improperly chosen $\boldsymbol{u}_0$ may excite only the modes already dominant in the passive state, failing to enlarge the smaller eigenvalues that are critical for system identification. The problem of how to design the optimal spatial pattern $\boldsymbol{u}_0$ to most effectively excite the weakly observable dynamics will be addressed in a later section.

Demonstration of Sinusoidal Inputs

In this section, we demonstrate how to design the frequency of oscillatory input by analyzing the eigenvalues of the covariance matrix. Theoretical formulations indicate that tuning the frequency of oscillatory perturbation inputs is more complex compared to that of impulse and step inputs. We generate neural dynamics governed by the connectivity matrix $\boldsymbol{A}$, which is characterized by designated dynamical modes and compare its eigenvalues λ_i with eigenvalues μ_i of the covariance matrix of the perturbed state vectors. The demonstration for impulse and step inputs can be found in the Supplementary Material B.4.

The simulation results show that an oscillatory input with the same frequency as a mode of the system matrix minimized the estimation error when the system had a single mode. Figure 4a shows the system matrix and its eigenvalues, which correspond to a single 10 Hz mode. We generated time series data using this system matrix and estimated the matrix. Figure 4b illustrates the changes in the sum of eigenvalues of the state covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, denoted as $\sum \mu_i$, the sum of reciprocal eigenvalues $1/\sum \mu_i$, and the estimation error defined by the Frobenius norm. The sum of eigenvalues is maximized by a 10 Hz sinusoidal input, while the sum of reciprocal eigenvalues is minimized. Consistent with these results, the estimation error is minimized by the 10 Hz sinusoidal input. As shown in Eq. 9, the estimation error is proportional to the sum of the inverse eigenvalues. This tendency can be explained more clearly by plotting the eigenvalues as ellipsoids. We visualized the eigenvalues using eigenvalue-scaled ellipsoids, as shown in Figs. 4c and 4d. Both the major and minor axes of the ellipsoids in Fig. 4c are extended by the sinusoidal inputs, especially by the 10 Hz input. In contrast, the major and minor axes of the ellipsoids in Fig. 4d are substantially shortened by the 10 Hz sinusoidal input.

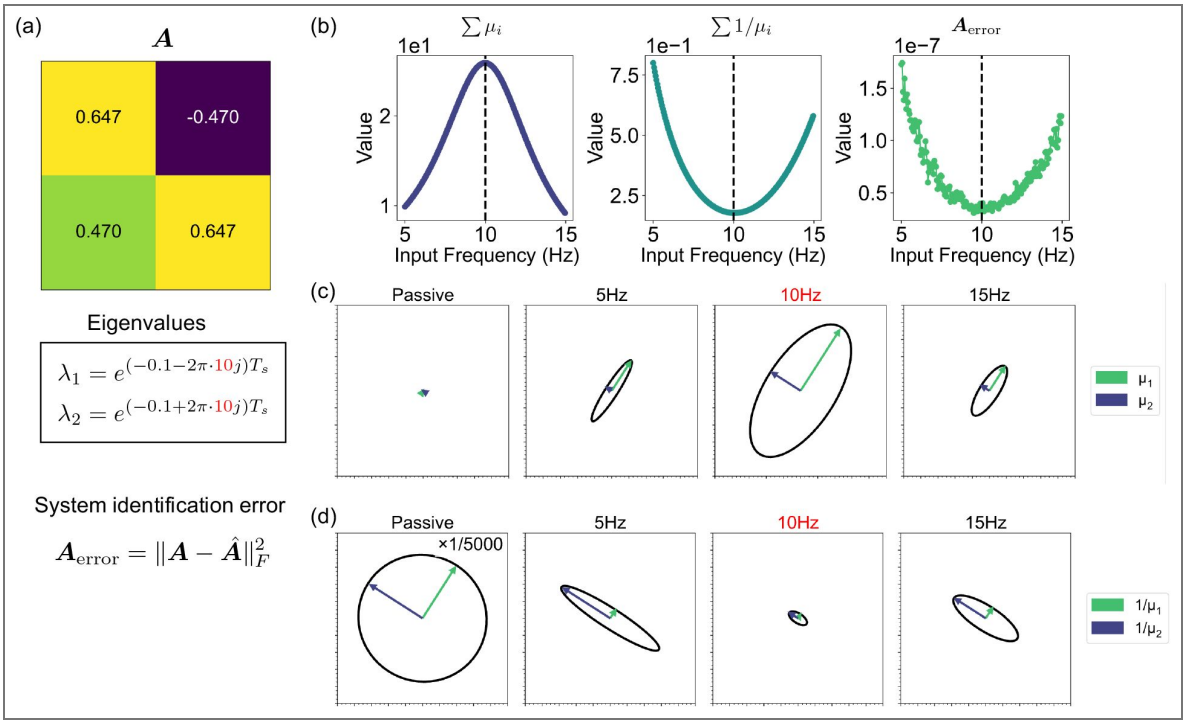


Fig. 4. Error in the system matrix and eigenvalue-scaled ellipsoids.

(a) State transition matrix $\boldsymbol{A}$ and its eigenvalues λ_i. (b) System frequency responses, showing the sum of eigenvalues of the state covariance matrix $\boldsymbol{\Sigma_X}$ (left), the sum of inverse eigenvalues of the state covariance matrix $\boldsymbol{\Sigma_X}$ (middle), and the system matrix error (right). (c) Eigenvalue ellipsoids of $\boldsymbol{\Sigma_X}$ under passive, 5 Hz, 10 Hz, and 15 Hz conditions. (d) Reciprocal-eigenvalue ellipsoids of $\boldsymbol{\Sigma_X}$ under passive, 5 Hz, 10 Hz, and 15 Hz conditions.

When the system exhibits multiple modes, the input frequencies should be designed to reflect all these modes. We demonstrate this using the system matrix shown in Fig. 5a, along with simulated time series and the corresponding matrix estimation. Because the system matrix contains two distinct pairs of eigenvalues (10 Hz and 20 Hz), we constructed the evaluation input $\boldsymbol{u}$ as a combination of two frequencies. The sum of eigenvalue reciprocals is minimized only when the input contained both 10 Hz and 20 Hz components; inputs with a single frequency yield larger values (Fig. 5b). To further evaluate the effect of input design, we compared the two-frequency input with a flat-spectrum input (Fig. 5c). When the total input energy was normalized, the two-frequency input achieves better identification performance than the flat-spectrum input, which is commonly employed in control and system identification studies. This result highlights the importance of tailoring the input spectrum to the system's intrinsic dynamics rather than relying on uniform excitation. The superior performance of the two-frequency combination can be understood through the eigenvalue-scaled ellipsoids in Figs. 5d–g. Figure 5d illustrates ellipsoids determined by the three dominant eigenvalues. The ellipsoid volume is expanded by three types of sinusoidal inputs (10 Hz, 20 Hz, and the combined input). However, the third axis (blue line) is not extended under single-frequency inputs, as shown in Fig. 5e, leading to larger estimation errors. Small eigenvalues disproportionately increase the sum of reciprocal eigenvalues, as seen in Figs. 5f and 5g. In these cases, the reciprocal eigenvalue-scaled ellipsoids are enlarged along the third axis (blue line) when only a single-frequency input is applied. By contrast, the combined input uniformly increases all eigenvalues, including the third, thereby successfully minimizing the reciprocal eigenvalue-scaled ellipsoid volume. These findings indicate that sinusoidal inputs should be designed to match multiple system modes rather than single modes. It should be noted that the true system modes cannot generally be known a priori. They must be identified through iterative experiments and refinement of the estimated system matrix. This practical procedure for optimal perturbation design is described in a later section (see Practical Designing Procedure for Optimal Perturbation Inputs).

Demonstration of Location Tuning

This section illustrates how our theoretical framework enables location-specific tuning of perturbation inputs within neural networks. As established earlier, perturbations should be designed so that the variance becomes large in all directions, rather than being confined to specific modes. Two candidate strategies arise naturally: directly stimulating the nodes associated with heavily damped modes to target those specific modes, or stimulating a hub node whose outgoing connections propagate the perturbation across the entire network.

To examine this relationship in practice, we constructed a network as shown in Figs. 6a and 6b. The network consists of seven nodes, structured into three oscillatory modes (Nodes 1–6) and one hub node (Node 7). It is designed to have three oscillatory modes at 15, 25, and 35 Hz. Fig. 6c shows the damping rates $|\lambda_A|$ of the eigenvalues of $\boldsymbol{A}$: the mode formed by Nodes 1 and 2 (15 Hz) is heavily damped, while those formed by Nodes 3–6 are moderately damped. Node 7 serves as a hub with only outgoing edges. Each node was individually perturbed by an impulse input. Here, an impulse input is defined as a Kronecker delta at $t = 0$ with fixed amplitude $\alpha = 10$, with no external input applied at any subsequent time step.

Stimulating Node 7 (hub node) minimizes the total estimation error of $\boldsymbol{A}$ (Fig. 6h), because its outgoing connections propagate the perturbation to all nodes and re-excite every dynamical mode. Figure 6d–f illustrate how the choice of input node shapes the smallest eigenvalues of $\boldsymbol{\Sigma_X}$. Because the 15 Hz mode (Nodes 1–2) is heavily damped, it decays rapidly and contributes almost no variance to $\boldsymbol{\Sigma_X}$ under passive conditions; as a result, the fifth through seventh eigenvalues μ_5–μ_7 are small and their reciprocals $1/\mu_5$–$1/\mu_7$ are large, producing the elongated ellipsoid seen in Fig. 6d. When Node 1 receives an impulse (Fig. 6e), the eigenvalue associated with the 15 Hz mode grows and the $1/\mu_5$ and $1/\mu_6$ axes shrink; however, the $1/\mu_7$ axis—which correspond to directions that Node 1 alone does not excite— remains large. Stimulating Node 7 (Fig. 6f) re-excites all six oscillatory nodes through its outgoing connections, simultaneously reducing all three axes. Figure 6g confirms this pattern across all candidate nodes: the contribution $1/\mu_i$

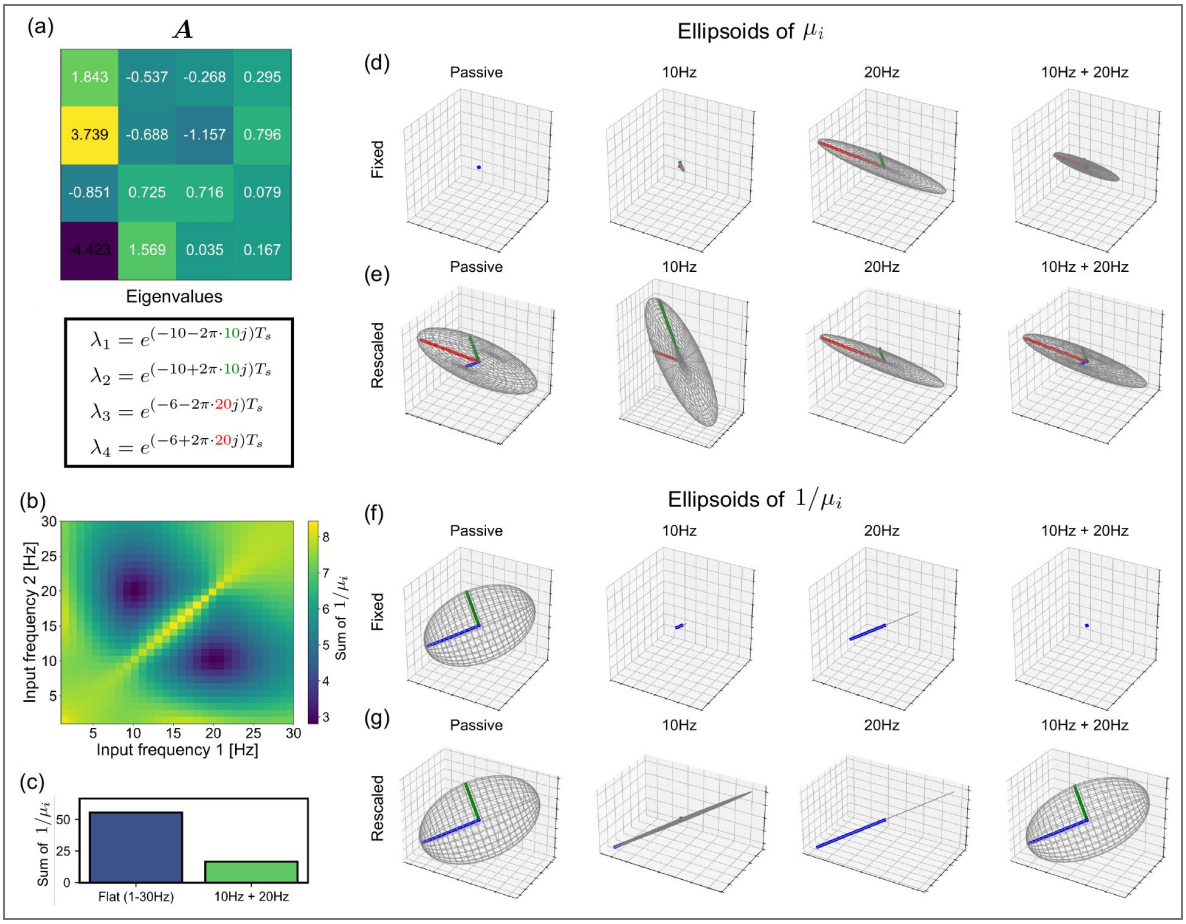


Fig. 5. Simulation of a multi-mode system and eigenvalue-scaled ellipsoids.

(a) System matrix ***A*** and its eigenvalues; the system exhibits two damped oscillatory modes at 10 Hz and 20 Hz. (b) Theoretical estimation error, defined as $\sum (1/\mu_i)$ for two-frequency input combinations; cooler colors indicate smaller estimation errors. (c) Comparison of the reciprocal eigenvalue sum for flat-spectrum versus composite-frequency inputs. (d) Eigenvalue-scaled ellipsoids (μ-scaled) under four input conditions (columns): Passive, 10 Hz, 20 Hz, and 10 Hz + 20 Hz; ellipsoid axes align with the eigenvectors. (e) Relative-scale view of (d). (f) Eigenvalue-scaled ellipsoids (inverse-scaled, $1/\mu_i$) for the same four conditions. (g) Relative-scale view of (f).

