



**HAL**  
open science

# Delayed closed-loop neurostimulation for the treatment of pathological brain rhythms in mental disorders: a computational study

Thomas Wahl, Joséphine Riedinger, Michel Duprez, Axel Hutt

► **To cite this version:**

Thomas Wahl, Joséphine Riedinger, Michel Duprez, Axel Hutt. Delayed closed-loop neurostimulation for the treatment of pathological brain rhythms in mental disorders: a computational study. *Frontiers in Neuroscience*, 2023, 10.3389/fnins.2023.1183670 . hal-04037276v4

**HAL Id: hal-04037276**

**<https://hal.science/hal-04037276v4>**

Submitted on 5 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Delayed closed-loop neurostimulation for the treatment of pathological brain rhythms in mental disorders: a computational study

Thomas Wahl\*, Joséphine Riedinger\*,<sup>†</sup>, Michel Duprez\*, Axel Hutt\*

July 5, 2023

## Abstract

Mental disorders are among the top most demanding challenges in world-wide health. A large number of mental disorders exhibit pathological rhythms, which serve as the disorders characteristic biomarkers. These rhythms are the targets for neurostimulation techniques. Open-loop neurostimulation employs stimulation protocols, which are rather independent of the patients health and brain state in the moment of treatment. Most alternative closed-loop stimulation protocols consider real-time brain activity observations but appear as adaptive open-loop protocols, where e.g. pre-defined stimulation sets in if observations fulfil pre-defined criteria. The present theoretical work proposes a fully-adaptive closed-loop neurostimulation setup, that tunes the brain activities power spectral density (PSD) according to a user-defined PSD. The utilized brain model is non-parametric and estimated from the observations via magnitude fitting in a pre-stimulus setup phase. Moreover, the algorithm takes into account possible conduction delays in the feedback connection between observation and stimulation electrode. All involved features are illustrated on pathological  $\alpha$ - and  $\gamma$ -rhythms known from psychosis. To this end, we simulate numerically a linear neural population brain model and a non-linear cortico-thalamic feedback loop model recently derived to explain brain activity in psychosis.

**Keywords:** neurostimulation, closed-loop, control, real-time, delay, EEG

## 1 Introduction

Electrical neurostimulation is an old human idea, and has been a well-established therapy for mental disorders for few decades. Caius Plinius during Antiquity and Scribonius Largus, who lived in the first century AD, proposed respectively contacts with the Electric ray (Torpedo Fish) for the treatment of post-partum pain and severe headaches. In the 19th century, electrical stimulation was commonly prescribed by neurologists for nervous disease [1]. Today, various electrical stimulation techniques exist to modulate neuronal systems and novel techniques for an optimal clinical treatment of a specific pathology gain more and more attention [2, 3]. They could be used as an additional therapeutic lever or as an alternative to pharmacological medication, thus representing a hope for pharmaco-resistant forms of disease.

Brain oscillations result from coordinated electrical neuronal tissues activity within and between structures and networks. Implicated in various neural processes, such as perception, attention and cognition, their disruption yields pathological rhythms, which reflect abnormal activity of the implicated brain network, notably at the cellular and molecular level [4]. These pathological rhythms serve as good biomarkers for neuropathologies. For instance, neurophysiological studies have revealed that a large number of mental disorders exhibit pathological rhythms, which do not occur in healthy patients [5]. Neurostimulation techniques have identified such

---

\*ICube, MLMS, University of Strasbourg; MIMESIS Team, Inria Nancy - Grand Est, Strasbourg, France

<sup>†</sup>INSERM U1114, Neuropsychologie cognitive et physiopathologie de la schizophrénie, Strasbourg, France

pathological rhythms as good stimulation targets for the treatment of brain oscillatory disorders. Neurostimulation induces electric currents in neuronal tissue. Depending on the stimulation protocol, i.e. the temporal stimulation current shape, its duration and pause and the number of repetitions, neurostimulation can lead to neural plasticity effects or to pacemaker-like brain stimulation, respectively.

For example, Deep Brain Stimulation (DBS) is an invasive technique and proposed for patients suffering from severe pharmaco-resistant Parkinson’s disease (PD) or obsessive-compulsive disorders. In PD patients aberrant hypersynchronicity and hyperactivity in the  $\beta$ -frequency band (12-30 Hz) of the basal ganglia-thalamocortical network can be addressed by the pharmacological medication (e.g. Levodopa) or DBS. The conventional DBS protocols focus on the subthalamic nucleus or globus pallidus stimulation continuously at a temporally constant frequency about 130 Hz. The suppression of the pathological beta oscillations was correlated with improving motor symptoms [6]. Recent techniques [7, 8] propose to apply an adaptive closed-loop stimulation protocol based on observed intracranial brain activity. In addition to this intracranial neurostimulation technique, transcranial electrical stimulation (TES) and transcranial magnetic stimulation (TMS) are non-invasive neuromodulation approaches in which, respectively, a low electrical current and a magnetic field are applied to the cortical tissues. The TES current modalities include direct currents (tDCS), i.e. constant currents, alternating current (tACS), i.e. typically oscillatory currents, and random noise-shape currents (tRNS), which typically includes frequencies above the  $\beta$ -frequency band. It was shown that tDCS can improve cognitive performance in healthy subjects [9] and patients [10] and it is applied as a therapeutic means to target brain network dysfunctions, such as Attention-Deficit/Hyperactivity Disorder [11] and major depressive disorder [12].

Although the neurostimulation techniques mentioned above may permit to alleviate mental disorder patients from symptoms, the success rate of these treatments is still limited [13]. This underperformance results from non-optimal choices of the stimulation protocol originating from the lack of understanding of the underlying neural response to stimulations and the non-patient specific stimulation protocol. In other words, typically, the stimulation protocol (including size, duration, repetition cycle of the stimulation signal) is open-loop, i.e. pre-defined by the clinician based on heuristic criteria before the stimulation starts [14]. This non-optimal approach is inferior to so-called closed-loop techniques, which automatically adapt to the patients current brain/health state. Such an adaptive, or closed-loop, approach has been introduced for intracranial [15, 16, 17] and transcranial stimulation [18] and has been shown to improve neurostimulation in major depression patients [19], epilepsy [20] and affective and anxiety disorders [21]. These proposed closed-loop methods are adaptive in the sense that a pre-defined stimulation signal is applied when observed brain activity fulfills certain criteria, such as passing an amplitude or power threshold. While this adaptive approach improves existing open-loop methods, the pre-defined stimulation signal may still be non-optimally chosen. Recently proposed methods produce better results by using reference signal tracking control schemes such as proportional integral (PI) controller [22, 23, 24, 25], linear quadratic regulator (LQR) [26, 27, 28] or predictive model control (MPC) [29] and [30] which uses an LQR in a MPC framework. However, this form of control requires to pre-define a reference signal which is often non-trivial to provide in a patient-specific manner. Furthermore, because of the stochastic nature of brain signals, forcing the resting state signal to follow a reference signal with its own independent noise creates an unnecessary constraint for the stimulation signal when we only want to regulate the power in given frequency bands.

We propose to estimate a stimulation signal on the basis of observed brain activity without the need to track a reference signal. The target stimulation signal is computed directly via a linear controller synthesized using a user-defined filter that encodes the desired frequency-domain modifications. We argue that it is the natural choice for a closed-loop optimization in the presence of pathological rhythms: typically, the pathology is identified by an abnormal power in a certain frequency band and the closed-loop control aims to modify this power value in such a way that the final brain activity power spectral distribution resembles the distribution of a healthy subject. Examples are pathological too strong  $\beta$ -rhythm magnitudes in Parkinson disease [31] and too weak  $\alpha$ -rhythm [32] and too strong  $\gamma$ -rhythm in schizophrenia [33] This approach implies the hypothesis that modifying the observed pathological brain rhythms of a patient to resemble brain rhythms of a healthy subject renders the patients brain state and improves the patients health situation. This assumption was motivated by

the impressive improving impact of DBS in psychiatric disorders [34].

Technically, the proposed method aims to reshape the spectral distribution of observed data, such as electroencephalographic data (EEG). For illustration, we consider pathological brain rhythms observed in psychosis in the  $\alpha$ - [32] and  $\gamma$ -band [35]. Our method relies on the extraction and the filtering in real-time of the brain resting state activity signal, using the EEG and an estimated brain response model. The underlying brain model is fully non-parametric and estimated from observed resting state EEG. Moreover, we consider the fact that the closed-loop feedback exhibits a certain conduction delay between measurement and stimulation. This conduction delay results from the transmission delay in the hardware and the numerical computation time of the stimulation signal. Very first estimates of this delay time are in the range of a few tens of milliseconds [Private communication, Isope, 2020], i.e. in the range of EEG signal time scales. Consequently, the present feedback delay in real-world systems may affect the methods performance. [30] presents a method to increase the robustness of an adaptive closed-loop controller against delay by reducing the sensitivity of the closed-loop to high-frequency disturbances. However, while this generally decreases the error in the closed-loop output, this also prevents to apply the control signal specifically to high-frequency ranges like the  $\gamma$ -range. To our best knowledge, the present study is the first providing a method to effectively compensate the frequency-domain errors created by the feedback delay in closed-loop neurostimulation systems without sacrificing the controllability of high-frequency ranges.

The remaining article is organized as follows : Section 2 presents the neurostimulation setup and the closed-loop circuit studied in the rest of this paper. Then, we propose a model-based controller design to apply desired modifications to the observed activity signal. Subsequently, we propose a model estimation method to extract the brain input response model needed for the controller design. Later, we address the problem of the closed-loop delay by designing an additional system to approximate the future values of the observations. Finally, we present two brain models, which illustrate and validate the proposed method. Then, Section 3 presents the simulation results of our circuits, including the accuracy of the model estimation step and the delay compensation. Lastly, in section 4, we discuss the results of the method presented in the paper compared to the state of the art, mention limitations and pinpoint some perspectives and possible experimental tests.

## 2 Material and methods

### 2.1 Simulated neurostimulation

We build a theoretical plant as a circuit containing a stimulation element and an observation element, both connected to the model brain system under study. In real practice, the stimulation element corresponds to the neurostimulation device, such as a TES system or a TMS coil. In contrast, the observation element may represent electro-/magneto-encephalographic electrodes (in the following called EEG) or electrodes observing Local Field Potential. We define the time-dependent functions  $u : \mathbb{R} \rightarrow \mathbb{R}$  and  $y : \mathbb{R} \rightarrow \mathbb{R}$  as the input stimulation current and the output EEG signal, respectively.