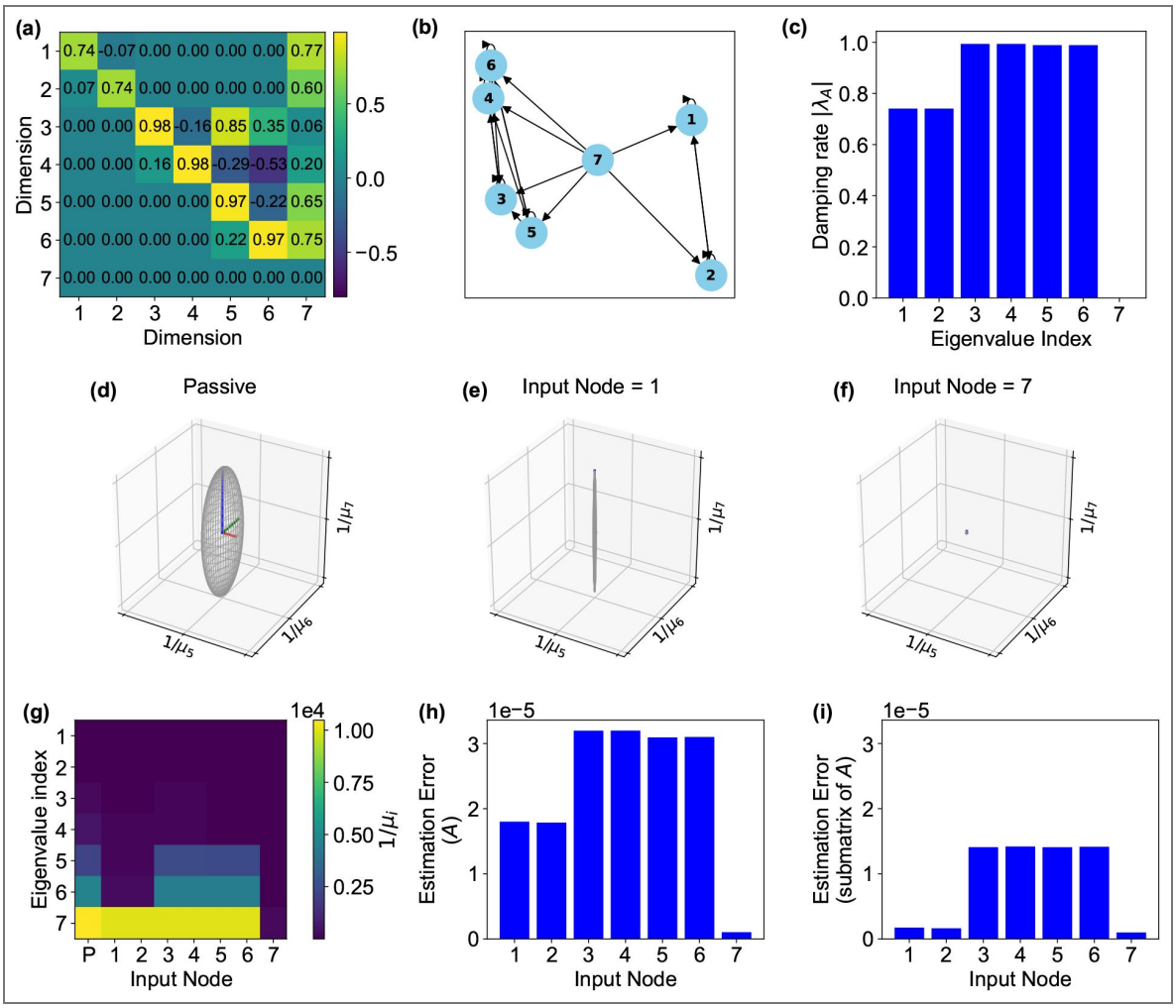


Fig. 6. Optimal stimulation location (a hub node or the heavily-damped subnetwork) minimizes estimation error.

A 7-node network with three oscillatory modes (15, 25, 35 Hz) is used to identify effective perturbation locations. The 15 Hz mode (Nodes 1–2) is heavily damped; Node 7 is a hub with only outgoing edges. (a) The 7 × 7 connectivity matrix $\boldsymbol{A}$. (b) Graph representation. (c) Damping rates $|\lambda_{A|}$ of the eigenvalues of $\boldsymbol{A}$. (d–f) 3D ellipsoids whose axes are proportional to $1/\mu_5$, $1/\mu_6$, $1/\mu_7$ (reciprocals of the three smallest eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$; larger axes indicate greater estimation error): (d) passive, (e) impulse to Node 1, (f) impulse to Node 7. (g) Heatmap of $1/\mu_i$ across input nodes (P = passive); uniformly small values appear only for Node 7. (h) Full-matrix estimation error per input node. (i) Estimation error for the submatrix excluding Node 7; Nodes 1, 2, and 7 achieve comparable errors, confirming that direct stimulation of the damped subnetwork matches hub stimulation.

remains uniformly small only when Node 7 receives the impulse. Figures 6h and 6i show that Node 7 achieves the smallest full-matrix estimation error, followed by Nodes 1 and 2, which directly drive the heavily damped 15 Hz mode.

As this simulation represents only one example of location design, its limitations and the corresponding countermeasures should be stated. In actual experiments, the most effective perturbation location depends on factors such as hub-node connectivity and modal damping rates, and $\boldsymbol{B}$ itself may not always be known a priori. In such cases, approaches such as iterative optimization of $\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right)$ (described in a later section) and joint estimation of $\boldsymbol{A}$ and $\boldsymbol{B}$ (Supplementary Material B.3) provide systematic alternatives. Nevertheless, the results presented here provide an intuitive guideline: stimulation directed at hub nodes or at nodes driving heavily damped modes effectively excites the full set of dynamical modes and minimizes estimation error.

Neural State Classification

We demonstrate the effectiveness of our framework by applying it to a neural state classification problem. Neural state classification is crucial for diagnosing illnesses and assessing brain conditions in many experimental neuroscience settings [31, 36, 50, 47]. Some of these studies have already applied arbitrary perturbation inputs for classification, but not optimal perturbation inputs. By simulating such real-world applications, we demonstrate how well optimal perturbation design can contribute to neuroscience research.

We designed a neural network with clearly distinct task conditions and considered a simulation setting in which these conditions are classified using signals of a fixed duration. These distinct task conditions consist of five types, each defined by a unique linear dynamical system characterized by differing eigenvalue spectra and connectivity topologies of matrix $\boldsymbol{A}$ (Fig. 7a). These task conditions are intended to mimic different cognitive or behavioral contexts. For example, in a typical motor task experiment, such conditions could correspond to motor execution or imagery involving the left or right hand, or resting state [50, 47]. The neural signals were simulated under five different task conditions and two stimulation conditions: passive observation and external perturbation. Perturbation was applied as impulse-type inputs, such as TMS. The stimulus location was determined for each task condition by applying an impulse to each node and selecting the one that minimized $\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right)$. The resulting time-series data are shown in Fig. 7b. Using this data, we estimated the underlying dynamical system via a control-based identification approach presented in Eq. 4, which corresponds to an estimation of functional connectivity.

The simulation demonstrates that classification performance under perturbation is superior to that under the passive condition, both in visual inspection and in quantitative evaluation. As revealed by linear discriminant analysis (LDA) projection, task clusters in the passive condition exhibit substantial overlap, whereas perturbation leads to clear separation among the tasks (Fig. 7c). Classification with LDA and 10-fold cross-validation shows the average accuracy is substantially higher in the perturbation condition (78.20%) compared to the passive condition (24.20%), as shown in Fig. 7d. ROC curve analysis further confirms that perturbation substantially improves discriminability across all tasks and outperforms the random perturbation condition (Fig. 7e). These results can be explained by Eq. 8: the designed perturbation inputs increase the eigenvalues of the covariance matrix, including even the smallest ones, which in turn leads to a decrease in the estimation error of $\boldsymbol{A}$.

To obtain estimates under the passive condition that are comparable to those derived under perturbation, it is necessary to experimentally observe extensive time-series data. Figure 7f illustrates the LDA projection and classification accuracy for different time-series lengths. The leftmost LDA plot (T = 20) corresponds to the passive condition shown in Fig. 7c, indicating that the estimation performance in the passive condition becomes comparable to that in the perturbation condition only when the time window reaches approximately T = 200, a 10-fold increase compared to T = 20. While the absolute duration depends on the interpretation of the time unit, such time windows may not be prohibitive in some experimental settings. Nevertheless,

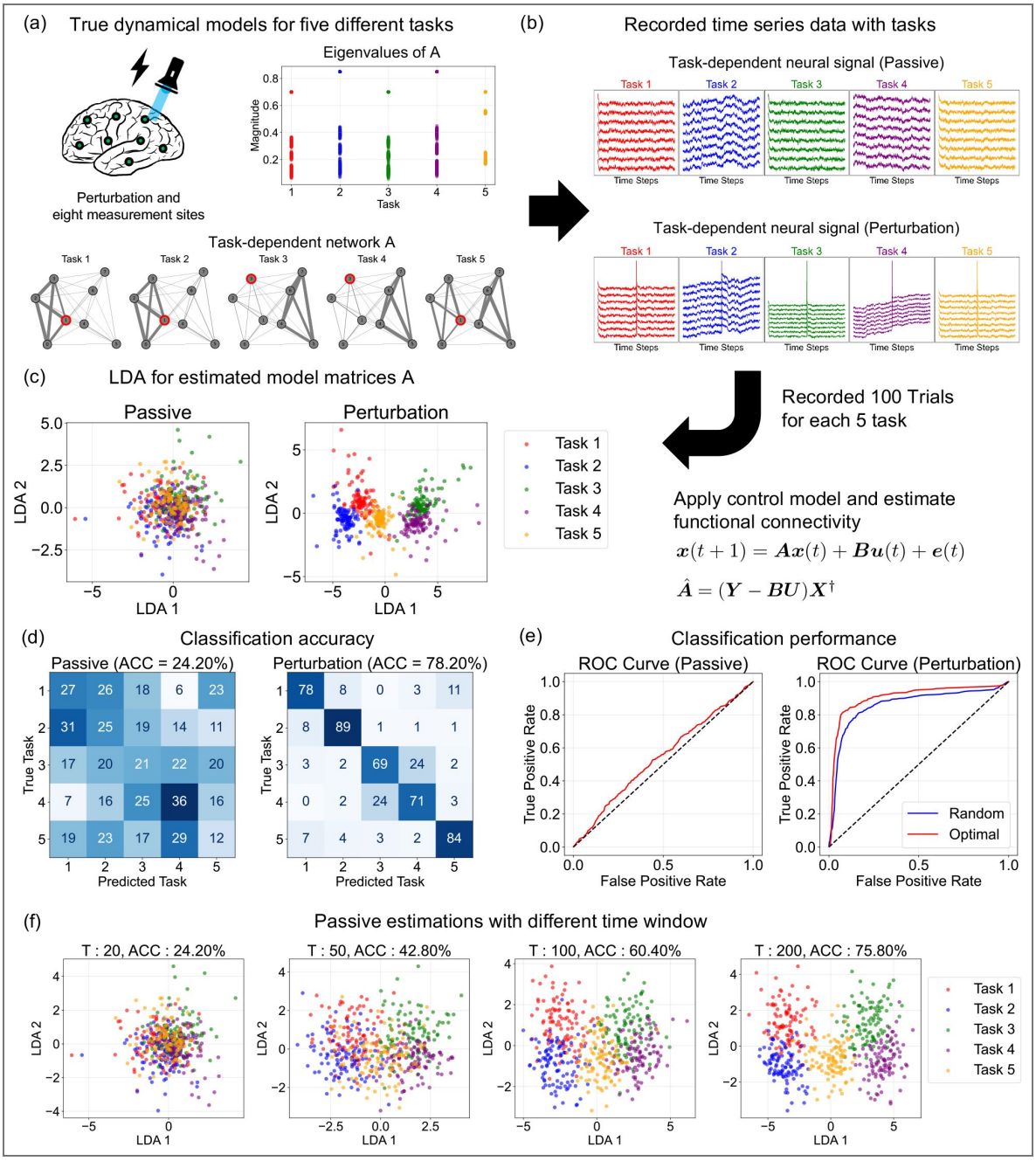


Fig. 7. Neural state classification enhanced by perturbation-aided model estimation.

(a) Ground-truth dynamics models defining five distinct task conditions, each with unique eigenvalue distributions and network structures of matrix ***A***. The optimally excited node is highlighted in red. (b) Simulated neural states (activity) induced by the five task conditions, shown separately for the passive (top) and perturbed (bottom) regimes. (c) Estimated model matrices ***A*** visualized using LDA. (d) Confusion matrices showing classification accuracy for the passive condition (left) and the optimally designed perturbation condition (right). (e) ROC curves for the passive condition (left) and a comparison of random and optimal perturbation conditions (right). (f) LDA plots for passive condition with increasing time window length, showing clustering and classification accuracy as the time window length *T* increases from 20 to 200.

our results consistently show that passive observation requires substantially longer recordings to achieve comparable performance, highlighting the efficiency of the perturbation-based approach when the available data length is limited.

Neural State Transitions via Optimal Control

This section demonstrates the practical utility of our framework by applying it to a neural state control problem [42, 51]. Here, we estimate the system parameters from both passive and perturbation states. Then, we determine optimal control inputs which control the neural states to desired targets, and associated control costs. By using perturbation inputs, $\boldsymbol{A}$ is accurately estimated, which in turn allows precise determination of the optimal inputs and the corresponding controlled state transitions (Fig. 8a). Moreover, use of the perturbation approach to derive the model allows the controllability Gramian [42, 47], as well as the estimation of control costs for each network area, to be more reliably assessed (Fig. 8b and 8c). This simulation underscores the importance of accurate system identification for achieving optimal control and reliable estimation of neural functions. Throughout this section, we use the term *perturbation input* to refer to the input used for system identification and *control input* to refer to the input used for state transitions.

We designed a network as shown in Fig. 9a to clearly show the differences in system identification between passive and perturbations states. The network consists of two disconnected groups. Nodes 1 and 2, as well as nodes 3, 4, and 5, are each connected separately, with no connections between these groups, as shown in matrix $\boldsymbol{A}$. We assumed a known input matrix $\boldsymbol{B}$ in which nodes 3 and 4 cannot be directly controlled. In this scenario, the optimal control strategy to manipulate nodes 3 and 4 is to apply inputs to node 5. However, if the system identification is inaccurate, there is the possibility that control will be attempted through nodes 1 and 2, which are disconnected from nodes 3 and 4. These matrices were estimated from the time series signals generated by simulation. We generated passive and perturbation states to estimate the parameter matrices. The perturbation condition uses an impulse input with sufficient intensity ($\alpha = 10^3$). By using the estimated system model, we determined optimal control inputs, then evaluated the state transitions and associated costs. The controlled transition test was run with $T = 50$. The control objective was to set nodes 3 and 4 to 25 while keeping all other nodes at 0 without any movement.

The simulation results show that perturbation-based estimation enables the precise control of neural states. If the estimation of $\boldsymbol{A}$ is inaccurate, the state transitions under optimal control cannot be executed correctly (Fig. 9b). Nodes 3 and 4 fail to reach the target, and unnecessary activations occur in nodes 1 and 2. Furthermore, unnecessary control inputs are required in nodes 1 and 2. On the other hand, when the model is correctly estimated using a high-strength impulse, the resulting control inputs enable accurate state transitions (Fig. 9c). In this case, the strength of the impulse is correctly concentrated on node 5. As the error in matrix $\boldsymbol{A}$ increases, the estimation errors of the controllability Gramian $\boldsymbol{W}$ lead to inaccuracies in estimating the control cost (Fig. 9d). Here, we adopt average controllability as an example measure of control cost [42, 46]. The key point here is that, since the controllability Gramian involves multiple products of $\boldsymbol{A}$ (Eq. 21), even small errors of $\boldsymbol{A}$ can result in significant discrepancies. Therefore, when estimating optimal control input for neural transitions or control costs, it is more appropriate to identify those model parameters with at least arbitrary perturbations, and ideally with designed perturbations.

Practical Designing Procedure for Optimal Perturbation Inputs

This section presents a practical framework for designing optimal perturbation inputs to estimate neural dynamics. The practical framework progressively refines the input signal through repeated cycles of model identification and perturbation input design. By leveraging this alternating scheme, each iteration utilizes the current model estimate to design a more informative input for the next round of identification. We also focus on perturbations characterized by a broad and uniform frequency distribution, a form commonly adopted in control engineering and

Fig. 8. Effects of system identification on network control theory.

(a) State transitions using models identified from the passive and perturbed conditions. (b) Controllability Gramians with two models. (c) Control costs computed from the estimated controllability Gramians.

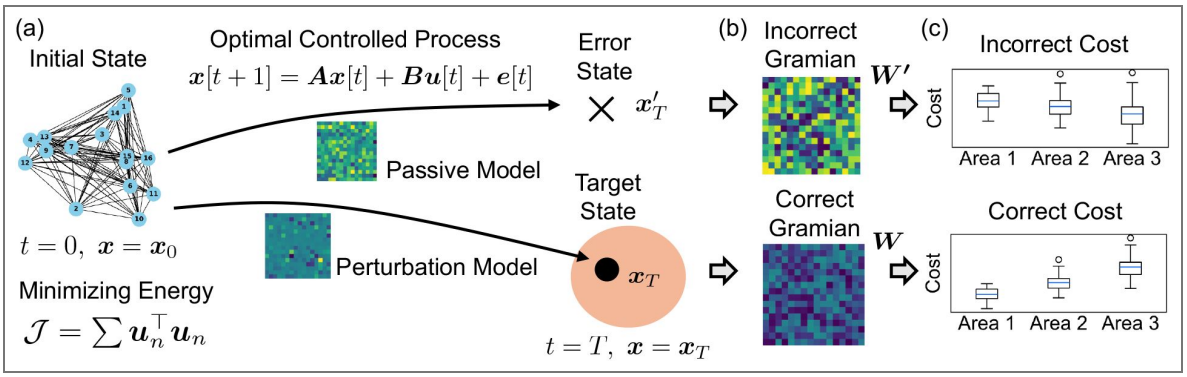
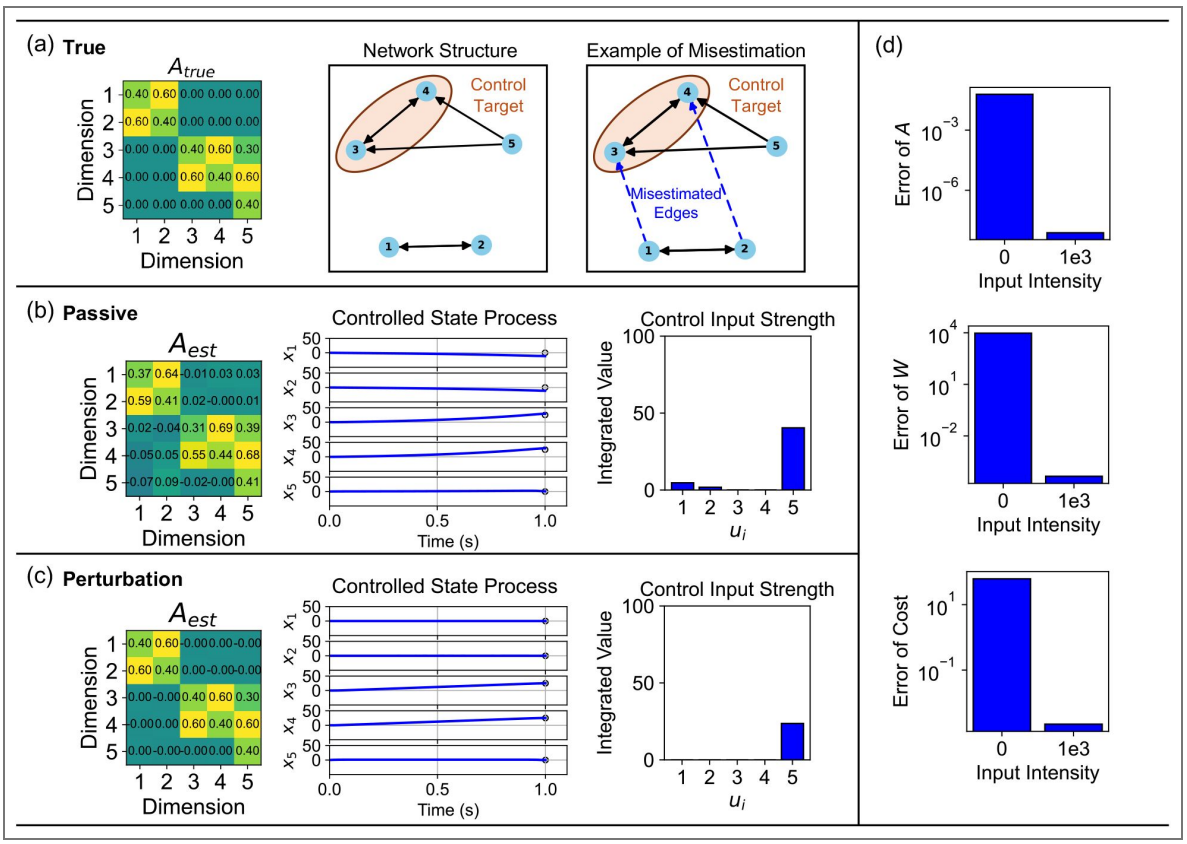


Fig. 9. Accurate system identification via perturbation enables precise optimal control.

(a) True matrix ***A*** and its network structure. Nodes 1, 2, and 5 are controllable nodes. Nodes 3 and 4 are targets controlled by the determined control input. When the connectivity matrix is misestimated, the structure has edges between nodes 1 and 2, and nodes 3 and 4. (b) Matrix estimation failure under passive condition and controlled state transitions. The control objectives are plotted as white circles. (c) Successful matrix estimations and controlled state transitions with sufficiently strong impulse inputs ($\alpha = 10^3$). (d) Errors in the estimated matrices ***A*** and ***W*** and the control cost (average controllability).



conceptually analogous to transcranial random noise stimulation (tRNS) in neuroscience [71, 72]. As shown in Figs. 5 and B.4, the designed composite-wave input allows for more accurate model estimation than the uniform-frequency input when their total input energies are equal.

Iterative refinement of both the perturbation design and the estimation process progressively improves the accuracy of $\boldsymbol{A}$. The time-series data is collected from 32 points, where the matrix $\boldsymbol{A}$ (Fig. 10a) is designed to have 16 oscillatory modes (i.e., 16 complex-conjugate eigenvalue pairs, yielding 32 eigenvalues in total). In this simulation, one stimulation session is applied to a single node at each design step. At each iteration, the stimulation is designed as a composite-frequency sinusoidal input encompassing all modes of the estimated $\hat{\boldsymbol{A}}$, and the target node of the perturbation is determined through numerical optimization that minimizes $\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right)$. Table B.1 summarizes the simulation conditions including measurement duration and number of trials used in each simulation. The stimulation duration was set equal to the measurement duration.

The following procedure outlines a practical approach (Fig. 10c) for this simulation, aiming to precisely estimate the matrix by designing an optimal perturbation input (assuming frequency stimulation, such as tACS).

1. Record spontaneous activity as the passive state and estimate $\boldsymbol{A}_{\mathrm{passive}}$.
2. Based on the estimated matrix, the optimal frequency and target node of the perturbation input are determined through numerical optimization following Eq. 9.
3. The designed perturbation $\boldsymbol{u}_{\mathrm{design1}}$ is applied to the system, and a more precise matrix $\boldsymbol{A}_{\mathrm{design1}}$ is identified.
4. The obtained matrix $\boldsymbol{A}_{\mathrm{design1}}$ is again used to design a new perturbation input $\boldsymbol{u}_{\mathrm{design2}}$ and estimate a more accurate matrix $\boldsymbol{A}_{\mathrm{design2}}$. This iterative process can be continued until convergence criteria, such as changes in $\boldsymbol{A}$, are met.

The proposed framework improved estimation accuracy iteration by iteration. As shown in Fig. 10b, the estimation error converges toward the theoretical optimum (dashed line) with each design iteration. Here, the theoretical optimum is obtained from a simulation in which the input is constructed as a composite sinusoid covering all dynamical modes of the true matrix $\boldsymbol{A}$, and the stimulation location is numerically selected to minimize $\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right)$. The error from the second iteration (2nd design) is smaller than that from the first (1st design), which in turn is substantially lower than the initial estimate. This convergence holds both when starting from the passive condition and when starting from a flat-spectrum input (a practical approach analogous to tRNS). While the flat-spectrum input (blue line, Iteration 1) provided a better starting point than passive data, the designed input outperformed it by the second iteration. Furthermore, the converged estimation error was lower than that of a random stimulation baseline—in which a single randomly chosen node was stimulated using a flat-spectrum (all-frequency) input throughout all iterations (green line)—whose estimation error remained approximately constant across iterations. This contrast directly demonstrates that the benefit of the proposed framework arises specifically from model-based optimization of both the spatial (node selection) and spectral (frequency composition) stimulation parameters. These results should be interpreted as a case-specific illustration rather than a universal gain, as the magnitude of improvement depends on factors such as network structure, recording duration, noise level, and stimulation design. This simulation demonstrates that our perturbation design framework enables the step-by-step refinement of system identification even without prior knowledge of the system.

Discussion

This study addressed the challenge of designing optimal perturbations for effectively identifying neural system dynamics. We introduced a framework for estimating the optimal perturbation input for identification of neural systems. Our findings demonstrated that incorporating perturbation inputs, including TMS, tDCS, and tACS, significantly improves identification accuracy. Specifically, alignment of their parameters with the intrinsic frequencies of matrix $\boldsymbol{A}$, high input intensity, and targeted inputs directed at nodes that efficiently excite heavily damped dynamical modes (either nodes directly associated with those modes or hub nodes whose outgoing

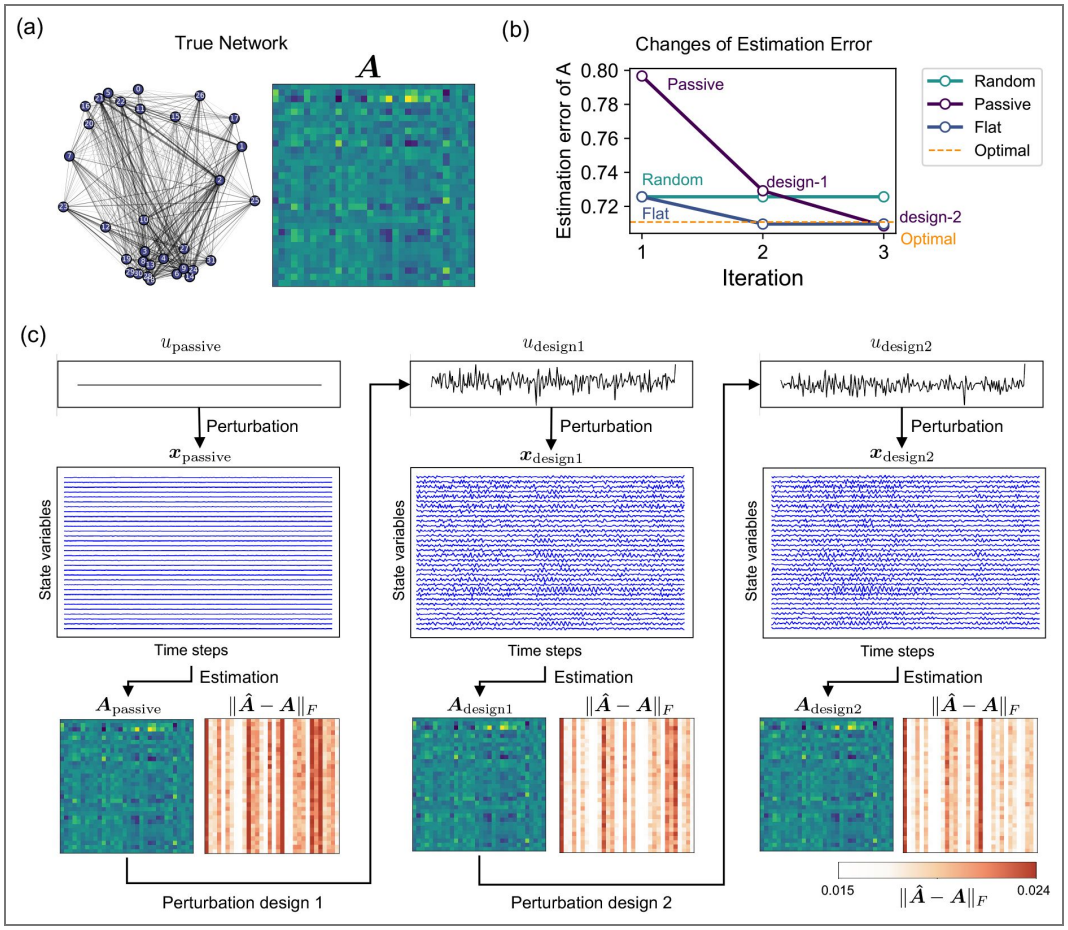


Fig. 10. Iterative perturbation design converges to the theoretical optimum in a 32-channel neural network simulation.

(a) True matrix $\boldsymbol{A}$ and its network structure, which was designed to include both oscillatory and attenuating components. (b) Comparison of estimation errors. The purple line indicates the iterative design process initialized from the passive-state estimate (Iteration 1). The blue line shows the iterative design starting from a flat-spectrum input (Iteration 1). Both processes converge towards the theoretically optimal error (dashed line) as the design is refined (design-1, design-2). The third line (labeled "Random") represents a baseline in which a single randomly chosen node is stimulated with a flat-spectrum input throughout all iterations. (c) Schematic of the iterative procedure. An initial estimated system matrix ($\boldsymbol{A}_{\mathrm{passive}}$) is used to design the first perturbation input ($\boldsymbol{u}_{\mathrm{design1}}$), which yields a better estimate ($\boldsymbol{A}_{\mathrm{design1}}$), and the cycle repeats.

connections propagate the perturbation throughout the network) enhances system identification. Furthermore, we outlined an approach for designing the optimal perturbation input by iterating system identifications with perturbations, and demonstrated that the approach progressively decreases parameter estimation error. Beyond the parameter identification of neural systems, our findings provide insights into how these estimated parameters influence optimal control theory in neuroscience. These results emphasize the potential of perturbation input design to advance understanding of neural dynamics.

The assumption of linearity and first-order autoregressiveness in neural dynamics warrants discussion. Biological signals often exhibit complex and nonlinear interactions that cannot be fully captured by linear models. Nevertheless, prior studies have demonstrated that nonlinear behaviors can frequently be approximated using linear models [9, 73, 42]. Linear models are particularly effective for providing accurate approximations of nonlinear systems within a specific operating range [74]. Therefore, establishing theoretical foundations based on linear assumptions can contribute to the development of theories for nonlinear systems or be effectively utilized in their advancement. To extend these theories into the nonlinear domain, it would be necessary to incorporate advanced frameworks such as the Koopman operator [75, 76] and system identification methods leveraging machine learning techniques [77, 78]. On the other hand, discussions regarding higher-order VAR models are relatively straightforward due to the simplicity of the estimation method. Previous studies involving actual functional data have employed VAR models of an order greater than one [79, 80, 81]. The framework proposed in this paper can be naturally and meaningfully extended to higher-order VAR models, broadening its applicability to more complex temporal dependencies.