If no input current is applied, the output is a non-zero stochastic signal  $y_0$  corresponding to the measured resting state EEG activity and a non-zero neurostimulation current alters the output signal as a linear response. This alteration is caused by a change in the brain activity in response to the neurostimulation input and a direct measurement of the input current. The latter is undesirable as it is not correlated with brain dynamics but only with neurostimulation and measurement devices. In the following, we assume that observations include brain dynamics correlated output only while direct current measurements are filtered out. A method to remove the direct current measurement from the EEG signal is discussed in Section 4.

Then, we define the plant  $\mathcal{P}$  as the system that takes  $u$  as its input and generates an output  $y$ , which is equal to  $y_0$  when no input is applied. By modeling the dynamics of  $\mathcal{P}$ , our goal is a neurostimulation signal  $u$  that causes predetermined changes in the spectral power amplitude of the output signal  $y$ . In our case, the goal is to increase the activity in the alpha band (8 – 12Hz) and decrease the activity in the gamma band (25 – 55Hz) motivated by aberrant power spectrum magnitudes in schizophrenia [31, 32]. A possible experimental setup

involving our method is sketched in Fig. 1A).

## 2.2 Linear time invariant model

We assume that the observed output response to a small neurostimulation input  $u$  is linear and time-invariant (LTI). This assumption is supported by multiple results across literature [36, 37, 38]. Thus, there is an underlying LTI system  $\mathcal{G}$  that produces an output  $y_u$  for any given input  $u$ . For this system, we can define a function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , which is the output produced by the plant input response system  $\mathcal{G}$  in response to a unit impulse signal  $\delta(t)$  where  $t \in \mathbb{R}$  is the time elapsed since the start of the signal. This function  $g$  is also called the unit impulse response of  $\mathcal{G}$  and we have

$$y_u(t) = g(t) * u(t) := \int_{-\infty}^{+\infty} g(t')u(t-t')dt'.$$

with time  $t$  and  $*$  denotes the convolution over time. It leads to the total plant output

$$y(t) = y_0(t) + y_u(t) = y_0(t) + g(t) * u(t). \quad (1)$$

With this choice of model, the contribution of the neurostimulation response to the total output is purely additive, allowing us to focus the analysis on  $\mathcal{G}$ , which represents the neurostimulation response part of the plant system. We also see that  $y_0$ , the resting state activity, contains the stochastic part of the output, while  $y_u$  can be predicted for any known input signal  $u$  if we have a model for the system  $\mathcal{G}$ . A method to estimate the plant input response model  $\mathcal{G}$  is presented in section 2.4.

## 2.3 Closed-loop control

In this section, we suppose that the function  $g$  is known. The estimation of  $g$  will be the aim of section 2.4.

To close the loop, we generate the plant input signal  $u$  as the output of a linear controller  $\mathcal{K}$  in response to the plant output  $y$

$$u(t) = k(t) * y(t),$$

where  $k : \mathbb{R} \rightarrow \mathbb{R}$  is the unit impulse response of the controller  $\mathcal{K}$ . We can now rewrite Eq. (1) as

$$y(t) = y_0(t) + g(t) * k(t) * y(t). \quad (2)$$

Here, we assume that no delay between observation and stimulation application is present. We will relax this condition in section 2.5. To solve Eq. (2), we apply the Laplace transform defined for each time-dependent function  $x : \mathbb{R} \rightarrow \mathbb{R}$  by

$$X(s) = \mathcal{L}\{x(t)\}(s) := \int_{0^-}^{+\infty} x(t)e^{-st}dt, \quad (3)$$

Thus, we define  $Y : \mathbb{C} \rightarrow \mathbb{C}$ ,  $Y_0 : \mathbb{C} \rightarrow \mathbb{C}$ ,  $G : \mathbb{C} \rightarrow \mathbb{C}$  and  $K : \mathbb{C} \rightarrow \mathbb{C}$  as the Laplace transforms of respectively  $y$ ,  $y_0$ ,  $g$  and  $k$ , allowing us to write Eq. (2) as

$$Y(s) = Y_0(s) + G(s)K(s)Y(s).$$

Hence

$$Y(s) = \frac{1}{1 - G(s)K(s)}Y_0(s). \quad (4)$$

We now have an equation for the closed-loop output in function of the resting state activity. A block diagram of the closed-loop circuit is shown in Fig. 1B). Hence to design the frequency distribution of  $y$  we tune the frequency distribution of the transfer function  $K$  of the controller  $\mathcal{K}$

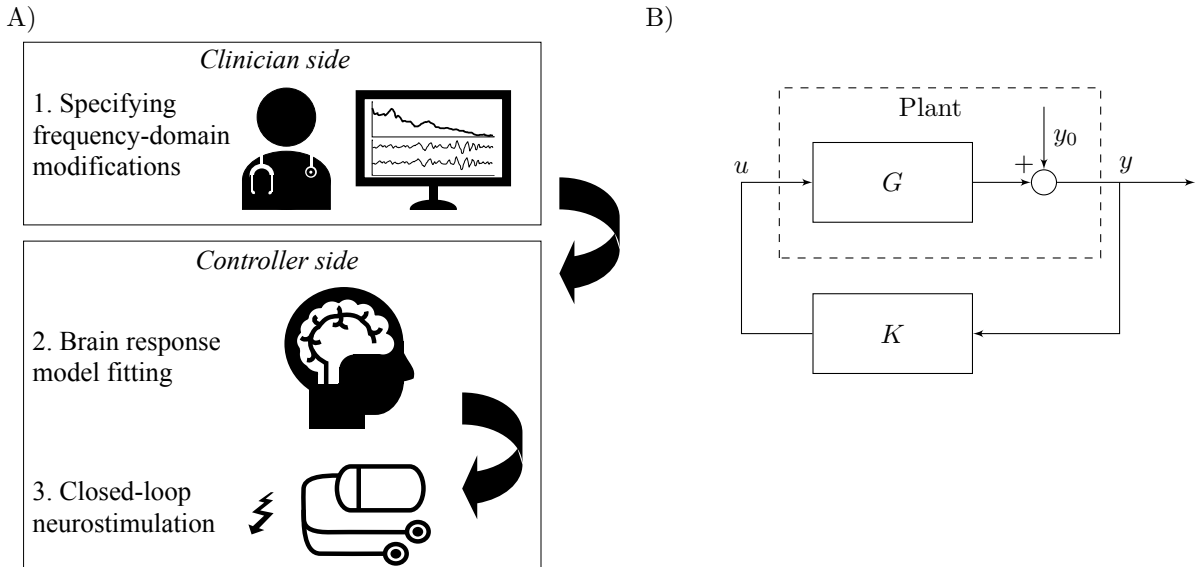


Figure 1: **Closed-loop neurostimulation setup.** **A)** The different steps of our method in a possible experimental setup. First, the clinician specifies the desired frequency domain modifications to apply e.g. increase  $\alpha$ -activity and decrease  $\gamma$ -activity. Second, the brain input response model is fitted to the measured EEG signal to match as closely as possible the current brain dynamics of the patient. Finally, the clinician’s specifications and the fitted model are used to synthesize the closed-loop controller  $\mathcal{K}$  that is directly used for closed-loop neurostimulation. **B)** The controller directly produces the stimulation signal  $u$  in function of the measured total brain output  $y$  in real-time.

### Controller synthesis

Our closed-loop setup aims to tune the observation power spectrum, or equivalently, the choice of  $Y(s)$  subjected to the resting state  $Y_0(s)$ . To this end, we define a linear filter  $\mathcal{H}$  with transfer function  $H : \mathbb{C} \rightarrow \mathbb{C}$  and

$$Y(s) = Y_0(s) + H(s)Y_0(s). \quad (5)$$

Specifically, we intend to restore the physiological state of the brain, e.g. of a schizophrenic patient as our motivation, with an observed EEG presenting low alpha activity and high gamma activity. The chosen filter  $\mathcal{H}$  is a weighted double bandpass filter with positive weight in the  $\alpha$ -frequency band to increase  $\alpha$ -power and negative weights in the  $\gamma$ -band to decrease the systems  $\gamma$ -activity. The filter’s transfer function is defined as

$$H(s) = c_1 \frac{2\pi B_1 s}{s^2 + 2\pi B_1 s + (2\pi f_1)^2} + c_2 \frac{2\pi B_2 s}{s^2 + 2\pi B_2 s + (2\pi f_2)^2}.$$

The exact parameters of  $\mathcal{H}$  are shown in table 1.

We can synthesize the closed-loop controller  $\mathcal{K}$ , by combining equations (4) and (5) and solving for  $K$  as

$$\begin{aligned} \frac{1}{1 - G(s)K(s)} Y_0(s) &= Y_0(s) + H(s)Y_0(s) \\ K(s) &= \frac{H(s)}{(1 + H(s))G(s)}. \end{aligned} \quad (6)$$

Therefore, if we know the plant input response transfer function  $G$ , we can find that desired controller transfer function  $K$  by Eq. (6). Once the transfer function is obtained, we can use it to find a corresponding state-space representation [39] for time domain simulations. The state-space system’s ordinary differential equations can then be implemented by any device that can measure the brain activity and produce a custom neurostimulation signal in real-time.

parameter	description	value
$f_1$	$\alpha$ -band natural frequency	10ms
$B_1$	$\alpha$ -band width	4Hz
$c_1$	$\alpha$ -band weight	1.0
$f_2$	$\gamma$ -band natural frequency	40ms
$B_2$	$\gamma$ -band width	30Hz
$c_2$	$\gamma$ -band weight	-0.5

Table 1: **Parameter set of the filter  $\mathcal{H}$ .** The frequency parameters are chosen based on the alpha frequency range (8-12Hz) and the gamma frequency range (25-55Hz) in an EEG. The weighting parameters  $c_1$  and  $c_2$ , respectively positive and negative, corresponding to the choice to increase the alpha activity and decrease the gamma activity.

## 2.4 Model estimation

The design of our closed-loop controller requires estimating the plant input response system  $\mathcal{G}$ , which in practice includes the brain dynamics, the neurostimulation device and the observation device. Our approach includes the estimation of  $\mathcal{G}$  directly from observed brain activity, such as EEG of the patient. This ensure that the estimated plant model will be as close as possible to the real brain dynamics in the corresponding experimental conditions. To this end, we first need to find a way to measure the plant input response without also measuring the plant resting state activity. This is not trivial since the observed signal is the sum of the resting state activity and the stimulation response.

### Signal extraction

Let us consider an open-loop setup with an arbitrary input  $u$  applied to the plant, which generates the output described by Eq. (1). In this equation, we only know  $u$  and  $y$ , and want to estimate the impulse response  $g$ . The problem is that we cannot observe  $y_0$  only during the stimulation. Hence, based on previous data recordings, we need to find a way to predict the dynamics of  $y_0$  during the stimulation.

First, we provide the following standard definitions that are important in the subsequent discussion. For any time domain signal  $x : \mathbb{R} \rightarrow \mathbb{R}$ , we denote the Fourier transform by

$$\hat{x}(f) = \mathcal{F}\{x(t)\}(f) := \int_{-\infty}^{\infty} x(t)e^{-2\pi ift} dt. \quad (7)$$

We define  $\alpha_0 : \mathbb{R} \rightarrow \mathbb{R}$  and  $\alpha_u : \mathbb{R} \rightarrow \mathbb{R}$  such as  $\alpha_0(t) = y_0(t) - \bar{y}_0$  and  $\alpha_u(t) = y_u(t) - \bar{y}_u$  where  $\bar{y}$ ,  $\bar{y}_0$  and  $\bar{y}_u$  are respectively the ensemble means of  $y$ ,  $y_0$  and  $y_u$ .

We assume that  $y_0$  is a wide-sense-stationary (WSS) random process, i.e. its mean and variance do not depend on time. According to the Wiener-Khinchin theorem [40, 41], the autocorrelation function of a wide-sense-stationary random process has a spectral decomposition given by the power spectrum of that process

$$S_{yy}(f) = |\hat{\alpha}(f)|^2,$$

where  $\hat{\alpha} : \mathbb{R} \rightarrow \mathbb{C}$  is the Fourier transform of  $\alpha(t) = y(t) - \bar{y} \in \mathbb{R}$  and  $S_{yy} : \mathbb{R} \rightarrow \mathbb{R}^+$  is the spectral density of  $y$ .

Then, we can write Eq. (1) as

$$\bar{y} + \alpha(t) = \bar{y}_0 + \alpha_0(t) + \bar{y}_u + \alpha_u(t),$$

where  $\bar{y} = \bar{y}_0 + \bar{y}_u$ . The equation then simplifies to

$$\alpha(t) = \alpha_0(t) + \alpha_u(t).$$

By application of the Fourier transform, we obtain

$$\hat{\alpha}(f) = \hat{\alpha}_0(f) + \hat{\alpha}_u(f)$$

and

$$|\hat{\alpha}(f)|^2 = |\hat{\alpha}_0(f)|^2 + |\hat{\alpha}_u(f)|^2 + 2\text{Re}[\hat{\alpha}_0(f)\hat{\alpha}_u(f)^*].$$

In the following, we compute the ensemble average of each term of this equation. Since  $\alpha$  and  $\alpha_u$  are two independent processes sampled at different times and  $\langle \hat{\alpha}_0 \rangle = \langle \hat{\alpha}_u \rangle = 0$ .

Hence

$$\langle 2\text{Re}(\hat{\alpha}_0(f)\hat{\alpha}_u(f)^*) \rangle = 2\text{Re}[\langle \hat{\alpha}_0(f)\hat{\alpha}_u(f)^* \rangle] = 0.$$

Here and in the following,  $\langle \cdot \rangle$  denotes the ensemble average. We point out that although Eq. (8) does hold when considering the ensemble average of the signals, fluctuations around 0 still remain in Eq. (8) for finite ensemble number of finite time signals.

We define the amplitude ratio (AR) as

$$\text{AR} = \left\langle \frac{|\hat{\alpha}(f)|^2}{|\hat{\alpha}_0(f)|^2} \right\rangle, \quad (8)$$

which quantifies the amplitude gain between the resting-state output and the stimulated output. The fluctuations mentioned above can also be reduced by increasing the input current leading to a higher AR and therefore, a higher contribution to the total signal of the  $\hat{\alpha}_u$  term (which is known) compared to the other terms.

Nevertheless, this yields

$$\langle |\hat{\alpha}_u(f)|^2 \rangle = \langle |\hat{\alpha}(f)|^2 \rangle - \langle |\hat{\alpha}_0(f)|^2 \rangle. \quad (9)$$

Using Eq. (1), we can express  $\hat{\alpha}_u$  in terms of the input impulse response  $g$  and the input  $u$

$$\begin{aligned} \hat{\alpha}_u(f) &= \mathcal{F}\{y_u(t) - \bar{y}_u\}(f) \\ &= \mathcal{F}\{g(t) * [u(t) - \bar{u}]\}(f) \\ &= \hat{g}(f)\mathcal{F}\{u(t) - \bar{u}\}(f). \end{aligned} \quad (10)$$

This equation permits to estimate the transfer function  $\hat{g}$ , see Section 3.

To express the transfer function  $\hat{g}$  in Laplace space, we use the fact that a unit impulse response function is non-zero only for positive time values  $t$ . Hence, based on equations (3) and (7), for  $s = 2\pi if$ , we can write the Laplace transform  $G$  as

$$G(2\pi if) = \int_{0^-}^{+\infty} g(t)e^{-2\pi ift} dt = \int_{-\infty}^{+\infty} g(t)e^{-2\pi ift} dt = \hat{g}(f),$$

where the unit impulse response function  $g$  directly relates the output  $y$  to the resting-state output  $y_0$  and the stimulation signal  $u$ , cf. Eq. (1).

We now need a method to generate a LTI system with a transfer function that matches the magnitude data computed with the formula. This is achieved by the magnitude vector fitting algorithm.

### Magnitude vector fitting

Our goal is now to find a transfer function  $G$  corresponding the magnitude data  $|\hat{g}(f)|^2$ . For this purpose, we use a variant of the vector fitting algorithm design to work even with only the magnitude data. This method is called magnitude vector fitting [42].



It allows to fit a passive LTI system to data by fitting the model transfer function. The system is synthesized such that the mean square error between the magnitude data sample and the transfer function evaluated at the same frequency points is minimized. [42] show that the transfer function of the fitted model reproduces both the magnitude and the phase shift of the original transfer function, although the fitting has been performed using sampled magnitude data only.

By minimizing the mean square error, the algorithm ensures that the transfer function of the fitted model accurately matches the original model as represented by the reconstructed gain data. Furthermore, to assess the accuracy of the reconstruction, we also compare the fitted model to the transfer function of the linearized brain model used for the simulation. This allows to double-check the validity of the reconstructed magnitude and also to verify if the reconstructed phase fits the phase of the original model as closely as possible cf. Fig 3C)D).

We define the root-mean-square error (RMSE) as

$$\text{RMSE} = \sqrt{\left\langle \left| \frac{\tilde{G}(2\pi i f) - G(2\pi i f)}{G(2\pi i f)} \right|^2 \right\rangle}, \quad (11)$$

where  $\tilde{G}$  is the fitted model's transfer function,  $G$  is the original model's transfer function and  $f \in \mathbb{R}^+$  are the frequency points used for the fitting. This allows us to quantify the accuracy of the fitting step.

## 2.5 Delay compensation

Realistic feedback loops exhibit conduction delays between the moment of observation and feedback stimulation. Reasons for such delays are finite conduction speeds in cables, electronic switches, interfaces and delays caused by the controller device to compute numerically adapted stimuli. In systems with large time scales, such as controlled mechanical devices on the centimeter or larger scale, such delays may be negligible. Conversely biological systems such as the brain evolve on a millisecond scale and conduction delays may play an important role. Preliminary estimation of input and output devices of desktop computers have revealed an approximate delay of  $\sim 10$ ms. By virtue of such delays, it is important to take them into account in the closed-loop between the moment of observation and stimulation.

The different sources of delay can be represented as plant input and output delays. Since the controller  $\mathcal{K}$  is LTI, the input and output delays can be concatenated into one single plant input delay. Hence, in our setup, we model the delay as an input delay  $\tau$  in the system  $\mathcal{G}$ , modifying  $y(t) = g(t)*u(t)$  in Eq. (1) to  $y(t) = g(t)*u(t-\tau)$ . The Smith predictor [43] [44] is a known method to compensate such delay times. However, in the present problem, this approach allows controlling a limited frequency band only (see Fig. 7A)B)). Consequently, it was necessary to invent another method. Since the plant input  $u$  is generated by the controller  $\mathcal{K}$ , we modify the controller to compensate the delay. To this end, the new controller  $\mathcal{K}$  is chosen to estimate the future value of  $u$  instead of the present value. The new proposed method to apply this controller modification is presented in the Results Section 3.2.

## 2.6 Comparison to the state of the art

Our method is tested against two main control schemes commonly used in adaptive closed-loop neurostimulation. These control schemes are closed-loop control with a PI controller [22, 23, 24, 25] and Linear Quadratic Gaussian (LQG) control [26, 27, 28] which refers to the combination of a Linear Quadratic Estimator (LQE) (or Kalman filter) with a Linear Quadratic Regulator (LQR) [45]. For both these methods, the tracked reference signal is generated from pre-recorded pathological resting state activities to which we apply a filter restoring the target  $\alpha$ - and  $\gamma$ -activities. In order to prevent closed-loop destabilization and regulate high frequency disturbances, a Smith predictor [43] is used for the PI and LQG controller to compensate the delay, while we use our own delay compensation method for our controller.

parameter	description	value
$\tau_{e,1,2}$	exc. synaptic time constant	5ms
$\tau_{i,1,2}$	inhib. synaptic time constant	20ms
$N_{11}$	first exc. linear coefficient	1.15
$N_{21}$	first inhib. linear coefficient	0.63
$N_{12}$	second exc. linear coefficient	2.52
$N_{22}$	second inhib. linear coefficient	6.6
$N$	number of neurons	1000
$\kappa_{1,2}^2$	noises variances (pathological)	$10^{-4}/N$
$(\kappa'_1)^2$	first noise variance (healthy)	$3.6 \cdot 10^{-4}/N$
$(\kappa'_2)^2$	second noise variance (healthy)	$2.5 \cdot 10^{-5}/N$
$b_{1,2}$	input coupling constants	0.18
$b_{3,4}$	input coupling constants	0.14

Table 2: **Parameter set of model (12)**. The choice of parameter is partially based on the paper in which it was developed (see [46]). The difference between the healthy and pathological state is modeled by changes in the amplitude of the finite size fluctuations of each population.

## 2.7 Brain models

Our closed-loop control method works for any LTI brain model. Furthermore, we want to show that it also produces good results on non-linear brain models, for which the neurostimulation input response behaves closely to an LTI system, when the input is sufficiently small. To this end, we present two models used to test our method. The first one is a linear neural population model of cortical activity, and the second one is a non-linear cortico-thalamic neural population model with cortico-thalamic delay.