Safety and experimental constraints play a pivotal role in the practical implementation of this framework, influencing both the scope and design of potential applications. In biological systems, stimulation parameters, such as amplitude, frequency, and location, must adhere to stringent safety thresholds to avoid adverse effects, such as tissue damage or unintended physiological responses [82, 17]. For stimulus intensity, our proposed framework suggested that a stronger stimulus improves system identification; however, an excessively strong stimulus poses safety risks [83]. With respect to sinusoidal input, stimuli are typically applied within ranges corresponding to neural activity. The optimal input frequencies derived from the proposed theory are expected to align with these ranges, suggesting that the proposed framework can be implemented without major complications [83, 84]. Furthermore, the selection of stimulation sites is often dictated by experimental accessibility or ethical considerations [85, 86]. Based on these constraints, it is necessary to determine the most appropriate stimulation parameters to achieve optimal system identification.

Two directions warrant further investigation: extending the framework to partial observability, and validating it through stimulation experiments. In experimental neuroscience, recordings are often limited to a subset of neural populations, resulting in partial observability. A growing body of work has leveraged delay-embedding techniques, represented by Takens' embedding theorem [87], to reconstruct hidden dynamics from partial observations [88, 89, 90, 91]. Applying such techniques enables the estimation of the full connectivity matrix, thereby extending our framework to settings with partial observability. The second direction concerns experimental validation. Validating a theoretical framework through experimental design is an essential in bridging the gap between theory and practice. Recent research involving some of the present authors has demonstrated that TMS can facilitate the discrimination of neural states [47]. Future experiments can build on this finding by incorporating passive, arbitrary, and designed optimal stimuli to estimate the connectivity matrix, and then comparing the ability of these matrices to distinguish different neural states. If these proposed evaluations demonstrate the practical effectiveness of our theory, it could then be applied to actual experiments. As shown in Fig. 10, a preliminary connectivity matrix is obtained through preliminary experiments to design optimal perturbations. These perturbations would then be applied in the main experiments, improving the

estimation of connectivity matrices and neural dynamics. Experiments using our approach will enable the more precise and comprehensive analysis of brain and neural function, and in turn facilitate valuable new insights into human cognition and behavior.

Methods

State Vector Simulation and Model Estimation

To investigate appropriate perturbation inputs for the identification of neural systems, we constructed neural dynamics using matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ to generate the temporal evolution of state variables. From the generated dynamics, we estimated the model matrix $\boldsymbol{A}$, while assuming that the model matrix $\boldsymbol{B}$ was known.

Given an initial condition $\boldsymbol{x}(0)$, the neural dynamics were generated according to the true matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and the governing equation for $\boldsymbol{x}(t)$ (Eq. 1), with a predefined noise time series $\boldsymbol{\xi}(t)$ added at each time step. For the discrete-time simulations, the sampling rate was set to an appropriate value for each simulation. The model matrix $\boldsymbol{A}$ was then estimated from the generated time-series states $\boldsymbol{x}(t)$ and applied perturbation inputs $\boldsymbol{u}(t)$ using the OLS.

Both the generation of neural dynamics and the model estimation were repeated for a predefined number of trials, with variations in initial conditions and noise sequences. The estimation errors of the matrices were quantified using the Frobenius norm. Details of the simulation parameters are provided in the Supplementary Material B.1.

Eigenvector Alignment for eigenvalue-scaled ellipsoids

To examine the relationship between the eigenvalues μ_i of the state covariance matrix $\boldsymbol{\Sigma_X}$ and the input frequencies, we plotted an eigenvalue-scaled ellipsoid, thereby providing an intuitive representation of optimal input design. However, the eigenvectors of $\boldsymbol{\Sigma_X}$ are not fixed; they may vary depending on the applied input, which complicates direct comparison across conditions. To enable consistent comparison of eigenvalues across input conditions, we reordered eigenvalues according to the similarity of their associated eigenvectors to those obtained in the passive condition, which served as the reference basis. For each reference eigenvector $\boldsymbol{t}_i$, we identified the most similar eigenvector $\boldsymbol{u}_j$ from the input condition by maximizing the absolute inner product

$$j^* = \arg\max_j \left| \boldsymbol{t}_i^\top \boldsymbol{u}_j \right| . \tag{16}$$

The eigenvalue μ_{j^*} corresponding to $\boldsymbol{u}_{j^*}$ was then assigned to the (i)-th position of the reordered list. Each $\boldsymbol{u}_j$ was used only once, ensuring a one-to-one correspondence between reference and input eigenvectors. This alignment resolves the permutation and sign ambiguities inherent in eigendecomposition, and allows eigenvalues to be compared across conditions along a common axis.

Optimal Control and Network Controllability

Accurately estimating dynamical models under perturbation not only facilitates precise inference of the neural dynamics but also contributes to the accurate estimation of control inputs for neural state transitions. Recently, the control theory which utilizes the controllability Gramian and control costs is applied for neuroscience, and provides insights into the efficiency and feasibility of inducing specific neural state transitions [42, 51] with some limitations [92, 93]. For clarity, we refer to the input for system identification as the *perturbation input* and the input for state control in the optimal control theory as the *control input*.

The optimal control input is derived under the condition of assuming the linear system is controllable. We consider the sequence of control inputs that minimizes the following cost function (input energy minimization) while driving the state from the initial state $\boldsymbol{x}_0$ to the terminal state $\boldsymbol{x}_T$:

$$\mathscr{M}\left(\{\boldsymbol{u}(t)\}_{t=0}^{T-1}\right)=\sum_{t=0}^{T-1}\boldsymbol{u}(t)^{\top}\boldsymbol{u}(t), \tag{17}$$

under

$$\boldsymbol{x}(0)=\boldsymbol{x}_0,\quad \boldsymbol{x}(T)=\boldsymbol{x}_T. \tag{18}$$

Here, the initial and terminal state conditions represent constraints on the optimization problem. This problem is expressed as follows:

$$\boldsymbol{u}(t)^{*}=\arg\min_{\boldsymbol{u}}\mathscr{M}\left(\{\boldsymbol{u}(t)\}_{t=0}^{T-1}\right). \tag{19}$$

The constrained optimization problem can be solved using Lagrange multipliers. The optimal control input is given by

$$\boldsymbol{u}(t)^{*}=\boldsymbol{B}^{\top}\left(\boldsymbol{A}^{\top}\right)^{T-t-1}\boldsymbol{W}^{-1}\left(\boldsymbol{x}(T)-\boldsymbol{A}^{T}\boldsymbol{x}(0)\right), \tag{20}$$

where $\boldsymbol{W}$ is the controllability Gramian, defined as:

$$\boldsymbol{W}=\sum_{t=0}^{T-1}\boldsymbol{A}^{t}\boldsymbol{B}\boldsymbol{B}^{\top}\left(\boldsymbol{A}^{\top}\right)^{t}. \tag{21}$$

The controllability Gramian plays a pivotal role in determining the optimal control and control input cost. It has been used to identify the functional roles of individual brain regions [42, 47].

Average Controllability for Control Cost

While the optimal control problem (Eq. 19) calculates the specific input energy for a given state transition, a more general, state-independent metric is often used to characterize the system's overall controllability. This metric, average controllability, is a state-independent metric used to quantify the system's overall ease of control [42, 46]. It is defined as the trace of the controllability Gramian:

$$\text{Average Controllability}\ =\operatorname{tr}(\boldsymbol{W})=\sum_{i=1}^{N}\lambda_i(\boldsymbol{W}) \tag{22}$$

A larger trace indicates that the system is, on average, more controllable (i.e., can be moved to various states with less input energy). This metric is therefore used as a measure of control efficiency.

Data availability

The current manuscript is a computational study, so no biological data have been generated. The code for the simulations and analyses presented in this paper is openly accessible at *https://github.com/mikito-ogino/NeuroPerturbID*.

Acknowledgements

We are grateful to Yumi Shikauchi, Shunsuke Kamiya, and Daiki Kiyooka for their insightful feedback and valuable discussions, which significantly contributed to the development of this research. This work was supported by JST Moonshot R&D Grant Number JPMJMS2012, JSPS KAKENHI Grant Number 24K20462 and JSPS KAKENHI Grant Number 23KJ0799.

Additional files

Supplementary Material

Additional information

Funding

Funder	Grant reference number	Author
MEXT | Japan Science and Technology Agency (JST)	https://doi.org/10.52926/jpmjms2012	Masafumi Oizumi
Japan Society for the Promotion of Science (JSPS)	24K20462	Mikito Ogino
Japan Society for the Promotion of Science (JSPS)	23KJ0799	Daiki Sekizawa

Author ORCID iDs

Mikito Ogino: https://orcid.org/0000-0003-1089-4469

References

[1] **Waites AB,** Briellmann RS, Saling MM, Abbott DF, Jackson GD (2006) Functional connectivity networks are disrupted in left temporal lobe epilepsy. *Ann. Neurol* **59**:335-343 https://doi.org/10.1002/ana.20733 | PubMed

[2] **Friston KJ** (2011) Functional and effective connectivity: a review. *Brain Connect* **1**:13-36 https://doi.org/10.1089/brain.2011.0008 | PubMed

[3] **Boly M,** et al. (2012) Brain connectivity in disorders of consciousness. *Brain Connect* **2**:1-10 https://doi.org/10.1089/brain.2011.0049 | PubMed

[4] **Sporns O** (2014) Contributions and challenges for network models in cognitive neuroscience. *Nat. Neurosci* **17**:652-660 https://doi.org/10.1038/nn.3690 | PubMed

[5] **Petersen SE,** Sporns O (2015) Brain networks and cognitive architectures. *Neuron* **88**:207-219 https://doi.org/10.1016/j.neuron.2015.09.027 | PubMed

[6] **Sporns O** (2016) *Networks of the brain* London, England: The MIT Press.

[7] **Seguin C,** et al. (2023) Communication dynamics in the human connectome shape the cortex-wide propagation of direct electrical stimulation. *Neuron* **111**:1391-1401.e5 https://doi.org/10.1016/j.neuron.2023.01.027 | PubMed

[8] **Stevens MC** (2009) The developmental cognitive neuroscience of functional connectivity. *Brain Cogn* **70**:1-12 https://doi.org/10.1016/j.bandc.2008.12.009 | PubMed

[9] **Honey CJ,** et al. (2009) Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U. S. A* **106**:2035-2040 https://doi.org/10.1073/pnas.0811168106 | PubMed

[10] **Rubinov M,** Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**:1059-1069 https://doi.org/10.1016/j.neuroimage.2009.10.003 | PubMed

[11] **Friston K,** Moran R, Seth AK (2013) Analysing connectivity with granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol* **23**:172-178 https://doi.org/10.1016/j.conb.2012.11.010 | PubMed

[12] **Siegle J,** et al. (2021) Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**:86-92 https://doi.org/10.1038/s41586-020-03171-x | PubMed

[13] **Yan Y,** Murphy TH (2024) Decoding state-dependent cortical-cerebellar cellular functional connectivity in the mouse brain. *Cell Rep* **43**:114348 https://doi.org/10.1016/j.celrep.2024.114348 | PubMed

[14] **Park AH,** et al. (2016) Optogenetic mapping of functional connectivity in freely moving mice via insertable wrapping electrode array beneath the skull. *ACS Nano* **10**:2791-2802 https://doi.org/10.1021/acsnano.5b07889 | PubMed

[15] **Kucyi A,** et al. (2018) Intracranial electrophysiology reveals reproducible intrinsic functional connectivity within human brain networks. *J. Neurosci* **38**:4230-4242 https://doi.org/10.1523/jneurosci.0217-18.2018 | PubMed

[16] **Greicius MD,** et al. (2007) Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiatry* **62**:429-437 https://doi.org/10.1016/j.biopsych.2006.09.020 | PubMed

[17] **Rocchi F,** et al. (2022) Increased fMRI connectivity upon chemogenetic inhibition of the mouse prefrontal cortex. *Nat. Commun* **13**:1056 https://doi.org/10.1038/s41467-022-28591-3 | PubMed

[18] **Nentwich M,** et al. (2020) Functional connectivity of EEG is subject-specific, associated with phenotype, and different from fMRI. *Neuroimage* **218**:117001 https://doi.org/10.1016/j.neuroimage.2020.117001 | PubMed

[19] **Bogéa Ribeiro L,** da Silva Filho M (2023) Systematic review on EEG analysis to diagnose and treat autism by evaluating functional connectivity and spectral power. *Neuropsychiatr Dis Treat* **19**:415-424 https://doi.org/10.2147/ndt.s394363 | PubMed

[20] **Reid AT,** et al. (2019) Advancing functional connectivity research from association to causation. *Nat. Neurosci* **22**:1751-1760 https://doi.org/10.1038/s41593-019-0510-4 | PubMed

[21] **Goebel R,** Roebroeck A, Kim D.-S, Formisano E (2003) Investigating directed cortical interactions in timeresolved fMRI data using vector autoregressive modeling and granger causality mapping. *Magn Reson Imaging* **21**:1251-1261 https://doi.org/10.1016/j.mri.2003.08.026 | PubMed

[22] **Ding M,** Chen Y, Bressler SL (2006) *Handbook of time series analysis: Recent theoretical developments and applications* (1) Weinheim, Germany: Wiley-VCH Verlag.