### 2.7.1 Linear brain model

We describe neural population activity with a noise-driven linear model [46]. The model is composed of two pairs of interacting excitatory and inhibitory populations. Here we have  $V_{e,i}^{(1,2)} : \mathbb{R} \rightarrow \mathbb{R}$ , representing the mean activity of the associated population, where  $V_e^{(1,2)}$  and  $V_i^{(1,2)}$  correspond respectively to excitatory and inhibitory populations. Each population is driven by noise  $\xi_{1,2} : \mathbb{R} \rightarrow \mathbb{R}$  and the external input  $u : \mathbb{R} \rightarrow \mathbb{R}$ , according to the following differential equations:

$$\begin{cases} \tau_{e,1} \frac{dV_e^{(1)}(t)}{dt} &= (-1 + N_{11})V_e^{(1)}(t) - N_{11}V_i^{(1)}(t) + b_1u(t) + \xi_1(t), \\ \tau_{i,1} \frac{dV_i^{(1)}(t)}{dt} &= N_{21}V_e^{(1)}(t) + (-1 - N_{21})V_i^{(1)}(t) + b_2u(t), \\ \tau_{e,2} \frac{dV_e^{(2)}(t)}{dt} &= (-1 + N_{12})V_e^{(2)}(t) - N_{12}V_i^{(2)}(t) + b_3u(t) + \xi_2(t), \\ \tau_{i,2} \frac{dV_i^{(2)}(t)}{dt} &= N_{22}V_e^{(2)}(t) + (-1 - N_{22})V_i^{(2)}(t) + b_4u(t), \end{cases} \quad (12)$$

where the noise  $\xi_{1,2}$  is uncorrelated Gaussian distributed with zero mean and variance  $\kappa_{1,2}^2 = 10^{-7}$ , and the stimulation  $u$  is weighted by the coupling constants  $b_i > 0$  of the corresponding population. In addition,  $\tau_{(e,i),(1,2)}$  are the synaptic time constants of the populations, and constants  $N_{ij} > 0$  are interaction gains of the respective population. Table 2 provides the parameters employed in subsequent simulations.

The observed output

$$y(t) = V_e^{(1)}(t) - V_i^{(1)}(t) + V_e^{(2)}(t) - V_i^{(2)}(t)$$

is a sum of the effective field potential  $V_e^{(j)} - V_i^{(j)}$  of both populations  $j = 1, 2$ , cf. Fig. 2A).

The simulation of the linear brain model in time domain is done using the library `control` of `python`. The numerical integration is computed thanks to matrix exponential [47], with a simulation sampling time of 1ms. The resting state activity of the linear brain model is shown in Fig. 2A).

### 2.7.2 Cortico-thalamic brain model

A different model considers the cortico-thalamic feedback circuit [48]. It describes the cortex layers I-III and the cortico-thalamic loop between cortical layers IV-VI, the thalamic relay cell population and the reticular structure. The cortical layer I-III exhibits mean activity of excitatory cells  $v$  and inhibitory cells  $w$ . Similarly, layer IV-VI exhibits the mean activity  $V_e$  and  $V_i$  and thalamic relay cell populations the mean activity  $V_{th,e}$  and  $V_{th,i}$ . Moreover, the reticular structure has the mean activity  $V_{ret}$ . The fibers between the cortex and thalamus and the cortex and reticular structure exhibit a finite conduction delay  $\tau$  [48, 49]. The 7-dimensional dynamical system of the brain state  $\mathbf{x} = (v, w, V_e, V_i, V_{th,e}, V_{th,i}, V_{ret}) \in \mathbb{R}^7$  obeys

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t), \mathbf{x}(t - \tau)) + \boldsymbol{\xi}(t) + \mathbf{B}u(t), \\ y(t) = \mathbf{C}\mathbf{x}(t), \end{cases} \quad (13)$$

where the superscript  $t$  denotes transposition,  $\mathbf{F} \in \mathbb{R}^7$  is a nonlinear vector function,  $\mathbf{B} \in \mathbb{R}^{7 \times 1}$  is the input coupling matrix and  $\mathbf{C} \in \mathbb{R}^{1 \times 7}$  is the observation matrix. We mention that  $\mathbf{B} = (b_1, b_2, b_3, b_4, 0, 0, 0)^t$ ,  $b_i > 0$ , i.e. only the cortical layers are stimulated with weights  $b_i$ . The observation  $y$  captures the activity of the cortical excitatory populations [48, 50] with  $\mathbf{C} = (c_1, 0, c_3, 0, 0, 0, 0)$ ,  $c_i > 0$ . For more details, please see the Appendix.

The time domain simulations of the cortico-thalamic model is done by numerical integration using the fourth-order Runge-Kutta method implemented by the `scipy` library in `python` with a maximum simulation time step of 1 ms. The resting state activity of the cortico-thalamic brain model is shown in Fig. 2B).

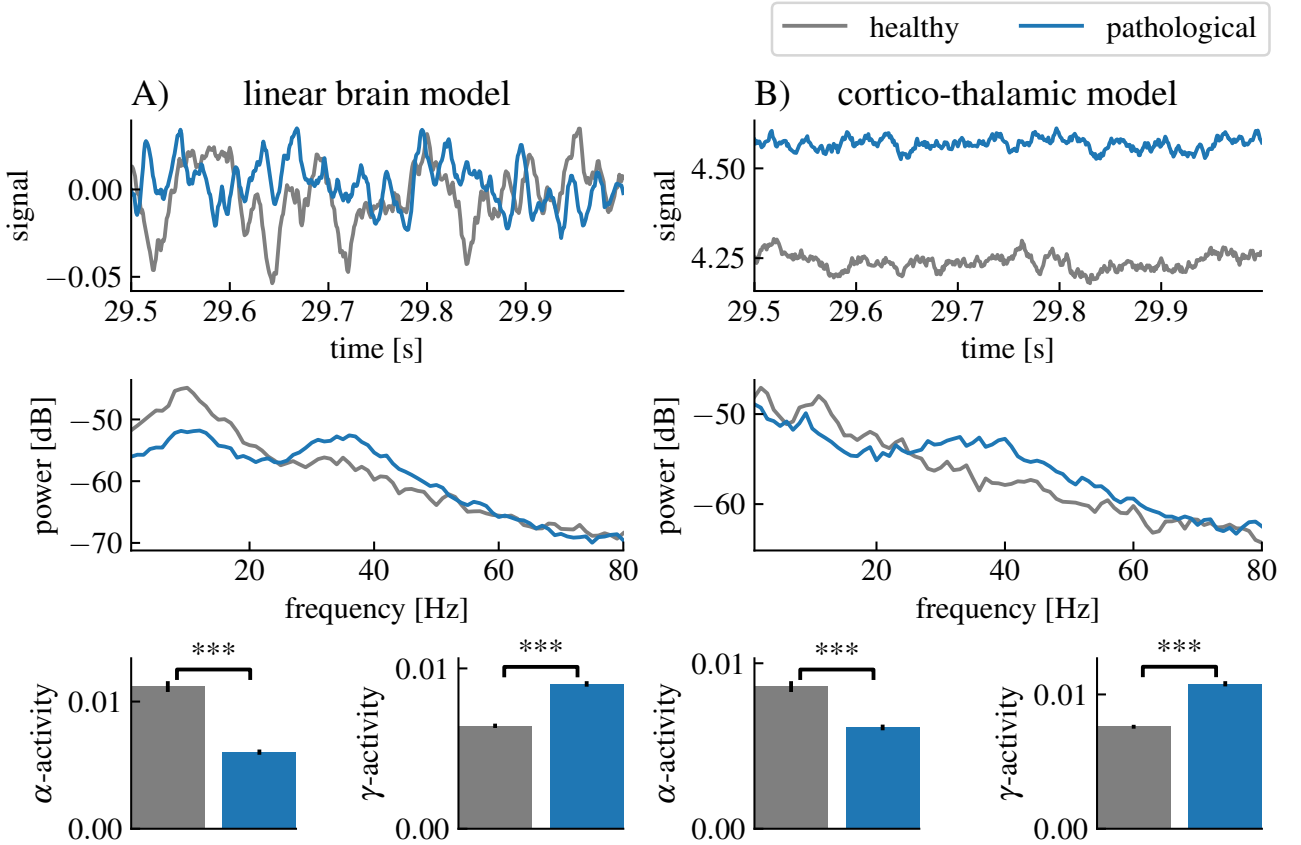


Figure 2: **Healthy and pathological resting state activity of the linear and cortico-thalamic brain models.** The pathological state is characterized by a decreased  $\alpha$ -activity and an increased  $\gamma$ -activity. **A)** Comparison of the healthy and pathological state of the linear brain model. the first row shows the last 500ms of the simulated time series. The second row shows the power spectral densities estimated from the time series and the last row shows the estimated  $\alpha$ - and  $\gamma$ -activities computed from the spectral densities and averaged over 50 simulations. **B)** same as A) but using the cortico-thalamic brain model. “\*\*\*\*” corresponds to a  $p$ -value less than 0.0005 using Welch’s  $t$ -test.

## 2.8 Measuring brain activity

The general activity in EEG measurement is measured by first estimating the power spectral density of the signal using frequency bins of 1Hz and then summing all the frequency bins up to the Nyquist frequency. In practice, the sum will be determined mainly by the activity in low frequencies and, more precisely, near the  $\alpha$ - and  $\gamma$ -activity peaks. To measure the results of our method, we define the  $\alpha$ - and  $\gamma$ -activities as the sum of 1Hz frequency bins only in their respective frequency bands, i.e. respectively 8-12Hz and 25-55Hz. Meanwhile, the total activity of the neurostimulation signal, which we call the mean amplitude of  $u$  is computed from 0Hz to the Nyquist frequency.

## 3 Results

The present work addresses two major problems in closed-loop control: the correct model choice of the systems dynamics and the present conduction delay. The subsequent sections propose solutions for both problems and illustrate them in some detail by applying them to the linear brain activity model from section 2.7.1. The final section demonstrates the closed feedback loop for the cortico-thalamic brain model from Section 2.7.2.

### 3.1 Model estimation

Equations (9) and 10 permit to express the magnitude of  $\hat{g}(f)$  in terms of the spectral densities of observable signals

$$\begin{aligned} |\hat{g}(f)|^2 |\mathcal{F}\{u(t) - \bar{u}\}(f)|^2 &= |\hat{\alpha}(f)|^2 - |\hat{\alpha}_0(f)|^2 \\ |\hat{g}(f)|^2 S_{uu}(f) &= S_{yy}(f) - S_{y_0y_0}(f) \\ |\hat{g}(f)|^2 &= \frac{S_{yy}(f) - S_{y_0y_0}(f)}{S_{uu}(f)}. \end{aligned} \quad (14)$$

The spectral density functions  $S_{y_0y_0}$  and  $S_{yy}$  may be estimated numerically from output data before and during a stimulation with a known chosen stimulation function  $u$ . The estimation may be performed by applying conventional methods, such as the Welch method [51]. These estimations provide the magnitude of the transfer function  $|\hat{g}|$  by utilizing Eq. (14). In detail, at first, we considered the linear model (12) and injected a white noise current into the plant gaining the system's response signal together with the resting state activity, cf. Fig. 3A). The subsequent estimation of  $S_{yy}(f)$ ,  $S_{y_0y_0}(f)$  and  $S_{uu}(f)$  (see Fig. 3B)) from the data permitted to compute the brain input response model  $\hat{g}(f)$  by Eq. (14). We observe a very good accordance of the original model response function and its estimation in magnitude (see Fig. 3C)) and phase (see Fig. 3D)).

### Robustness

The remaining error in the estimated model compared to the original model depends on the amplitude ratio between the stimulated output  $y$  and the resting state output  $y_0$ , cf. Fig. 4. Low stimulation currents or high driving noise can also cause the magnitude vector fitting algorithm not to converge, leading to a non-minimal mean-square error between the fitted and the original models when evaluated at the frequency sample points used for the algorithm.

This problem can be solved by increasing the amplitude of the input current  $u$  that we inject in the plant, which decreases the contribution of the resting state driving noise  $\xi$  to the output signal relative to the input current. Although the remaining dominant input current is also noisy, its value at any time or frequency is known, meaning that it is canceled out in the ratio  $\frac{S_{yy}}{S_{uu}}$  in Eq. (14). This effectively leads to lower noise in the transfer function magnitude data extracted with Eq. (14). The limitation is then set by the maximum amplitude of the current we are allowed to inject into the brain in a given neurostimulation setup. Indeed, the amplitude of the current is limited both for safety reasons that are beyond the scope of this paper and

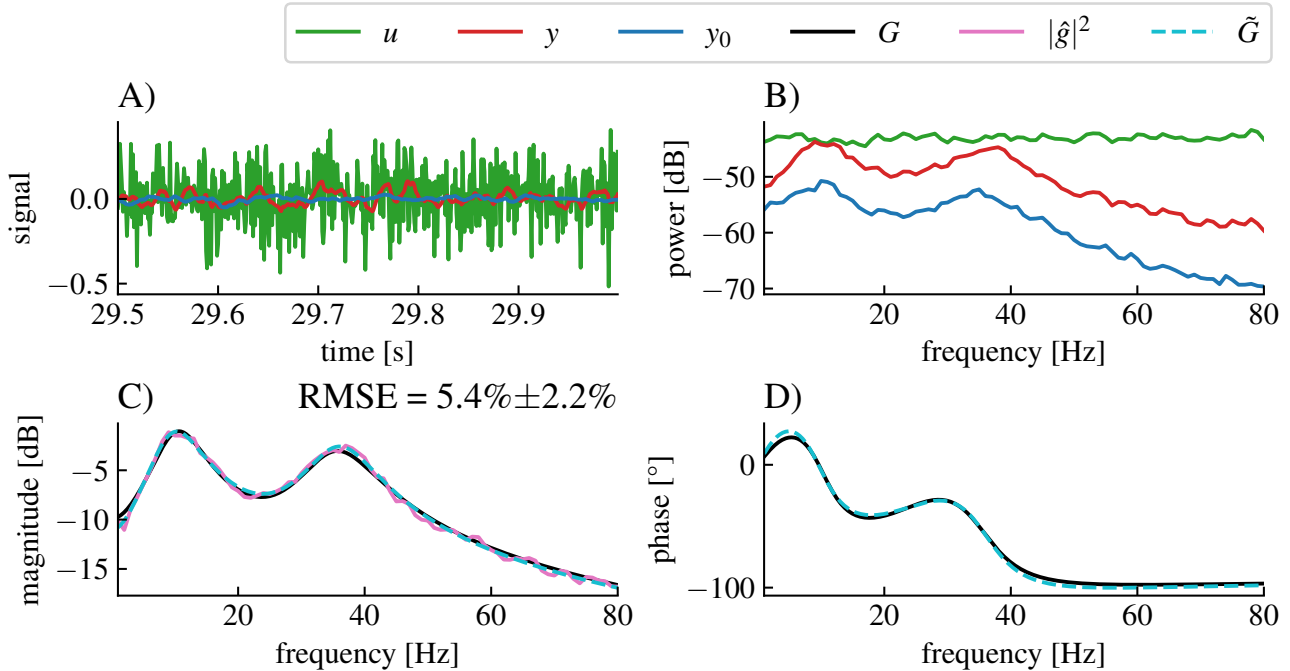


Figure 3: **The magnitude vector fitting algorithm successfully reconstructs the transfer function  $G$  from magnitude-only data.** **A)** Time series of the resting state activity (blue), the input signal (green) and the stimulation response (red). **B)** Spectral densities of the simulated input signal (green), the resting state activity (blue) and the stimulation response (red). The input signal is a white noise with chosen standard deviation 0.005. **C)** Reconstructed gain  $|\hat{g}|$  of the plant input response. The fitted model (dashed cyan) accurately matches the original model (black) with a root mean square error (RMSE) of  $5.4\% \pm 2.2\%$  (Confidence Interval (CI) 95%). The RMSE represents the error percentage between the fitted model's transfer function and the original model's transfer function averaged over 50 trials. The pink curve is the raw data used for fitting, computed from the spectral density data in panel A) using Eq. (14). **D)** Reconstructed phase of the plant input response  $\hat{g}$ .

because of the assumption of linearity on which our method is based and which requires small currents. On the other hand, we can also decrease the noise in the spectral density data by increasing the stimulation time, and hence increasing the amount of data which decreases the contribution of the noise in the power spectral density estimation. Therefore, the accuracy of the model fitting step can be optimized by finding a trade-off between the maximum amplitude of the stimulation current and the maximum duration of the stimulation.

The accuracy of the fitting is generally easily assessed by computing the root mean square error between the data and the fitted model's transfer function. In practice, this could be used as an indicator to evaluate if sufficient stimulation amplitude and time were chosen and then possibly reiterate this step with different parameters. The root mean square error could also be used to directly quantify the error between the transfer function of the estimated model and the original model, which can be an important parameter regarding the stability and the robustness of the closed-loop while performing delay compensation, as discussed in the next section.

### 3.2 Delay compensation

Delay compensation is achieved by adding another LTI system at the output of the controller  $\mathcal{K}$  cf. Fig. 5, whose purpose is to reproduce the transfer function of a negative delay. We call this system the predictor  $\phi$ .

However, perfectly reproducing the transfer function of a negative delay would be impossible since the associated time-domain system would then be a perfect predictor, which is a non-causal, i.e. un-physical, system. Nonetheless, we can build a causal and stable system that behaves almost like a perfect predictor, however only in the frequency ranges of interest.

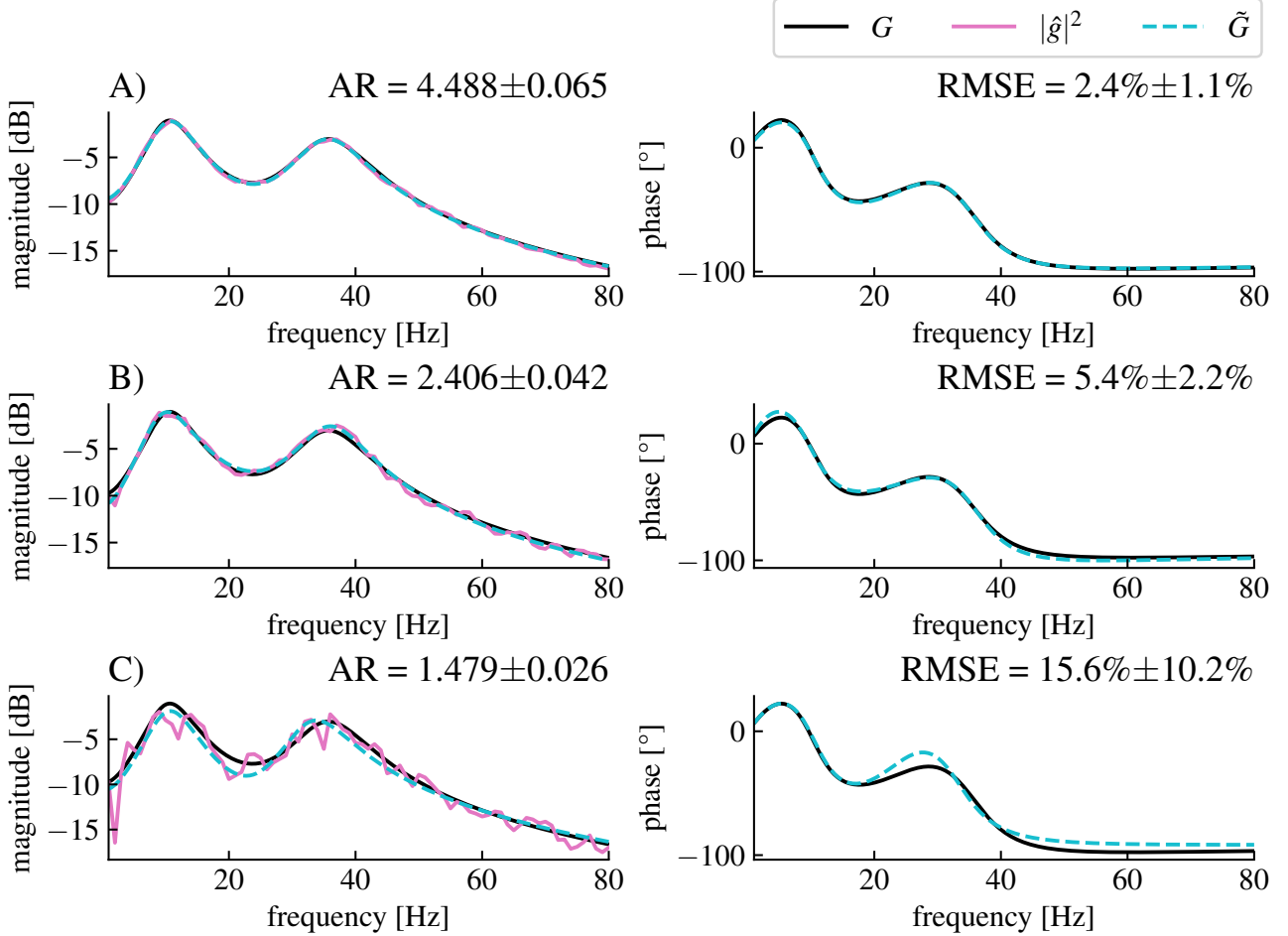


Figure 4: **The magnitude vector fitting algorithm’s performances depend on the amplitude ratio of the stimulation current and the driving noise.** **A)** Reconstructed magnitude and phase data for an amplitude ratio (AR) of  $4.488 \pm 0.065$ . The AR is computed as the ratio between the mean amplitude of  $y$  and the mean amplitude of  $y_0$  averaged over 50 simulations (CI 95%). The fitted model has a corresponding RMSE of  $2.4\% \pm 1.1\%$  (CI95%). **B)** Same as A), but with an AR of  $2.406 \pm 0.042$ , the presented data are the same as in Figure 3C, D. **C)** Same as A), but with an AR of  $1.479 \pm 0.026$  resulting in a RMSE of  $15.6\% \pm 10.2\%$  for the fitted model. We see that the noise level in magnitude data is higher for smaller AR, which leads to higher RMSE (CI 95%) between the fitted model’s transfer function and the original model’s transfer function. The fitted model is coded in dashed cyan and deviates more from the original model for higher noise levels. All data have been computed from 30s long simulated time series.