[23] **Ting C,** Seghouane A, Khalid MU, Salleh S (2015) Is first-order vector autoregressive model optimal for fMRI data?. *Neural Comput* **27**:1857-1871 https://doi.org/10.1162/neco_a_00765 | PubMed

[24] **Seth A,** Barrett A, Barnett L (2015) Granger causality analysis in neuroscience and neuroimaging. *J. Neurosci* **35**:3293-3297 https://doi.org/10.1523/jneurosci.4399-14.2015 | PubMed

[25] **Cekic S,** Grandjean D, Renaud O (2018) Time, frequency, and time-varying granger-causality measures in neuro-science. *Stat Med* **37**:1910-1931 https://doi.org/10.1002/sim.7621 | PubMed

[26] **Ruiz S,** Buyukturkoglu K, Rana M, Birbaumer N, Sitaram R (2014) Real-time fMRI brain computer interfaces: self-regulation of single brain regions to networks. *Biol. Psychol* **95**:4-20 https://doi.org/10.1016/j.biopsycho.2013.04.010 | PubMed

[27] **Bassett D,** Sporns O (2017) Network neuroscience. *Nat. Neurosci* **20**:353-364 https://doi.org/10.1038/nn.4502 | PubMed

[28] **Lee M,** Yoon J.-G, Lee S.-W (2020) Predicting motor imagery performance from resting-state EEG using dynamic causal modeling. *Front. Hum. Neurosci* **14**:321 https://doi.org/10.3389/fnhum.2020.00321 | PubMed

[29] **Bergmann TO,** et al. (2021) Concurrent TMS-fMRI for causal network perturbation and proof of target engagement. *Neuroimage* **237**:118093 https://doi.org/10.1016/j.neuroimage.2021.118093 | PubMed

[30] **Siddiqi SH,** Kording KP, Parvizi J, Fox MD (2022) Causal mapping of human brain function. *Nature reviews neuroscience* **23**:361-375 https://doi.org/10.1038/s41583-022-00583-8 | PubMed

[31] **Casali AG,** et al. (2013) A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med* **5** https://doi.org/10.1126/scitranslmed.3006294 | PubMed

[32] **Stepaniants G,** Brunton BW, Kutz JN (2020) Inferring causal networks of dynamical systems through transient dynamics and perturbation. *Phys. Rev. E* **102**:042309 https://doi.org/10.1103/physreve.102.042309 | PubMed

[33] **Lepperød ME,** Stöber T, Hafting T, Fyhn M, Kording KP (2023) Inferring causal connectivity from pairwise recordings and optogenetics. *PLoS Comput. Biol* **19**:e1011574 https://doi.org/10.1371/journal.pcbi.1011574 | PubMed

[34] **Steinberg EE,** Janak PH (2013) Establishing causality for dopamine in neural function and behavior with optogenetics. *Brain Res* **1511**:46-64 https://doi.org/10.1016/j.brainres.2012.09.036 | PubMed

[35] **Lepperød ME,** Stöber T, Hafting T, Fyhn M, Kording KP (2023) Inferring causal connectivity from pairwise recordings and optogenetics. *PLoS Comput. Biol* **19**:e1011574 https://doi.org/10.1371/journal.pcbi.1011574 | PubMed

[36] **Hallett M,** et al. (2017) Contribution of transcranial magnetic stimulation to assessment of brain connectivity and networks. *Clin. Neurophysiol* **128**:2125-2139 https://doi.org/10.1016/j.clinph.2017.08.007 | PubMed

[37] **Tik M,** et al. (2023) Acute TMS/fMRI response explains offline TMS network effects - an interleaved TMS-fMRI study. *Neuroimage* **267**:119833 https://doi.org/10.1016/j.neuroimage.2022.119833 | PubMed

[38] **Massimini M,** et al. (2007) Triggering sleep slow waves by transcranial magnetic stimulation. *Proc. Natl. Acad. Sci. U. S. A* **104**:8496-8501 https://doi.org/10.1073/pnas.0702495104 | PubMed

[39] **George MS** (2019) Whither TMS: A one-trick pony or the beginning of a neuroscientific revolution?. *Am. J. Psychiatry* **176**:904-910 https://doi.org/10.1176/appi.ajp.2019.19090957 | PubMed

[40] **Bernal-Casas D,** Lee HJ, Weitz AJ, Lee JH (2017) Studying brain circuit function with dynamic causal modeling for optogenetic fMRI. *Neuron* **93**:522-532.e5 https://doi.org/10.1016/j.neuron.2016.12.035 | PubMed

[41] **Grimm C,** et al. (2024) Tonic and burst-like locus coeruleus stimulation distinctly shift network activity across the cortical hierarchy. *Nat. Neurosci* **27**:2167-2177 https://doi.org/10.1038/s41593-024-01755-8 | PubMed

[42] **Gu S,** et al. (2015) Controllability of structural brain networks. *Nat. Commun* **6**:8414 https://doi.org/10.1038/ncomms9414 | PubMed

[43] **Deng S,** Li J, Thomas Yeo B, Gu S (2022) Control theory illustrates the energy efficiency in the dynamic reconfiguration of functional connectivity. *Commun. Biol* **5**:295 https://doi.org/10.1038/s42003-022-03196-0 | PubMed

[44] **Kawakita G,** Kamiya S, Sasai S, Kitazono J, Oizumi M (2022) Quantifying brain state transition cost via schrödinger bridge. *Netw Neurosci* **6**:118-134 https://doi.org/10.1162/netn_a_00213 | PubMed

[45] **Kamiya S,** Kawakita G, Sasai S, Kitazono J, Oizumi M (2023) Optimal control costs of brain state transitions in linear stochastic systems. *J. Neurosci* **43**:270-281 https://doi.org/10.1523/jneurosci.1053-22.2022 | PubMed

[46] **Moradi Amani A,** et al. (2024) Controllability of functional and structural brain networks. *Complexity* **2024** https://doi.org/10.1155/2024/7402894

[47] **Shikauchi Y,** et al. (2025) Quantifying state-dependent control properties of brain dynamics from perturbation responses. *J. Neurosci* e0364252025 https://doi.org/10.1523/JNEUROSCI.0364-25.2025 | PubMed

[48] **Minai Y,** Smith M, Soldado-Magraner J, Yu B (2024) MiSO: Optimizing brain stimulation to create neural activity states. In: Advances in Neural Information Processing Systems 37. **37** pp. 24126-24149 https://doi.org/10.52202/079017-0760

[49] **Wagenmaker A,** et al. (2024) Active learning of neural population dynamics using two-photon holographic optogenetics. *Adv. Neural Inf. Process. Syst* **37**:31659-31687 PubMed

[50] **Angulo-Sherman IN,** Rodríguez-Ugarte M, Sciacca N, Iáñez E, Azorín JM (2017) Effect of tDCS stimulation of motor cortex and cerebellum on EEG classification of motor imagery and sensorimotor band power. *J. Neuroeng. Rehabil* **14**:31 https://doi.org/10.1186/s12984-017-0242-1 | PubMed

[51] **Karrer TM,** et al. (2020) A practical guide to methodological considerations in the controllability of structural brain networks. *J. Neural Eng* **17**:026031 https://doi.org/10.1088/1741-2552/ab6e8b | PubMed

[52] **Betzel RF,** Gu S, Medaglia JD, Pasqualetti F, Bassett DS (2016) Optimally controlling the human connectome: the role of network topology. *Sci. Rep* **6**:30770 https://doi.org/10.1038/srep30770 | PubMed

[53] **Kim JZ,** et al. (2018) Role of graph architecture in controlling dynamical networks with applications to neural systems. *Nat. Phys* **14**:91-98 https://doi.org/10.1038/nphys4268 | PubMed

[54] **Braun U,** et al. (2021) Brain network dynamics during working memory are modulated by dopamine and diminished in schizophrenia. *Nat. Commun* **12**:3478 https://doi.org/10.1038/s41467-021-23694-9 | PubMed

[55] **Ahmed S,** Nozari E (2022) On the linearizing effect of spatial averaging in large-scale populations of homogeneous nonlinear systems. In: 2022 IEEE 61st Conference on Decision and Control (CDC). https://doi.org/10.1109/cdc51059.2022.9993260

[56] **Nozari E,** et al. (2024) Macroscopic resting-state brain dynamics are best described by linear models. *Nat. Biomed. Eng* **8**:68-84 https://doi.org/10.1038/s41551-023-01117-y | PubMed

[57] **Momi D,** Wang Z, Griffiths JD (2023) TMS-evoked responses are driven by recurrent large-scale network dynamics. *eLife* **12** https://doi.org/10.7554/elife.83232 | PubMed

[58] **Tapsell LC,** et al. (2025) What are the optimal transcranial direct current stimulation parameters and design elements to modulate corticospinal excitability? a systematic review and longitudinal meta-analysis. *Neurol. Res. Pract* **7**:86 https://doi.org/10.1186/s42466-025-00449-1 | PubMed

[59] **Grover S,** Fayzullina R, Bullard BM, Levina V, Reinhart RMG (2023) A meta-analysis suggests that tACS improves cognition in healthy, aging, and psychiatric populations. *Sci. Transl. Med* **15**:eabo2044 https://doi.org/10.1126/scitranslmed.abo2044 | PubMed

[60] **Ogata K** (2010) *Modern Control Engineering* Prentice Hall.

[61] **Nise NS** (2020) *Control Systems Engineering* (8) John Wiley & Sons.

[62] **Lozano AM,** et al. (2019) Deep brain stimulation: current challenges and future directions. *Nat. Rev. Neurol* **15**:148-160 https://doi.org/10.1038/s41582-018-0128-2 | PubMed

[63] **Deisseroth K** (2011) Optogenetics. *Nat. Methods* **8**:26-29 https://doi.org/10.1038/nmeth.f.324 | PubMed

[64] **Tong L,** et al. (2023) Single cell in vivo optogenetic stimulation by two-photon excitation fluorescence transfer. *iScience* **26**:107857 https://doi.org/10.1016/j.isci.2023.107857 | PubMed

[65] **LaFosse PK,** et al. (2024) Single-cell optogenetics reveals attenuation-by-suppression in visual cortical neurons. *bioRxiv* https://doi.org/10.1101/2023.09.13.557650 | PubMed

[66] **Green M,** Moore JB (1986) Persistence of excitation in linear systems. *Syst Control Lett* **7**:351-360 https://doi.org/10.1016/0167-6911(86)90052-6

[67] **Shimkin N,** Feuer A (1987) Persistency of excitation in continuous-time systems. *Syst Control Lett* **9**:225-233 https://doi.org/10.1016/0167-6911(87)90044-2

[68] **Jenkins BM,** Annaswamy AM, Lavretsky E, Gibson TE (2018) Convergence properties of adaptive systems and the definition of exponential stability. *SIAM J Control Optim* **56**:2463-2484 https://doi.org/10.1137/15m1047805 | PubMed

[69] **Lu X,** Cannon M (2023) Robust adaptive model predictive control with persistent excitation conditions. *Automatica* **152**:110959 https://doi.org/10.1016/j.automatica.2023.110959

[70] **Hamilton JD** (1994) *Time Series Analysis* Princeton: Princeton University Press.

[71] **Paulus W** (2011) Transcranial electrical stimulation (tES - tDCS; tRNS, tACS) methods. *Neuropsychol Rehabil* **21**:602-617 https://doi.org/10.1080/09602011.2011.557292 | PubMed

[72] **Paulus W,** Nitsche MA, Antal A (2016) Application of transcranial electric stimulation (tDCS, tACS, tRNS): From motor-evoked potentials towards modulation of behaviour. *Eur. Psychol* **21**:4-14 https://doi.org/10.1027/1016-9040/a000242

[73] **Fernández Galán R** (2008) On how network architecture determines the dominant patterns of spontaneous neural activity. *PLoS One* **3**:e2148 https://doi.org/10.1371/journal.pone.0002148 | PubMed

[74] **Khalil HK** (2002) *Nonlinear systems* Upper Saddle River, NJ: Prentice-Hall.

[75] **Koopman BO** (1931) Hamiltonian systems and transformation in hilbert space. *Proc. Natl. Acad. Sci. U. S. A* **17**:315-318 https://doi.org/10.1073/pnas.17.5.315 | PubMed

[76] **Chow C,** Dan T, Styner M, Wu G (2024) Understanding brain dynamics through neural koopman operator with structure-function coupling. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 509-518

[77] **Suk H.-I,** Wee C.-Y, Lee S.-W, Shen D (2016) State-space model with deep learning for functional dynamics estimation in resting-state fMRI. *Neuroimage* **129**:292-307 https://doi.org/10.1016/j.neuroimage.2016.01.005 | PubMed

[78] **Glaser JI,** Benjamin AS, Farhoodi R, Kording KP (2019) The roles of supervised machine learning in systems neuroscience. *Prog. Neurobiol* **175**:126-137 https://doi.org/10.1016/j.pneurobio.2019.01.008 | PubMed

[79] **Tseng SY,** Chen RC, Chong FC, Kuo TS (1995) Evaluation of parametric methods in EEG signal analysis. *Med. Eng. Phys* **17**:71-78 https://doi.org/10.1016/1350-4533(95)90380-t | PubMed

[80] **Chang J.-Y,** et al. (2012) Multivariate autoregressive models with exogenous inputs for intracerebral responses to direct electrical stimulation of the human brain. *Front. Hum. Neurosci* **6**:317 https://doi.org/10.3389/fnhum.2012.00317 | PubMed

[81] **Shakeel A,** Onojima T, Tanaka T, Kitajo K (2021) Real-time implementation of EEG oscillatory phase-informed visual stimulation using a least mean square-based AR model. *J. Pers. Med* **11**:38 https://doi.org/10.3390/jpm11010038 | PubMed

[82] **Bikson M,** et al. (2016) Safety of transcranial direct current stimulation: Evidence based update 2016. *Brain Stimul* **9**:641-661 https://doi.org/10.1016/j.brs.2016.06.004 | PubMed

[83] **Antal A,** et al. (2017) Low intensity transcranial electric stimulation: Safety, ethical, legal regulatory and application guidelines. *Clin. Neurophysiol* **128**:1774-1809 https://doi.org/10.1016/j.clinph.2017.06.001 | PubMed

[84] **Mager T,** et al. (2018) High frequency neural spiking and auditory signaling by ultrafast red-shifted optogenetics. *Nat. Commun* **9**:1750 https://doi.org/10.1038/s41467-018-04146-3 | PubMed

[85] **Rossi S,** Hallett M, Rossini PM, Pascual-Leone A (2009) Safety of TMS Consensus Group. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin. Neurophysiol* **120**:2008-2039 https://doi.org/10.1016/j.clinph.2009.08.016 | PubMed

[86] **Heinrichs J.-H** (2012) The promises and perils of non-invasive brain stimulation. *Int. J. Law Psychiatry* **35**:121-129 https://doi.org/10.1016/j.ijlp.2011.12.006 | PubMed

[87] **Takens F** (1981) Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence, Warwick 1980, vol. 898 of Lecture Notes in Mathematics. pp. 366-381 https://doi.org/10.1007/bfb0091924

[88] **Arbabi H,** Mezić I (2017) Computation of transient koopman spectrum using hankel-dynamic mode decompoisition. In: Aps G1.009.

[89] **Brunton SL,** Brunton BW, Proctor JL, Kaiser E, Kutz JN (2017) Chaos as an intermittently forced linear system. *Nat. Commun* **8**:19 https://doi.org/10.1038/s41467-017-00030-8 | PubMed

[90] **Ostrow M,** Eisen A, Fiete I (2024) Delay embedding theory of neural sequence models. *arXiv* https://doi.org/10.48550/arxiv.2406.11993

[91] **Huang A,** et al. (2025) InputDSA: Demixing then comparing recurrent and externally driven dynamics. *arXiv* https://doi.org/10.48550/arxiv.2510.25943

[92] **Tu C,** et al. (2018) Warnings and caveats in brain controllability. *Neuroimage* **176**:83-91 https://doi.org/10.1016/j.neuroimage.2018.04.010 | PubMed

[93] **Suweis S,** et al. (2019) Brain controllability: Not a slam dunk yet. *Neuroimage* **200**:552-555 https://doi.org/10.1016/j.neuroimage.2019.07.012 | PubMed

Peer reviews

Joint Public Review:

Summary:

Inferring so-called "functional connectivity" between neurons or groups of neurons is important both for validating models and for inferring brain state, including in human patients. This study aims to enhance this inference process by using closed-loop perturbation-based approaches. To this end, the authors develop a framework based on linear dynamical models that minimizes the estimation error. Based on this framework, the authors provide a practical guide for applying it in realistic experiments. Modalities include non-invasive ones, such as fMRI, iEEG, and invasive ones, such as optogenetic perturbations combined with neuropixel probes or calcium imaging.