The numerical implementation of the controller necessitates discretization in time. Consequently, it is reasonable to choose the predictor design as a discrete-time system, meaning that for any input signal at  $x_t : \mathbb{R} \rightarrow \mathbb{R}$  at an instant  $t \in \mathbb{R}$ , it approximately predicts the future signal  $x_{t+\Delta t}$  where  $\Delta t \in \mathbb{R}$  is the sampling time chosen when building the predictor. Since  $x$  is a discrete sequence, its transfer function is obtained using the Z-transform, defined as

$$X(z) = \mathcal{Z}\{x_{n\Delta t}\}(z) := \sum_{n=0}^{\infty} x_{n\Delta t} z^{-n},$$

with  $z \in \mathbb{C}$  and  $X : \mathbb{C} \rightarrow \mathbb{C}$ . Then the transfer function  $\Phi : \mathbb{C} \rightarrow \mathbb{C}$  of a negative delay of one step  $\Delta t$  applied to  $x$  would simply be  $\Phi(z) = z$ , the Z-transform of a one-step delay. However, this choice would be non-causal, which is not implementable numerically in time. Nevertheless, to obtain a stable and implementable system

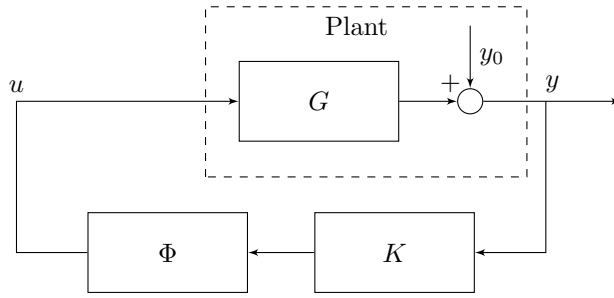


Figure 5: **Closed-loop neurostimulation circuit with predictor**

with a transfer function as close as possible to  $z$ , we chose the ansatz

$$\Phi(z_0) = \frac{b_0 z_0 + b_1}{z_0 - a} = z_0, \quad (15)$$

for a fixed value  $z = z_0$  and where  $a \in \mathbb{R}$  is the pole of the system and  $b_0 \in \mathbb{R}$  and  $b_1 \in \mathbb{R}$  are the polynomial coefficients of the numerator of  $\Phi$ . This equation corresponds to the transfer function of a discrete LTI system with exactly one pole and one zero, which is the closest form of a proper rational function to the identity function of  $z$  in the sense that it has only one more pole. We add the additional constraints that  $|a| < 1$ , since this is the necessary and sufficient condition for the discrete predictor  $\phi$  to be stable.

We choose to reformulate this problem by setting  $a$  as a free parameter. This way, we can select any  $a$  between  $-1$  and  $1$ , and the remaining parameters are found by solving the linear equation  $b_0 z_0 + b_1 = z_0(z_0 - a)$ , where  $z \in \mathbb{C}$  is a chosen complex frequency point at which we want this equation to hold. Since there are two unknowns, we can write a second equation in which we want the derivative of each side of the equation also to be equal, yielding  $b_0 = 2z_0 - a$ . By replacing  $b_0$  in the first equation, we obtain

$$\begin{aligned} z_0(2z_0 - a) + b_1 &= z_0(z_0 - a) \\ b_1 &= -z_0^2. \end{aligned}$$

In the  $z$ -domain, the zero frequency corresponds to  $z_0 = 1$ . We choose to solve this equation for this point, hence we can replace  $a$ ,  $b_0$  and  $b_1$  in Eq. (15) which yields

$$\Phi(z) = \frac{(2 - a)z - 1}{z - a}. \quad (16)$$

This transfer function can then be converted to an associated state-space representation and used for time domain simulations with a sampling time  $\Delta t$ . The output of this system will then be  $y_t \approx u_{t+\Delta t}$  for any input signal  $u_t$ . Simulating delays greater than the system sampling time is simply achieved by concatenating multiple times this predictor system. Here the delay has to be a multiple of the sampling time. This predictor can then be appended to the output of the digital controller  $\mathcal{K}$ .

To avoid closed-loop instability, we must limit the amplitude of the feedback signal computed from the controller input signal. This amplitude is determined by the three systems  $\mathcal{G}$ ,  $\mathcal{H}$  and  $\mathcal{K}$ . Since  $\mathcal{G}$  is defined by the system under study and  $\mathcal{H}$  is the chosen filter defining the desired modifications in the frequency distribution of the observed signal,  $\phi$  (or equivalently parameter  $a$ ) is the only degree of freedom. Figure 6 shows the region of closed-loop stability as a function of the predictor pole  $a$  and the delay.

Because the predictor has a gain that is still slightly greater than one in the frequency ranges of interest, we reduce the weights of the filter  $\mathcal{H}$  to compensate for the excess gain at the  $\alpha$  and  $\gamma$ -peaks. To do this, we simply divide the weight of each band by the magnitude of the predictor system evaluated at the band's natural frequency. This reduces the errors in the closed-loop transfer function in the  $\alpha$  and  $\gamma$ -ranges.

Figure 7C) shows results combining the model estimation by vector fitting and the delay compensation. The proposed closed-loop control yields an increase in  $\alpha$ -power and a decrease in  $\gamma$ -power according to the

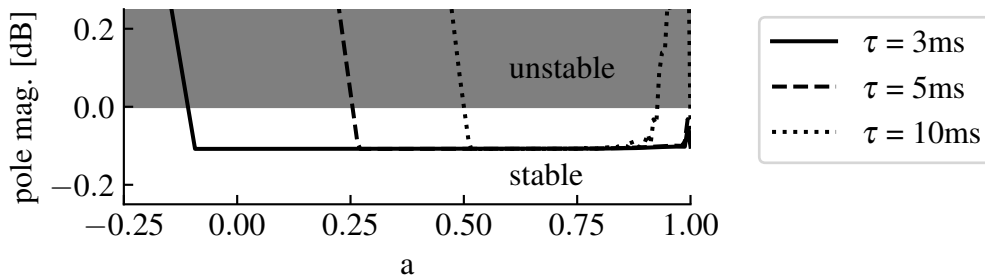


Figure 6: **The predictor pole location affects the closed-loop stability.** The magnitude of the pole with the highest magnitude in the closed-loop transfer function parameterizes the stability of the closed-loop. Indeed, if this value is less than 0 dB, then all the poles of the closed-loop transfer function have a magnitude less than 0 dB, meaning that the system is stable. The system is unstable otherwise. Here the full curve, the dashed curve and the dotted curve correspond to predictors for delays of 3 ms, 5 ms and 10 ms, respectively. The higher the delay is, the lower is the size of the region of closed-loop stability for  $a$ .

employed target filter  $\mathcal{H}$ . The application of PI and LQG control with a Smith predictor for delay compensation (Fig. 7A)B)) has poor performances for higher  $\gamma$ -frequency activity.

### 3.2.1 Accuracy

The error between the achieved closed-loop output activity levels and the target activity levels is highly affected by the delay cf. Fig. 8. This is caused by the phase shift between the input and the output signal, which changes the effect of the control signal on the output in a frequency-dependent manner. Our frequency range of interest is limited to frequencies below 55 Hz, which is the higher limit we use for the  $\gamma$ -range. In this case, the effect of delays of the order of milliseconds will be more visible for higher frequencies and higher delays cf. Fig. 8. However, using our predictor design, we significantly mitigate the effects of the delay in the frequency ranges of interest cf. Fig. 8 red curve. Nonetheless, these effects are still present, creating a limit to the maximum delay our predictor is able to compensate, which in our case is situated around 10ms.

### 3.2.2 Stability and robustness

As discussed earlier, delay compensation can destabilize the closed-loop system depending on the parameters of its components. However, if the correct predictor pole is chosen based on Fig. 6, the closed-loop will remain stable. These values are computed under the assumption that there are no model estimation errors. If we take into account the inaccuracies in the fitted brain model compared to the original brain model, extra gain can add up in the feedback signal, introducing again the risk of destabilizing the closed-loop cf. Fig. 8. The solution is either to simply reduce the amplitude of the spectral density modification that we want to apply by reducing the amplitude of the transfer function of filter  $\mathcal{H}$ , or to reduce the amplitude of the predictor  $\oplus$  reducing its accuracy and possibly increasing delay errors. There is then a trade-off between how much we want to change the gain of the closed-loop transfer function while also compensating delay errors and how much we want to avoid closed-loop destabilization caused by model uncertainties. In any case, the inaccuracies in the estimated brain model create errors in the closed-loop transfer function regardless of the delay, which makes them the main determinant of the performance limits of our method in any given setup. Nevertheless, our method produces smaller  $\alpha$ - $\gamma$ -activities errors than LQG control and produces the smallest error for  $\gamma$ -activities for delays of 1 to 12ms cf. Fig. 8. Comparisons lead to  $p$ -values less than 0.0005 using Welch's  $t$ -test, except for LQG control in row D) where the variance of the data was too high to find any significant difference with this test.

## 3.3 Application to cortico-thalamic circuit model

To extend the analysis to a biologically more realistic model, we employed a nonlinear cortico-thalamic brain model (cf. section 2.7.2). Fitting a linear transfer function to the brain model activity as described above, we





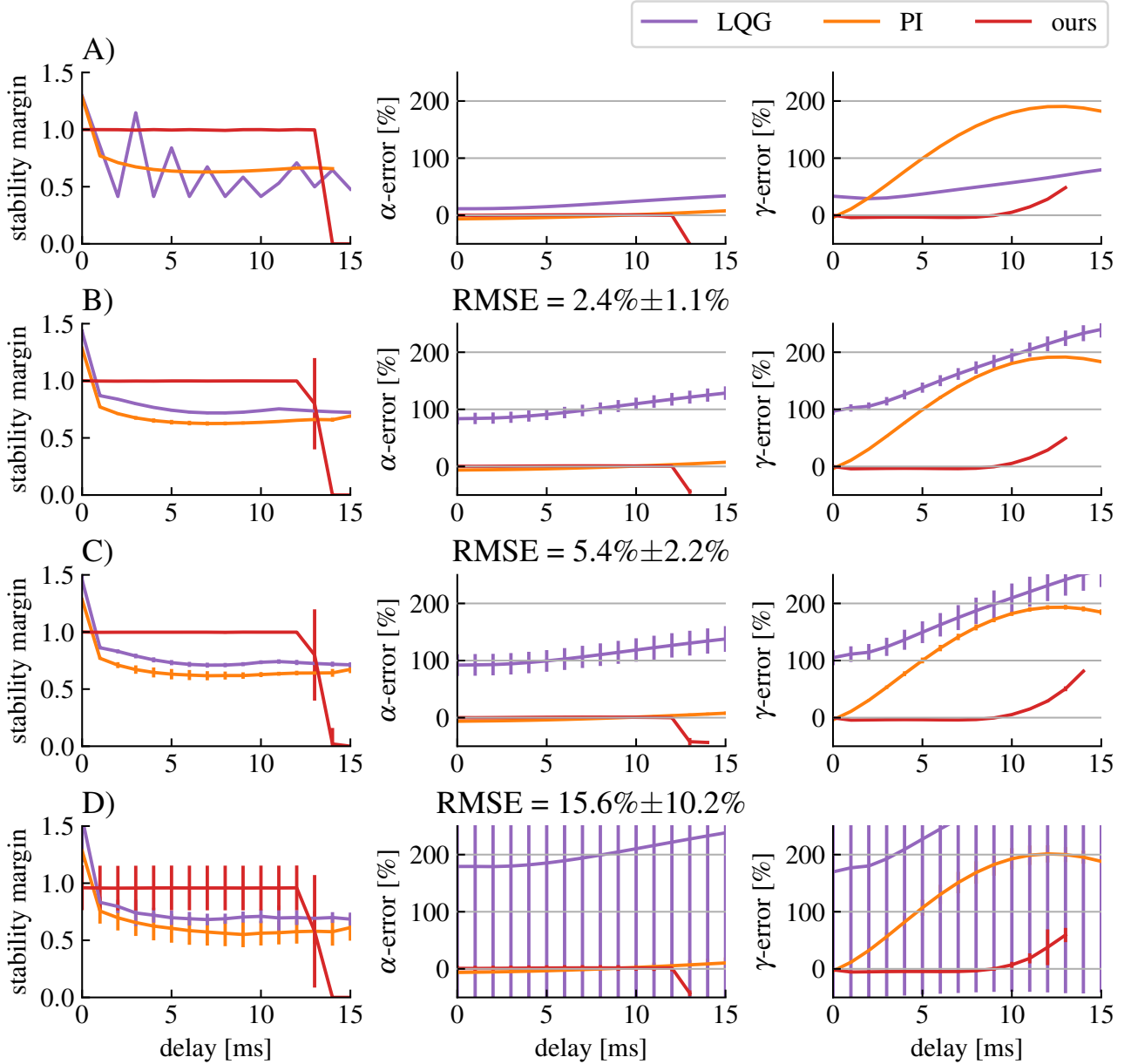


Figure 8: **Delay affects the stability and robustness of the closed-loop transfer function.** **A)** Stability margin and  $\alpha$ - and  $\gamma$ -activities errors in function of the delay with no model errors. The stability margin is computed as the shortest distance between the Nyquist plot and the point  $-1$ . The higher the stability margin, the more robust the closed-loop is to model uncertainties. The  $\alpha$ - and  $\gamma$ -errors represent the error percentage between the measured activity and the target activity **B)****C)****D)** Same as row A) but with model errors (quantified as the RMSE of the model fitting step) corresponding respectively to the model fits shown in Fig. 4A)B)C). The vertical bars represent the standard deviation of their corresponding point.

In fact, the model-based control enhances  $\alpha$ -activity and diminishes  $\gamma$ -activity in good accordance to the imposed filter  $\mathcal{H}$  (see Fig. 9C)). The closed-loop transfer function deviates from the target transfer function for large frequencies beyond the  $\gamma$ -frequency range. This results from the employed conduction delay.

To elucidate better the functions of the different elements of the proposed method, we applied a second closed-loop setup, where the neurostimulation input was applied to the first three layers of the cortex modeled by  $u$  and  $v$  and to the reticulum modeled by  $V_{ret}$  (Fig. 10). In this setting, the response in the high-frequency ranges are mainly produced by the cortex, while the response in low-frequency ranges originates mainly from the reticulum and the thalamic relay structure, with a gap approximately between 10Hz and 20Hz. The weak response between 10 Hz and 20 Hz observable cf. Fig. 10A) is compensated by the controller, which produces

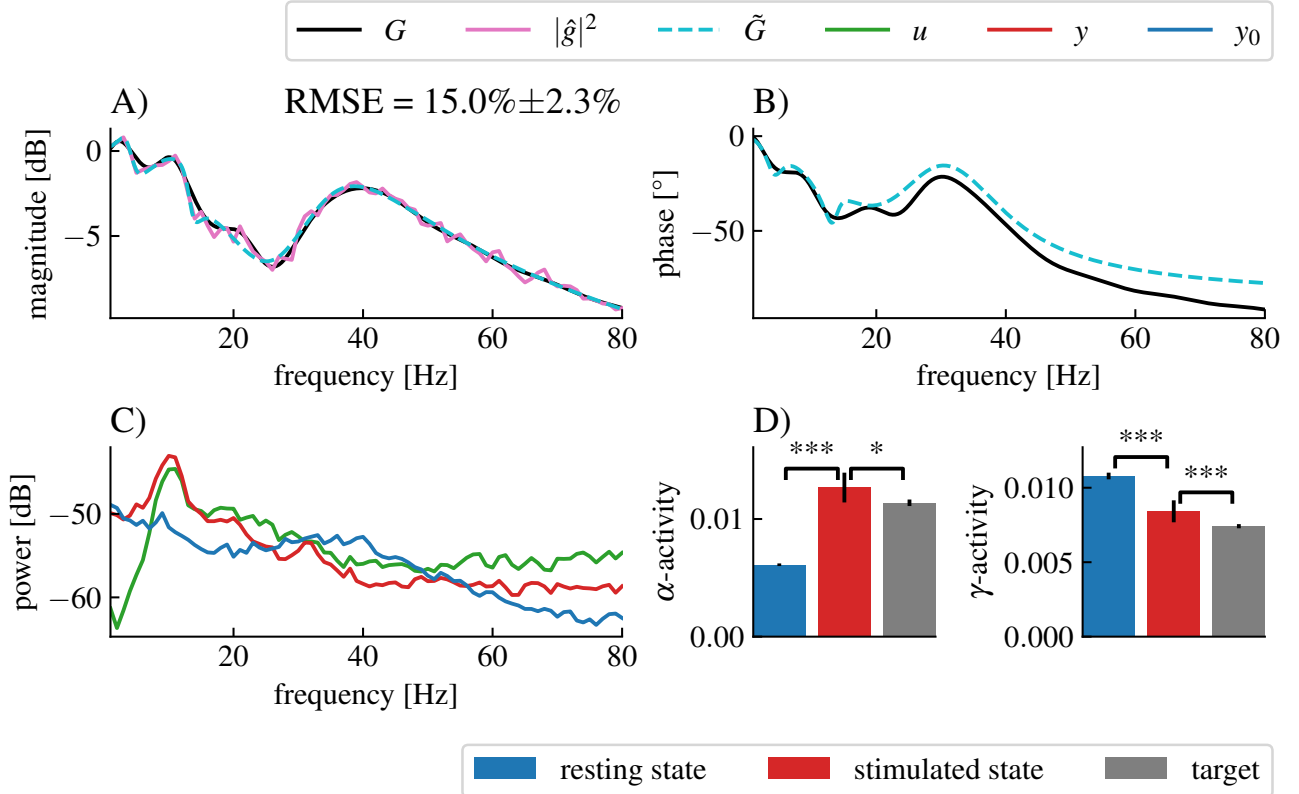


Figure 9: **Fitted model-based control using the cortico-thalamic brain model successfully reproduces the target transfer function in the frequency domains of interest.** **A)** Magnitude of the fitted brain model transfer function (dashed cyan) obtained from the power spectral density data (pink) compared to the magnitude of the original cortico-thalamic brain model transfer function (black). **B)** Phase shift of the fitted transfer function (dashed cyan) compared to the phase shift of the original transfer function (black). **C)** Spectral densities of the resting state activity signal (blue), the stimulated brain output (red) and the stimulation signal (green). **D)**  $\alpha$ - and  $\gamma$ -activities of the closed-loop output averaged over 50 trials for which the fitting step was repeated each time. “\*\*\*” corresponds to a  $p$ -value less than 0.0005 with Welch’s  $t$ -test while “\*” corresponds to a  $p$ -value less than 0.05. The  $\alpha$ - and  $\gamma$ -activities are respectively increased and decreased after application of the closed-loop.

a high magnitude stimulation in the closed-loop for these frequencies cf. Fig. 10C). The second consequence is the inaccuracy of the closed-loop output in the low-frequency ranges, this is caused by the rather long cortico-thalamic internal delay. This delay yields a larger phase shift at low-frequencies and originates from the fact that we observe signals in the cortex, but stimulate in the reticulum.

## 4 Discussion

The goal of the proposed method was to design a delayed closed-loop control method to apply defined modifications to the spectral distribution of an observed signal, such as EEG or LFP. Under the assumption of linear brain stimulation response, the presented work explicitly describes all the steps needed to build a delayed closed-loop neurostimulation setup to restore the physiological brain state of a patient [52]. Since the controller is modeled as a linear time-invariant system, its implementation is lightweight, straightforward, and easily applicable in most embedded systems. Applications to a simple neural populations model (Fig. 7) and to a biologically plausible cortico-thalamic feedback system (Fig. 9 and 10) demonstrate its elements and their impact on the control performance.