Strengths:

A main strength of this paper is the application and adaptation of an explicit error expression to system dynamics estimation from evoked neural responses, bringing a useful theoretical tool into computational neuroscience for, as far as we know, the first time. Importantly, while the analytical derivation assumes the neural dynamics is linear and the control signal is known, these assumptions do not appear to be essential: their method outperforms passive observation even when the true dynamics is nonlinear or the control input is not known perfectly. Moreover, the relative simplicity of the method makes its practical applications straightforward, as the authors illustrate in the context of brain state classification and neural control.

Besides being of practical importance, simply pointing out that passive observation can lead to large mis-estimation of functional connectivity should serve as a wakeup call to anybody engaged in this endeavor.

Weaknesses:

None.

https://doi.org/10.7554/eLife.110030.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Joint Public Review:***
>
> *Weaknesses:*
>
> *(1) The derivation of the main error term misses some important steps, which complicates peer review at this stage. In particular, factorisation of the covariance into*

> *noise and the inverse of the observation covariance matrix needs a more thorough justification. The cited sources do not contain the derivation for a noise term with full covariance, which is essential for deriving this error term.*

The derivation of the main error term misses some important steps, which complicates peer review at this stage

We thank the reviewers for this careful observation. This concern is associated with the error term. Thus, we first clarified the noise assumption explicitly. We assume that $\xi(t)$ is i.i.d. over time with zero mean and an arbitrary (not necessarily diagonal) positive definite covariance matrix $\mathbf{\Sigma}_\xi = \mathbb{E}\left[\boldsymbol{\xi}(t)\boldsymbol{\xi}(t)^\top\right] \in \mathbb{R}^{n\times n} \succ 0$.

In particular, factorisation of the covariance into noise and the inverse of the observation covariance matrix needs a more thorough justification. The cited sources do not contain the derivation for a noise term with full covariance, which is essential for deriving this error term.

The cited sources do contain the derivation for a noise term with full covariance. We have updated the citation that directly supports Eq. (S.2): Proposition 11.1 of Hamilton (1994, TimeSeries Analysis), which establishes the asymptotic distribution $\sqrt{T}\,\mathrm{vec}\left(\hat{\mathbf{\Pi}}_T - \pi\right) \xrightarrow{L} \mathcal{N}\left(0, \Omega \otimes Q^{-1}\right)$ The proof, given in Appendix 11.A of Hamilton (1994), proceeds via a CLT for martingale difference sequences. See Theoretical Details of the Supplementary Materials.

> *(2) The practical recommendation at the end of the paper also requires clearer guidance on how the design perturbations are constructed, and how many times and for how long the system is stimulated in each iteration of the experiment.*

Thank you for this helpful suggestion. We agree that the practical implementation of the experimental design should be explained more clearly. We have addressed this concern in two ways. First, we have revised the manuscript to explicitly describe the parameter design procedure. Second, we have revised the manuscript to clearly provide a reference to the detailed experimental condition table in the supplementary material. See Results - Main Manuscript.

> *(3) Finally, there is no analysis of model mis-specification. In particular, the true dynamics are unlikely to be linear; the noise is unlikely to be either Gaussian or uncorrelated across time; and the B matrix is unlikely to be known perfectly. We're not suggesting that the authors consider a more complex model, but it's important to know how sensitive their method is to model mismatch. If nothing can be done analytically, then simulations would at least provide some kind of guide.*

We thank the reviewer for raising this important point regarding model mis-specification. We agree that it is important to run simulations to assess the impact of these model mismatches, therefore we conducted preliminary simulations to assess the sensitivity when two primary assumptions are violated: linear state dynamics and a perfectly known input matrix B. We added these preliminary results in the Supplementary Material, and revised the main manuscript to include a mention of these results. In summary, the simulations showed:

The model estimation error increases with the strength of the nonlinearity; however, perturbation can increase the information and lead to accurate estimation of the hidden mode even under nonlinearity.

The matrix A can be estimated roughly even if the assumed input matrix B differs from the true matrix in some cases.

The matrices A and B can be estimated jointly without bias, provided the stimulation pattern excites the full state space.

In such joint estimation, preferentially exciting the hidden modes directly leads to more accurate estimation of A than stimulating other modes. See Background and Results - Main Manuscript, Experimental Conditions and Results - Supplementary Material.

> ***Recommendations for the authors:***
>
> *(1) Please tell us what tACS, tDCS, and TMS are, and how much control the experimenter has over them. That's important, because they are going to be used as control signals, so we need to know how accurately u(t) can be specified, and what its range is.*
>
> *Please tell us what tACS, tDCS, and TMS are, and how much control the experimenter has over them.*

We appreciate the reviewer's helpful comment. We agree that it is important to describe these stimulation methods with appropriate references. We have added a new section titled "Neural Stimulation as Control Inputs" to the background, which connects our theoretical framework to practical experimental settings.

> *We need to know how accurately u(t) can be specified, and what its range is.*

The accuracy and range of control inputs vary substantially depending on the specific stimulation technique and experimental setup, and a thorough discussion would require a dedicated review beyond the scope of this manuscript. Instead, we have added a sentence acknowledging the gap between practical experimental implementations and the theoretical formulation, and cited relevant references for readers interested in further details.

Background - Main Manuscript

"Neural Stimulation as Control Inputs

This section describes how commonly used neural stimulation techniques can be related to input signals in control theory. Their adjustable parameters vary depending on how the stimulation inputs are modulated.

Three non-invasive electrical stimulation methods illustrate how stimulation paradigms map onto basic control inputs. Transcranial magnetic stimulation (TMS) induces brief and transient perturbations via electromagnetic pulses [19], which are naturally represented as a sequence of impulse-like inputs, where the timing and intensity of each pulse are the primary controllable parameters. Transcranial direct current stimulation (tDCS) primarily modulates neural activity through approximately constant inputs [26], which can be viewed as a step-like signal whose main controllable parameter is the amplitude of the applied current. Transcranial alternating current stimulation (tACS) delivers oscillatory inputs [7], corresponding to sinusoidal signals characterized by amplitude, frequency, and phase. In control theory, impulse, step, and sinusoidal inputs are the basic components used to characterize system responses and dynamics [22, 21].

The control input framework extends beyond non-invasive techniques to invasive and optogenetic stimulation. Invasive electrical stimulation, including intracranial microstimulation and deep brain stimulation (DBS), enables direct delivery of electrical inputs to neural tissue [17], providing flexible control over amplitude and timing through pulse trains or temporally structured waveforms. Optogenetic stimulation allows genetically targeted activation or inhibition of specific neurons using light [5], providing fine-grained control over multiple input dimensions, including amplitude (light intensity), temporal pattern, and cell-type specificity. In particular, recent developments enable stimulation at the level of individual neurons with high temporal precision [27, 16], allowing flexible construction of spatiotemporal input patterns.

These stimulation examples demonstrate that the theoretical framework developed in this paper connects to practical experimental settings. While a substantial gap remains between idealized control inputs in theory and experimentally realizable stimulation, the core principles established in the following sections provide a foundation that naturally extends to these practical stimulation paradigms."

> *(2) Is the solid curve in the last panel of Figure 4b the prediction? If not, are the data points consistent with the prediction in Equation 8? This should be clear.*

Thank you for this question. The solid curve shows the empirical eigenvalues of the state covariance matrix, not the eigenvalues of matrix *A*. As shown in Equation 9, the estimation error is proportional to the inverse of the sum of the eigenvalues of the state covariance matrix. We have clarified these points by revising the main text and the figure captions. See Results - Main Manuscript.

> *(3) Page 10, "Nodes 7 and 8 have only outgoing edges". It looks like node 7 has an incoming edge from node 8 (Fig. 6b). Or are we misinterpreting something?*

Thank you for pointing out this inconsistency. You are correct. Node 7 did have an incoming edge from Node 8 and contradicted the statement in the text. Moreover, we now think that two hub nodes (Node 7 and 8) are not necessary for demonstrating our primary theory. To resolve these issues, we have revised the simulation so that the network now contains a single hub node with only outgoing edges. Please refer to Fig. 6.

> *(4) Figure 6f, g: why isn't the estimation error proportional to tr[Sig_x^{-1}]?*

We thank the reviewer for this observation. In the original manuscript, the estimation error in Figure 6f, g was plotted on a logarithmic scale, which obscured the proportional relationship with $\mathrm{tr}\left[\boldsymbol{\Sigma}_X^{-1}\right]$. The underlying values are indeed proportional, consistent with our theoretical prediction.

In the revised manuscript, we have substantially reorganized Figure 6 to make the theoretical reasoning more transparent by using $1/\mu$ instead of $\mathrm{tr}\left[\boldsymbol{\Sigma}_X^{-1}\right]$ following Eq. 9. The sum across the column in panel (g) is proportional to the estimation error shown in panel (h).

> *(5) Page 11, "The simulation was conducted with a single node receiving an impulse-shaped perturbation input." What's an "impulse-shaped perturbation input"? A delta function? Please make this clear.*

Thank you for this clarifying question. We clarified the explanation as follows.

Results - Main Manuscript

"Each node was individually perturbed by an impulse input. Here, an impulse input is defined as a Kronecker delta at t = 0 with fixed amplitude α = 10, with no external input applied at any subsequent time step."

> *(6) Page 11, "Fig. 6d shows that the system possesses modes with small absolute eigenvalues." According to Figure 6d, all the absolute eigenvalues are between 9.9 and 9.98. So this statement appears not to be correct. Are we missing something?*

Thank you for pointing this out. You are correct. The previous statement was inconsistent with the figure. In the revised simulation, we have redesigned the network so that it clearly contains modes with distinct damping characteristics: heavily damped modes with absolute eigenvalues below 0.8 and lightly damped modes with absolute eigenvalues close to 1.0 (see revised Fig. 6c). This makes the relationship between mode damping and perturbation effectiveness much more transparent.

Results - Main Manuscript

“Figs. 6c shows the damping rates $|\lambda A|$ of the eigenvalues of A: the mode formed by Nodes 1 and 2 (15Hz) is heavily damped, while those formed by Nodes 3–6 are moderately damped. Node 7 serves as a hub with only outgoing edges.”

> *(7) Page 11, "Crucially, the eigenvectors corresponding to these rapidly decaying modes (e.g., evec 7, evec 8) have their largest components concentrated at Nodes 7 and 8." If "nodes" are the same as "eigenvalue index", then the components on node 8 are zero (Figure 6e, bottom). In any case, it should be clear what you mean.*

Thank you for this comment. We agree that the previous description regarding eigenvectors was unclear and inconsistent with the figure. We now think that explaining with eigenvectors is not necessary for demonstrating our primary theory. We have therefore replaced the plots of eigenvectors with the reciprocals of the eigenvalues of Σ_X, which are directly and rigorously explained by Equation (9). These reciprocals clearly show that the modes associated with Nodes 1 and 2 are heavily damped and applying perturbations to these nodes contribute to the estimation error and the perturbations to hub node (Node 7) broadly increases all eigenvalues of Σ_X, thereby reducing the estimation error across all modes. This change ensures that all simulation results are grounded in the theory presented in the paper. Please refer to the revised Fig. 6d–g for details.

> *(8) It’s not clear to us what’s plotted in Figure 6e. The real part of the eigenvectors? Which would explain why some of the eigenvectors are the same (e.g., 1 and 2). But that does not seem like a good idea, since the eigenvectors can be rotated by an arbitrary complex phase. Also, eigenvectors 7 and 8 seem totally opposite, and there’s no weight at all on index 6 and index 8. Could there be a mistake in the figure? In addition, Figure 6e is explained and interpreted in two different paragraphs, discussing the same observation. It would be easier to understand if they were moved to the same paragraph. Also, in the second paragraph where plot 6e is referenced (page 11, line 19), what does ‘concentrated’ mean?*

Thank you for this comment. As described in our response to Comment (7), we have removed the eigenvector-based analysis from the revised simulation to prevent confusing readers. The revised results focus on quantities that are directly explained by Equation (9). Please refer to the revised Fig. 6 for the updated results.

> *(9) Page 11, "This simulation demonstrates that, given a tentative connectivity matrix, an effective perturbation input (such as TMS or tDCS) can be designed by targeting the node with the highest weighted out-degree." "Demonstrates" seems strong. There is only one simulation, and that wasn’t totally convincing: a perturbation applied to node 7 did almost the same as a perturbation applied to nodes 1-6, even though it had a much higher outdegree than those nodes. It would be very helpful if you provided theoretical reasoning for why perturbing nodes with high-weight connections (Figure 6) minimises the prediction error. In particular, which properties of a hub node make it effective as a stimulation target? Why is it just its total output weights and not also its number of edges/centrality? How is the high output weight of a node related to its alignment with other eigenvectors, and is this always the case or just in this example?*
>
> *Demonstrates seems strong.*

Thank you for this important comment. We agree that "demonstrates" was too strong and have replaced it with "illustrates."

> *It would be very helpful if you provided theoretical reasoning for why perturbing nodes with high-weight connections (Figure 6) minimises the prediction error.*

We agree with this comment. We have revised the simulation and accompanying text to connect the results directly to the main theory based on $\mathrm{tr}\left[\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right]$. As responded to Comments (7) and (8), we have removed the eigenvector-based analysis and instead plotted the reciprocals of the eigenvalues of Σ_X, which are directly related to the estimation error via Equation (9). This change allows us to provide a clear theoretical explanation for why perturbing certain nodes minimizes the prediction error.

> *Is this always the case or just in this example?*

We have also explicitly stated the limitations: whether a hub node or a specific subnetwork node is more effective depends on factors such as the outgoing edge weights from the hub and the individual damping rates of each mode. The revised text emphasizes that this simulation presents one example of perturbation location design, and the optimal strategy must be evaluated case by case using the theoretical framework of Equation (9).

Results - Main Manuscript

“As this simulation represents only one example of location design, its limitations and the corresponding countermeasures should be stated. In actual experiments, the most effective perturbation location depends on factors such as hub-node connectivity and modal damping rates, and *B* itself may not always be known a priori. In such cases, approaches such as iterative optimization of $\mathrm{tr}\left[\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right]$ (described in a later section) and joint estimation of *A* and *B* (Supplementary Material B.3) provide systematic alternatives. Nevertheless, the results presented here provide an intuitive guideline: stimulation directed at hub nodes or at nodes driving heavily damped modes effectively excites the full set of dynamical modes and minimizes estimation error.”

> *(10) What are the physical units for the time scales and stimulation amplitudes? For instance, on page 13, there is an argument that "In practical experiments, such long windows are unrealistic because neural states change rapidly over time." However, it is unclear whether T=100 a.u. or T=1000 a.u. etc. is realistic. By relating it to the eigenspectrum of A, which is supposed to be physiologically realistic, one can estimate the length of the stimulation window and support the above statement. Similarly, the impulse amplitude on page 13 is alpha=10^{20} (a.u.). Also, Table 1 in the Supplementary has an extremely wide range of stimulation amplitudes. Is 10^{20} a.u. a feasible amplitude in practice? And finally, please tell us which nodes the input was applied to.*

Thank you for your incisive comments. We have addressed each comment as follows.

> *What are the physical units for the time scales and stimulation amplitudes?*

They don’t have physical units. This study is a theoretical investigation that focuses on the relative differences between passive observation and perturbation-based approaches, rather than providing precise predictions for specific experimental settings. The time scales and stimulation amplitudes are therefore expressed in arbitrary units.

> *For instance, on page 13, there is an argument that "In practical experiments, such long windows are unrealistic because neural states change rapidly over time." However, it is unclear whether T=100 a.u. or T=1000 a.u. etc. is realistic.*

We agree that the original expression “unrealistic” was not appropriate given the arbitrary units. We have revised the text to clarify that the time scales are in arbitrary units and that the main point is about the relative difference in required data length between passive observation and perturbation-based approaches, rather than making an absolute claim about feasibility.

Results - Main Manuscript

“To obtain estimates under the passive condition that are comparable to those derived under perturbation, it is necessary to experimentally observe extensive time-series data. Figure 7f illustrates the LDA projection and classification accuracy for different time-series lengths. The leftmost LDA plot (T = 20) corresponds to the passive condition shown in Fig. 7c, indicating that the estimation performance in the passive condition becomes comparable to that in the perturbation condition only when the time window reaches approximately T = 200, a 10-fold increase compared to T = 20. While the absolute duration depends on the interpretation of the time unit, such time windows may not be prohibitive in some experimental settings. Nevertheless, our results consistently show that passive observation requires substantially longer recordings to achieve comparable performance, highlighting the efficiency of the perturbation-based approach when the available data length is limited.”

> *Is 10^{20} a.u. a feasible amplitude in practice?*

Although we have already stated that the stimulation amplitudes are in arbitrary units, we agree that the original value of 10^{20} was excessively large and could be misleading. We have revised the impulse amplitude from 10^{20} to 10^{2}, as 10^{20} is physically unrealistic—it would imply a stimulus intensity many orders of magnitude beyond any conceivable experimental setting. The revised value of 10^{2} is more plausible: for reference, TMS stimulation voltages exceed typical EEG amplitudes by roughly 4–6 orders of magnitude. The classification accuracy decreased slightly; however, our main conclusion regarding the efficiency of the perturbation-based approach under limited data remains unchanged. See Table B.1.

> *Finally, please tell us which nodes the input was applied to.*

The stimulus location was optimized to minimize the $\mathrm{tr}\left[\boldsymbol{\Sigma}_X^{-1}\right]$. We have clarified this procedure in the main text and added a visual indication of the selected stimulation site (red circles) in Fig.7.

Results - Main Manuscript

“The neural signals were simulated under five different task conditions and two stimulation conditions: passive observation and external perturbation. Perturbation was applied as impulse-type inputs, such as TMS. The stimulus location was determined for each task condition by applying an impulse to each node and selecting the one that minimized $\mathrm{tr}\left[\boldsymbol{\Sigma}_X^{-1}\right]$. The resulting time-series data are shown in Fig. 7b. Using this data, we estimated the underlying dynamical system via a control-based identification approach presented in Eq. 5, which corresponds to an estimation of functional connectivity.

> *(11) Page 13: "The controlled transition test was run with T = 1." Previously, T referred to the number of time steps. Is that the case here? If so, that seems hard to justify. If not, please tell us what T is (and, ideally, use a different symbol).*
>
> *Is that the case here?*

No. In this context, T does not denote the number of time steps.

> *If not, please tell us what T is (and, ideally, use a different symbol).*

Thank you for your helpful comment. We have standardized the notation throughout the manuscript. In this paper, T consistently denotes the number of time steps (i.e., data length). In the sections “Neural State Classification” and “Neural State Transitions,” we had mistakenly used T to refer to time length. To resolve this inconsistency, we have added a separate column labeled “Data Length (T)” to Table B.1 for clarification and replaced the previous usage of T with “data length” where appropriate.

Results - Main Manuscript

“To obtain estimates under the passive condition that are comparable to those derived under perturbation, it is necessary to experimentally observe extensive time-series data. Figure 7f illustrates the LDA projection and classification accuracy for different time-series lengths. The leftmost LDA plot (T = 20) corresponds to the passive condition shown in Fig. 7c, indicating that the estimation performance in the passive condition becomes comparable to that in the perturbation condition only when the time window reaches approximately T = 200, a 10-fold increase compared to T = 20. While the absolute duration depends on the interpretation of the time unit, such time windows may not be prohibitive in some experimental settings. Nevertheless, our results consistently show that passive observation requires substantially longer recordings to achieve comparable performance, highlighting the efficiency of the perturbation-based approach when the available data length is limited.”

Results - Main Manuscript

“The controlled transition test was run with T = 50. The control objective was to set nodes3 and 4 to 25 while keeping all other nodes at 0 without any movement. See Table B.1.”

> *(12) Page 14: "where the matrix A (Fig. 10a) is designed to have 16 modes." What do you mean by has "16 nodes"?*

Thank you for pointing this out. By “16 modes,” we refer to 16 dynamical eigenmodes (i.e., 16 eigenvalue pairs). In the real-valued state-space representation used in the simulations, each complex conjugate pair corresponds to a 2-dimensional real block, resulting in a 32-dimensional system (32 nodes). We revised the wording to clearly distinguish between the number of dynamical modes and the dimensionality (number of nodes) of the state vector, to avoid confusion.

Results - Main Manuscript

“Iterative refinement of both the perturbation design and the estimation process progressively improves the accuracy of A. The time-series data is collected from 32 points, where the matrix A (Fig. 10a) is designed to have 16 oscillatory modes (i.e., 16 complex-conjugate eigenvalue pairs, yielding 32 eigenvalues in total).”

> *(13) In Figure 10b, the y-axis should start at zero; otherwise, it’s a bit misleading how much the active perturbation helps. This will make it clear that the estimation error drops by about 33%. It would be worth commenting on whether this is typical; after all, potential users of this method would want to know how much improvement they’re likely to see.*
>
> *It would be worth commenting on whether this is typical; after all, potential users of this method would want to know how much improvement they’re likely to see.*

Thank you for this insightful comment. We agree with the reviewer that quantifying the expected improvement would be valuable for experimental practice. However, this simulation is a theoretical demonstration. Its primary purpose was to show that an iterative active perturbation approach can progressively converge to an optimal perturbation design even without prior knowledge of the true system, rather than to quantify a universally expected improvement rate.

> *The y-axis should start at zero; otherwise, it’s a bit misleading how much the active perturbation helps.*

We believe that the y-axis should start at the accuracy with optimal perturbation because the primary purpose of this simulation was to demonstrate that an iterative active perturbation

approach can progressively converge. If we had started the y-axis at zero, the message would be visually obscured.

> *Potential users of this method would want to know how much improvement they're likely to see.*

We acknowledge that this is one example of the application of our method, and the magnitude of improvement depends on various factors such as network structure, noise level, and stimulation design. We should not mislead the readers. We have therefore maintained the y-axis starting point and added a clarifying statement in the revised manuscript to indicate that the simulation is case-specific rather than universal.

Results - Main Manuscript

"These results should be interpreted as a case-specific illustration rather than a universal gain, as the magnitude of improvement depends on factors such as network structure, recording duration, noise level, and stimulation design. This simulation demonstrates that our perturbation design framework enables the step-by-step refinement of system identification even without prior knowledge of the system."

> *(14) Please provide a derivation for the factorised covariance in Equation S.2. This is the equation that underpins the main result of the paper - the error in dynamical system estimation. Currently, it appears to be taken from [Hamilton, J. D. Time Series Analysis], yet we were unable to find this result in the book. Most of the derivations in Chapters 8.1 and 8.2 assume diagonal noise covariance, and even isotropic noise (cov = sigma*2 * I), which simplifies the particular case of the derivations. Could the authors provide a reference or the derivation for the case with full covariance?*

As described in our response to the Weakness above, the relevant result is Proposition 11.1 of Hamilton (1994), not Chapters 8.1–8.2. Proposition 11.1 states the asymptotic distribution of the vectorised OLS estimator in a VAR model with i.i.d. innovations whose covariance matrix Ω is an arbitrary positive definite matrix. The factorisation $\boldsymbol{\Sigma}_{\text{Asym}} = \boldsymbol{\Sigma}_{\xi} \otimes \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$ in our Eq. (S.2) corresponds directly to the revised manuscript we explicitly cite "Hamilton (1994), Proposition 11.1" at Eq. (S.1) and $\boldsymbol{\Sigma}_{\text{Asym}} = \boldsymbol{\Sigma}_{\xi} \otimes \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$ in Hamilton's Proposition 11.1, with $\boldsymbol{\Sigma}_{\text{Asym}} = \boldsymbol{\Sigma}_{\xi} \otimes \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$ and $\xi(t)$ in the preceding paragraph, so that the connection to this result is unambiguous. See Theoretical Details Supplementary.$\boldsymbol{\Sigma}_{\text{Asym}} = \boldsymbol{\Sigma}_{\xi} \otimes \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$. In state the i.i.d. assumption on

> *(15) Strongly perturbing/Exciting fast-decaying modes to give them more 'runway' and increase observed variability makes intuitive sense for a normal system. But what would happen in the case of non-normal dynamics, where stimulating one dimension only transiently amplifies it, but then excites other dimensions? Non-normality breaks the alignment between PC components and dynamic modes (see Kumar, Ankit, Loren M. Frank, and Kristofer E. Bouchard. "Identifying feedforward and feedback controllable subspaces of neural population dynamics." arXiv preprint arXiv:2408.05875 (2024)), so eigenvectors of A and Σ_X won't align for non-normal dynamics. This alignment appears to be a hidden assumption of this paper. Should the normal dynamics then be stated as an assumption/limitation of the framework? Is the example in Figure 6a highly non-normal? Does considering out-degree provide an empirical approach, an alternative to 'enlargement', to dealing with non-normality?*

We thank the reviewer for this insightful comment regarding non-normal dynamics.

> *Should normal dynamics be stated as an assumption/limitation of the framework?*

No. Our framework does not assume normal dynamics. For example, Figure 5 and the revised Figure 6 illustrate non-normal cases. In Figure 5, the row and column norms of *A* differ

substantially (row norms ≈ [1.96, 4.05, 1.33, 4.70]; column norms ≈ [6.14, 1.94, 1.39, 0.87]). In Figure 6, Node 7 is a hub with only outgoing edges, so row 7 of A is zero while column 7 has nonzero entries. In addition, the Nodes 5–6 and Nodes 3–4 are strictly one-way, leaving the corresponding off-diagonal block upper-triangular. Both features break the symmetry required for normality.

We additionally computed a commutator-based non-normality index $\|AA^\top - A^\top A\|_F / \|A\|_F^2$ which equals 0 for any normal matrix and approaches $\|AA^\top - A^\top A\|_F / \|A\|_F^2$ for the canonical maximally non-normal example (the 2×2 nilpotent Jordan block). The non-normality index is 1.34 for Figure 5 and 0.41 for Figure 6. These values confirm that both are clearly non-normal.

> *Is the example in Figure 6a highly non-normal?*

Yes. As stated above, the network structure of Figure 6a clearly indicates non-normality.

> *Does considering out-degree provide an empirical approach, an alternative to 'enlargement', to dealing with non-normality?*

No. In the previous manuscript, the results of out-degree and eigenvector structure were provided as supplementary intuition. However, the central contribution of our framework lies in maximizing the minimum eigenvalue μ of the observed state covariance (Equation 9), and for systems where non-normality is strong and subnetwork structure is less modular, the iterative optimization of $\mathrm{tr}\left[\Sigma_X^{-1}\right]$ provides a principled, assumption-free method. We clearly mentioned this point in the revised manuscript as follows. See Results in the main manuscript.

> *(16) In the iterative experiment at the very end of the paper (Figure 10), what was the strategy for designing 'u_{design}'? How many stimulations were applied in an iteration? Are you stimulating along eigenvectors? Do you sample from components randomly, or perturb each of them individually, with the amplitude proportional to reciprocals?*

Thank you for this important question. We have clarified the design rule and stimulation protocol in the revised manuscript, and address each sub-question below.

> *How many stimulations were applied in an iteration?*

One stimulation session was applied in an iteration.

> *Are you stimulating along eigenvectors?*

The optimal stimulation is designed as the target node of the perturbation is determined through numerical optimization that minimizes $\mathrm{tr}\left[\Sigma_X^{-1}\right]$

> *Do you sample from components randomly, or perturb each of them individually, with the amplitude proportional to reciprocals?*

The optimal stimulation is designed as a composite-frequency sinusoidal input encompassing all modes of the estimated $\hat{A}$, and the target node of the perturbation is determined through numerical optimization that minimizes $\mathrm{tr}\left[\Sigma_X^{-1}\right]$. See Results - Main Manuscript.

> *(17) While it is clear that the proposed active method performs better than passive observation, some results lack a comparison with stimulating random directions/nodes with a comparable control energy (Figure 7e & Figure 10b).*

Thank you for your helpful comment. Although the optimally designed stimulation outperforms random stimulation, the previous stimulation settings were not configured to explicitly demonstrate this difference. Therefore, we modified the network structure,

recording length, and stimulation intensity so that both the main messages and the difference from random stimulation can be shown simultaneously. Accordingly, we have added a comparison with random stimulation in both Figure 7e and Figure 10b.

In Figure 7e, we included a random stimulation condition where the target node is selected randomly, and the results show that the optimized stimulation outperforms random stimulation.

These additions strengthen the evidence for the effectiveness of our proposed method compared to non-optimized approaches.

> *(18) Is it reasonable to assume full observability of the system? It would be interesting to consider biases arising from the partial observability of the system, in the spirit of Figure 9, which looked at partial controllability.*

We thank the reviewer for this suggestion. We agree that partial observability is an important consideration, however it is out of scope for the current work. Thus, we have added a future direction in the Discussion addressing partial observability. We note that Takens' embedding theorem and Hankel DMD enable recovery of a system's eigenvalues from partial observations, and since our framework relies on the eigenvalue structure of A, the proposed perturbation design remains applicable under partial observability.

Discussion - Main Manuscript

"Two directions warrant further investigation: extending the framework to partial observability, and validating it through stimulation experiments. In experimental neuroscience, recordings are often limited to a subset of neural populations, resulting in partial observability. A growing body of work has leveraged delay-embedding techniques, represented by Takens' embedding theorem [25], to reconstruct hidden dynamics from partial observations [2, 3, 23, 11]. Applying such techniques enables the estimation of the full connectivity matrix, thereby extending our framework to settings with partial observability. The second direction concerns experimental validation. Validating a theoretical framework through experimental design is an essential in bridging the gap between theory and practice."

> *Recommendations for improving the writing and presentation.*
>
> *(1) The word 'state' is overloaded in Figure 7. When talking about neural state classification, the 'state' refers to a regime guided by a distinct dynamics A (should it be A_i? Figure 7A bottom). However, each dynamical system also has a 'state'. A different word should be used in Figure 7A and the corresponding text.*

We agree that the terminology is potentially confusing. To avoid the confusion, we replaced the term "state" with "task condition" and "neural signal" in the main text, and revised Fig. 7.

Results - Main Manuscript

"We designed a neural network with clearly distinct task conditions and considered a simulation setting in which these conditions are classified using signals of a fixed duration. These distinct task conditions consist of five types, each defined by a unique linear dynamical system characterized by differing eigenvalue spectra and connectivity topologies of matrix *A* (Fig. 7a). These task conditions are intended to mimic different cognitive or behavioral contexts. For example, in a typical motor task experiment, such conditions could correspond to motor execution or imagery involving the left or right hand, or resting state [1, 24]. The neural signal was simulated under five different task conditions and two stimulation conditions: passive observation and external perturbation. Perturbation was applied as impulse-type inputs, such as TMS. The stimulus location was determined for each task

condition by applying an impulse to each node and selecting the one that minimized $\mathrm{tr}\left[\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\right]$. The resulting time-series data are shown in Fig. 7b. Using this data, we estimated the underlying dynamical system via a control-based identification approach presented in Eq. 5, which corresponds to an estimation of functional connectivity."

> *(2) It would be helpful to provide dimensions of matrices around Equation S.2, since vectorization makes dimensions hard to track.*

We thank the reviewer for this helpful suggestion. We agree that explicitly stating the matrix dimensions improves readability, particularly around the Kronecker product where vectorization can obscure the size of the resulting covariance matrix. We have revised the text as follows (the equation S.2 is 3 now). See Theoretical Details of the Supplementary Materials.

> *(3) The background sections of the paper would benefit from referring to similar active perturbation methods: Wagenmaker, Andrew, et al. "Active learning of neural population dynamics using two-photon holographic optogenetics." Advances in Neural Information Processing Systems 37 (2024): 31659-31687. Minai, Yuki, et al. "MiSO: Optimizing brain stimulation to create neural activity states." Advances in Neural Information Processing Systems 37 (2024): 24126-24149.*

We thank the reviewer for these helpful suggestions. We have incorporated Wagenmaker et al. (2024) and Minai et al. (2024) into the Introduction. Their works focus on developing algorithmic approaches to active stimulation design for specific experimental platforms, while our work aims to establish a general theoretical framework for why and which perturbation inputs are effective has yet to be established. We cited these works and have clarified this distinction in the revised manuscript as follows.

Introduction - Main Manuscript

"In this paper, we propose a framework for designing the optimal perturbation input through control theory in neuroscience. We interpret neural dynamics as a control system [8, 6, 14, 12, 20, 24], and treat external perturbations as control inputs to design properties of neural stimulation (Fig. 1d). If the optimal perturbation input can be systematically designed, it becomes possible to steer the neural system toward states that are maximally informative (Fig. 1e), thereby enhancing the accuracy of the inferred connectivity (Fig. 1f). While recent studies have begun to develop algorithmic approaches to active stimulation design for specific experimental platforms [18, 28], a general theoretical framework for why and which perturbation inputs are effective has yet to be established. We first describe how to formulate neural dynamics as a control system and how to estimate the model parameters from observed data. Building upon this formulation, we derive a theoretical basis that enables us to design the optimal perturbation inputs for the neural system identification. We demonstrate the validity and utility of this theoretical basis by exploring its implications for optimizing parameters of common neurostimulation techniques and by applying it to practical examples, including neural state classification [4, 9, 1, 24] and control of neural states [8, 13, 14, 12]. In these demonstrations, we define concrete problems and apply the theory to validate its practical utility."

> *Minor corrections to the text and figures.*
>
> *(1) In Figure 3, it would be helpful to point out that the plots are in the subspace spanned by the first three principal components.*

Thank you for pointing out. We revised the caption of the Fig.3 as follows. See Results - Main Manuscript.

> *(2) Figure 4 caption: "eigenvectors" –> "eigenvectors of Σ_X", just to make it crystal clear (since A also has eigenvectors).*

Thank you for your suggestion. We have revised the caption of Figure 4. See Results - Main Manuscript.

> *(3) Page 5, Equation (8): Capital xi should be introduced in the main text of the paper as the covariance matrix for the noise term xi. Now it can only be understood after reading the Supplementary. Or maybe call the covariance matrix Σ_ξ rather than Σ_ξ? That will probably make it clearer.*

Thank you for this suggestion. We have made both changes. First, we introduced the noise covariance matrix explicitly in the main text immediately after Eq. (1), defining. Second, we replaced the notation $\Sigma_{\tilde{\xi}}$ with Σ_ξ (lowercase subscript matching the noise variable $\Sigma_\xi = \mathbb{E}\left[\xi(t)\xi(t)^T\right] > 0$ throughout the main text and supplementary.

> *(4) Page 7, Equation (14): The description of the equation states "The covariance matrices of x(t) for impulse inputs can be written as:... ". We assume this is supposed to be the "state vector," not "covariance matrices".*

Thank you for catching this. We have corrected the wording to “state vector” instead of “covariance matrices". See Results - Main Manuscript

> *(5) Page 10, after introducing Figures 6a-b, potentially a sentence is missing ('...' in the first line of the last paragraph).*

Thank you for pointing this out. We have removed the placeholder along with updating the stimulation settings for Figure 6.

> *(6) Page 10, Figure 6 (e): A clearer labelling would be helpful, e.g., a title for the legend (e.g. node index) and a more informative title (e.g. Eigenvector alignment with nodes), a caption (what is the take-home message?), and y-axis labels (what is the 'value'?).*

Thank you for this helpful suggestion. We have revised the Figure 6 taking your suggestion into account. The new figure includes a clearer title, axis labels, and an informative caption that highlights the key take-home message. See Results - Main Manuscript

> *(7) Page 11, paragraph 1: The sentence "...(as discussed in Section)" is missing a section reference.*

Thank you for pointing this out. We have replaced the section reference placeholder with the equation reference to Equation (9), which is the relevant theoretical result.

> *(8) Page 13, caption of Figure 8c: compputed -> computed.*

We have corrected this typo. We also reviewed the manuscript for any similar typographical errors and corrected them.

> *(9) Page 14 Figure 9: The y-axis label for "Controlled State Process" plots is missing.*

Thank you for catching this. The y-axis was hidden by other elements in the figure. We have revised the figure layout to ensure that the y-axis label is visible.

> *(10) Page 16: The sentence "As described in Section, a preliminary ... " is missing a section reference*

Thank you for pointing this out. We have corrected the missing reference. The sentence now reads “as shown in Fig. 10” rather than the incomplete “as described in Section.”

> *(11) Page 16, Fig. 10c: It looks like the errors have inconsistent color ranges. A shared colorbar would help.*

Thank you for your suggestion. We have revised Figure 10c to use a shared colorbar across all subplots.

> *(12) Page 17, Paragraph preceding eq. 16: "the eigenvectors of X" --> "the eigenvectors of Sigma_X".*

Thank you for catching this. We have revised the text. See Methods - Main Manuscript.

> *(13) S.1: This is a GLS, not an OLS estimator, if this assumes full noise covariance.*

As stated in our response to the Weakness above, we assume the noise term $\xi(t)$ is i.i.d. with covariance matrix Σ_ξ, and the estimator we analyze is the OLS estimator. See Theoretical Details - Supplementary Material.

> *(14) S.2: Unclear that ⊗ is the Kronecker product (not outer), as it is not defined.*

We thank the reviewer for pointing out this ambiguity. In the revised Supplementary Material, we have explicitly explained the Kronecker product with a reference to Hamilton [10, Appendix A.4, p. 732]. See Theoretical Details - Supplementary Material.

> *(15) S.51: There is an accidental comma between alpha and A after "xdiff(t) ="*

Thank you for pointing this out. We have removed the accidental comma. See Theoretical Details - Supplementary Material.

> *(16) Section B.2: It would be useful to have the simulation details for that section (as is given for the other section in B.1).*

We thank the reviewer for this helpful suggestion. We have added a dedicated “Simulation details” paragraph to Appendix B.5 (the section containing Fig. B.4) so that the setup is now described with the same level of specificity as the other appendix sections. See Experimental Conditions and Results - Supplementary Material.

References

(1) Irma N Angulo-Sherman, Marisol Rodríguez-Ugarte, Nadia Sciacca, Eduardo Iáñez, and José M Azorín. Effect of tDCS stimulation of motor cortex and cerebellum on EEG classification of motor imagery and sensorimotor band power. J. Neuroeng. Rehabil., 14(1):31, April 2017.

(2) Hassan Arbabi and I Mezić. Computation of transient koopman spectrum using hankeldynamic mode decompoisition. APS, page G1.009, November 2017.

(3) Steven L Brunton, Bingni W Brunton, Joshua L Proctor, Eurika Kaiser, and J Nathan Kutz. Chaos as an intermittently forced linear system. Nat. Commun., 8(1):19, May 2017.

(4) Adenauer G Casali, Olivia Gosseries, Mario Rosanova, Mélanie Boly, Simone Sarasso, Karina R Casali, Silvia Casarotto, Marie-Aurélie Bruno, Steven Laureys, Giulio Tononi, and Marcello Massimini. A theoretically based index of consciousness independent of sensory processing and behavior. Sci. Transl. Med., 5(198), August 2013.

(5) Karl Deisseroth. Optogenetics. Nat. Methods, 8(1):26–29, January 2011.

(6) Shikuang Deng, Jingwei Li, B T Thomas Yeo, and Shi Gu. Control theory illustrates the energy efficiency in the dynamic reconfiguration of functional connectivity. Commun. Biol., 5(1):295, April 2022.

(7) Shrey Grover, Renata Fayzullina, Breanna M Bullard, Victoria Levina, and Robert M G Reinhart. A meta-analysis suggests that tACS improves cognition in healthy, aging, and psychiatric populations. Sci. Transl. Med., 15(697):eabo2044, May 2023.

(8) Shi Gu, Fabio Pasqualetti, Matthew Cieslak, Qawi K Telesford, Alfred B Yu, Ari E Kahn, John D Medaglia, Jean M Vettel, Michael B Miller, Scott T Grafton, and Danielle S Bassett. Controllability of structural brain networks. Nat. Commun., 6:8414, October 2015.

(9) Mark Hallett, Riccardo Di Iorio, Paolo Maria Rossini, Jung E Park, Robert Chen, Pablo Celnik, Antonio P Strafella, Hideyuki Matsumoto, and Yoshikazu Ugawa. Contribution of transcranial magnetic stimulation to assessment of brain connectivity and networks. Clin. Neurophysiol., 128(11):2125–2139, November 2017.

(10) James Douglas Hamilton. Time Series Analysis. Princeton University Press, Princeton, 1994.

(11) Ann Huang, Mitchell Ostrow, Satpreet H Singh, Leo Kozachkov, Ila Fiete, and Kanaka Rajan. InputDSA: Demixing then comparing recurrent and externally driven dynamics. arXiv [q-bio.NC], November 2025.

(12) Shunsuke Kamiya, Genji Kawakita, Shuntaro Sasai, Jun Kitazono, and Masafumi Oizumi. Optimal control costs of brain state transitions in linear stochastic systems. J. Neurosci., 43(2):270–281, January 2023.

(13) Teresa M Karrer, Jason Z Kim, Jennifer Stiso, Ari E Kahn, Fabio Pasqualetti, Ute Habel, and Danielle S Bassett. A practical guide to methodological considerations in the controllability of structural brain networks. J. Neural Eng., 17(2):026031, April 2020.

(14) Genji Kawakita, Shunsuke Kamiya, Shuntaro Sasai, Jun Kitazono, and Masafumi Oizumi. Quantifying brain state transition cost via schrödinger bridge. Netw. Neurosci., 6(1):118– 134, February 2022.

(15) Hassan K Khalil. Nonlinear systems. Prentice-Hall, Upper Saddle River, NJ, 2002.

(16) Paul K LaFosse, Zhishang Zhou, Jonathan F O'Rawe, Nina G Friedman, Victoria M Scott, Yanting Deng, and Mark H Histed. Single-cell optogenetics reveals attenuationby-suppression in visual cortical neurons. bioRxivorg, page 2023.09.13.557650, May 2024.

(17) Andres M Lozano, Nir Lipsman, Hagai Bergman, Peter Brown, Stephan Chabardes, Jin Woo Chang, Keith Matthews, Cameron C McIntyre, Thomas E Schlaepfer, Michael Schulder, Yasin Temel, Jens Volkmann, and Joachim K Krauss. Deep brain stimulation: current challenges and future directions. Nat. Rev. Neurol., 15(3):148–160, March 2019.

(18) Yuki Minai, Matthew Smith, Joana Soldado-Magraner, and Byron Yu. MiSO: Optimizing brain stimulation to create neural activity states. In A Globerson, L Mackey, D Belgrave, A Fan, U Paquet, J Tomczak, and C Zhang, editors, Advances in Neural Information Processing Systems 37, volume 37, pages 24126–24149, San Diego, California, USA, 2024. Neural Information Processing Systems Foundation, Inc. (NeurIPS).

(19) Davide Momi, Zheng Wang, and John D Griffiths. TMS-evoked responses are driven by recurrent large-scale network dynamics. Elife, 12(e83232), April 2023.

(20) Ali Moradi Amani, Amirhessam Tahmassebi, Andreas Stadlbauer, Uwe Meyer-Baese, Vincent Noblet, Frederic Blanc, Hagen Malberg, and Anke Meyer-Baese. Controllability of

functional and structural brain networks. Complexity, 2024(1), January 2024.

(21) Norman S Nise. Control Systems Engineering. John Wiley & Sons, 8 edition, 2020.

(22) Katsuhiko Ogata. Modern Control Engineering. Prentice Hall, 2010.

(23) Mitchell Ostrow, Adam Eisen, and Ila Fiete. Delay embedding theory of neural sequence models. arXiv [cs.LG], June 2024.

(24) Yumi Shikauchi, Mitsuaki Takemi, Leo Tomasevic, Jun Kitazono, Hartwig R Siebner, and Masafumi Oizumi. Quantifying state-dependent control properties of brain dynamics from perturbation responses. J. Neurosci., page e0364252025, December 2025.

(25) Floris Takens. Detecting strange attractors in turbulence. In David Rand and Lai-SangYoung, editors, Dynamical Systems and Turbulence, Warwick 1980, volume 898 of Lecture Notes in Mathematics, pages 366–381. Springer, Berlin, Heidelberg, 1981.

(26) Liam C Tapsell, Matheus D Pinto, Ann-Maree Vallence, Casey Whife, Maria Luciana Perez Armendariz, Shaswat Senger, Jack Andringa-Bate, Dana Hince, and Myles C Murphy. What are the optimal transcranial direct current stimulation parameters and design elements to modulate corticospinal excitability? a systematic review and longitudinal meta-analysis. Neurol. Res. Pract., 7(1):86, November 2025.

(27) Lei Tong, Shanshan Han, Yao Xue, Minggang Chen, Fuyi Chen, Wei Ke, Yousheng Shu, Ning Ding, Joerg Bewersdorf, Z Jimmy Zhou, Peng Yuan, and Jaime Grutzendler. Single cell in vivo optogenetic stimulation by two-photon excitation fluorescence transfer. iScience, 26(10):107857, October 2023.

(28) Andrew Wagenmaker, Lu Mi, Marton Rozsa, Matthew S Bull, Karel Svoboda, Kayvon Daie, Matthew D Golub, and Kevin Jamieson. Active learning of neural population dynamics using two-photon holographic optogenetics. Adv. Neural Inf. Process. Syst., 37:31659–31687, 2024.

https://doi.org/10.7554/eLife.110030.2.sa0