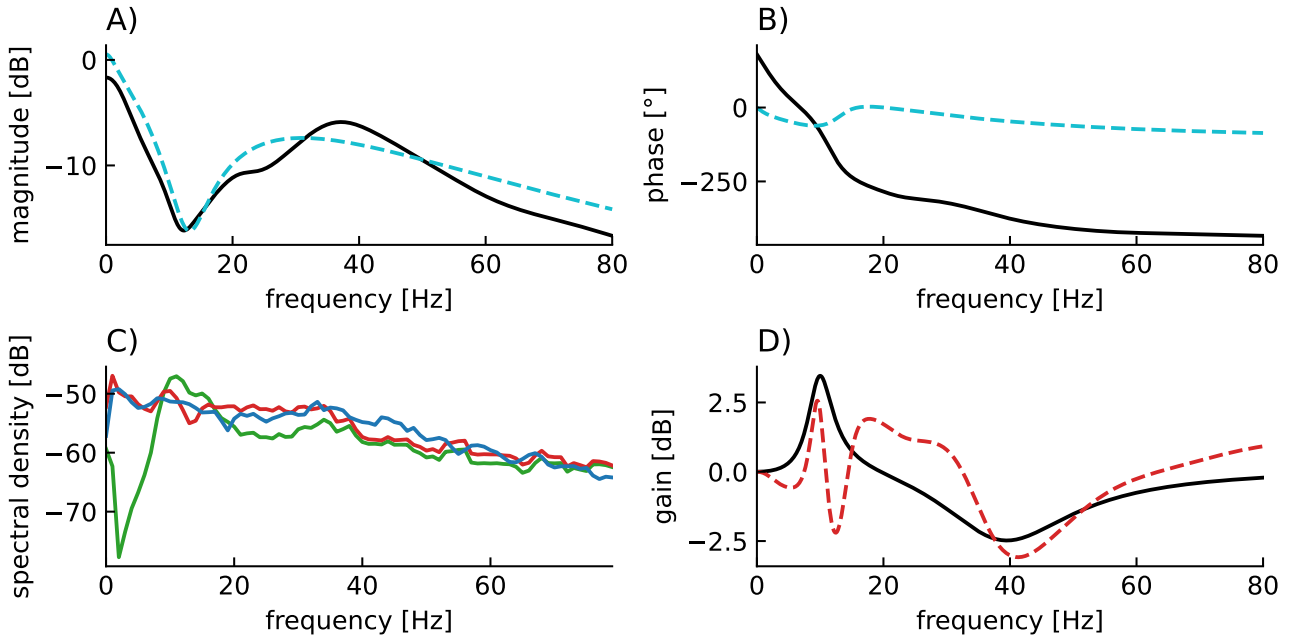


Figure 10: **Reticulum stimulation yields incorrect closed-loop gain in low-frequency ranges.** **A)** Magnitude of the fitted brain model transfer function (dashed cyan) compared to the magnitude of the original cortico-thalamic brain model transfer function (black). **B)** Phase shift of the fitted transfer function (dashed cyan) compared to the magnitude of the original transfer function (black). **C)** Spectral densities of the resting state activity signal (blue), the stimulated brain output (red) and the stimulation signal (green). **D)** Closed-loop transfer function (dashed red), compared to the target transfer function  $1 + H(s)$  (black).

## Main contributions

Our method allows to precisely specify the desired frequency-domain modifications we want to apply to the brain activity. The resulting closed-loop controller can then synthesize in real-time the required closed-loop neurostimulation signal necessary to reach the desired output without the need to track a pre-defined reference signal. This makes the method more flexible since it requires specifying relative rather than absolute signal modifications, which is preferable considering the intra- and inter-patients variability of the EEG spectrum. Furthermore, specifying a reference signal which is a stochastic signal uncorrelated with the noise of the current plant output introduces additional noise in the feedback signal since the controller needs to compensate for both the mean difference between the frequency distributions of the two signals and the difference between the driving noises of the two signals. Therefore, our method is able to track a target frequency distribution for the brain output with a lower current amplitude than classical methods cf. Fig. 7.

## Model estimation

We assume resting state activity signal driven by noise, when no neurostimulation is applied. Injecting a stimulation creates an additional response that adds to the resting state. Consequently, both the resting state signal and response signal can be observed separately in experimental practice and they serve to estimate a linear state-space model as outlined in section 3.1. This approach is successful for both simplified linear models (cf. Figs. 3,4) and neurophysiological realistic nonlinear models (cf. Fig. 9). This approximation is suitable for nonlinear systems whose dynamics evolve close to a stationary state. Several studies have already exposed evidence confirming that the measured brain dynamics behave mostly linearly at macroscopic scales [36], [37]. Moreover, in the case of the brain response to small neurostimulation input, our assumption of the linear brain response is supported by results of [38]. The authors of this study measured the controllability Gramian of their brain model with nonlinear sigmoid transfer function, similar to the cortico-thalamic brain model [48] used in this paper. If the system exhibits nonlinear dynamics far from any linear approximation, such as bistable

dynamics and chaotic evolution, the proposed vector fitting technique may yield a too large model error and thus instability of the closed-loop feedback. The hypothesis of macroscopically linear dynamics has also recently been tested against various nonlinear models [53]. While that work included fitting methods for both linear and nonlinear brain models, our work chose the paradigm of purely frequency domain model fitting with the magnitude vector fitting algorithm [42] and applied it to the brain input response system, which we could isolate thanks to a simple open-loop neurostimulation setup. While models have already been studied in application to neurostimulation [54], [55], we propose a straightforward black box modeling approach that is directly usable for adaptive closed-loop neurostimulation, and is technically applicable easily for each individual patients before any closed-loop neurostimulation sessions.

### Delay compensation

Conduction delays of a few milliseconds in the transmission between observation and stimulation may be negligible in systems evolving on time scales of seconds or longer, but may play an important role in neural systems. Our study demonstrates that such feedback delays may introduce control errors and we show how these errors can be avoided by a novel delay compensation method (section 3.2). Application to the linear model (7) demonstrated its superior performance compared to conventional delay compensation methods. Delay compensating systems have already been described in other work [56], [57]. However, we used a design primarily focused on the correction of a gain error in the closed-loop transfer function, whereas the majority of the current research is based on time domain criterion and stability enforcement [58], [59]. The methods performance, i.e. how well the total gain function fits to the pre-defined transfer function, is good for low-frequencies but weakens for frequencies exceeding a limit frequency. Note that frequency domain compensation has also already been achieved, notably via delay equalizers [60]. However, this would restrict the frequency range in which the delay is compensated, and create additional errors in the surrounding frequencies. Other designs include filters with negative group delays, however their applications are limited to band limited input signals [61], [62]. The predictor design we presented also relies on negative group delay, enabling delay compensation in a large frequency band, while still being applicable to the brain EEG, which is inherently not band limited, because of the noise. Nonetheless, while our predictor design allows to significantly decrease the delay errors in the closed-loop transfer function, the delay still imposes a limit on the controllable frequency range. The larger the delay, the smaller is this limit frequency. Low performance may induce instability in the feedback loop [63] and thus should be avoided. A corresponding stability criteria has been proposed, cf. Fig. 6. Better predictor designs could allow better performance of the closed-loop system for larger delays. The improvement of the accuracy of our closed-loop neurostimulation setup by building more efficient predictor designs is in progress and we refer the reader to future work.

## Limits of our methodology

### Experimental stimulation parameters and safety

Experimental stimulation protocols have to ensure the subjects safety [64] and thus avoid stimulus-induced health risks and complications. For instance, tDCS may be administered for a duration of 60 minutes and a maximum current of 4 mA without yielding health risks. However, parameters beyond these limits may yield adverse effects in subjects, such as skin lesions similar to burns and mania or hypomania in patients with depression [65]. The proposed method does not limit the stimulation duration *per se*, but of course the duration can be chosen accordingly without constraining the method. The method adapts the systems brain rhythms to the target rhythms very rapidly on a time scale of less than a second and hence permits rather short stimulation duration longer than a second.

Moreover, the proposed method does not specify absolute stimulation current magnitude applied. The impact of stimulation at certain magnitudes depends heavily on the stimulation type. In tDCS, anodal stimulation with positive currents have a different impact as cathodal stimulation with negative currents. In addition, currents

are thought to have to pass a certain threshold to yield a measurable effect. In tACS [66], stimulating in the  $\alpha$ -frequency range large and small magnitudes yield excitation and inhibition, respectively, while intermediate magnitudes yield weak effects. Stimulating with a range of frequencies, as in tRNS [67], a 1mA peak-to-peak amplitude for 10 minutes stimulation duration does not yield adverse effects. We conclude that it is not straight-forward to decide which stimulation magnitude applied in the presented method would be safe for human subjects, since the stimulation signal is neither constant, single frequency oscillation nor random noise. In sum, we argue that a maximum peak-to-peak amplitude of 1mA for few tens of minutes may not yield adverse effects, but still may evoke a measurable impact on observations and the brain state. Of course, future experimental studies will gain deeper insights.

### Model internal delay

The internal delay in the brain is not reproducible by the magnitude vector fitting algorithm, which relies on the time invariance of the signals. Hence, this will cause errors in the transfer function of the fitted model (cf. Fig. 9) that are larger for higher contribution of the delay in the output, cf. Fig. 10. To limit this effect, we must minimize the delay between the application of the neurostimulation input and the measurement of the response to this input as much as possible by taking into account the delay between the different brain regions.

### Estimating the closed-loop delay

For delay compensation, in this paper, we assumed that we know the conduction delay in the closed-loop. However, although it is a single constant parameter, we would need a method to measure it for a real closed-loop neurostimulation setup. A straightforward way to do this would be to inject any current into the plant and measure the time lag between the moment at which we inject the input current and the moment at which we measure the output signal. This estimated delay would then correspond to the total closed-loop delay except for the computation delay of the digital controller  $\mathcal{K}$ . This computation delay can be easily measured with the same software used for computation, as it corresponds to the delay needed to perform constant-size matrix multiplications. Moreover, several methods have already been developed to estimate the conduction delays in linear systems [68], [69].

### Direct input current measurements

One of the main challenges to solve for closed-loop neurostimulation is the elimination of direct transmission artifacts from the measured EEG signal [70]. Indeed, when measuring the plant output signal, a portion of the measured signal might be a direct measurement of the input current without any influence from the brain dynamics. In the ideal case, one intends to minimize the contribution of the stimulation input to the observed signal since it would mean that the measured EEG signal does not fully correspond to the brain activity. Hence, reading the EEG of the patient would be more difficult for the user of our closed-loop setup, and the contribution of the brain dynamics to the closed-loop would be smaller. A simple solution to this problem is discussed further below.

### Perspectives

The control proposed allows to perform accurate frequency shaping of the systems' activity spectral distribution. However, this approach is limited to linear models of the brain stimulation response. This may be disadvantageous if the systems dynamics exhibit nonlinear behavior (see e.g. [71]) as we want to represent the brain dynamics realistically. Furthermore, in real-case scenarios, we would also have to take into account the noise in the acquisition of the signal by the sensor and in the application of the input signal by the actuator.

## Filtering out direct input current measurements

Filtering out the direct input current measurements is achievable with our setup removing the strictly proper system requirement while using the magnitude vector fitting algorithm to measure the brain input response. In other words, while fitting the brain input response system, we want the fitted model to be able to contain a direct transmission term corresponding to the direct current measurement. Hence, if the real plant input response contains a significant direct transmission term, it will be identified by the magnitude vector fitting algorithm when synthesizing the estimated plant input response. The second step is then simply to subtract the feedthrough term multiplied by the input current to the plant output signal. Thus, the remaining part of the signal would only correspond to the brain dynamics.

## Application to multiple inputs multiple outputs plants

For now, we only focused on plant with a signal input signal and a single output signal. However, in a real setup, the EEG measurement is typically composed of multiple channels corresponding to different electrodes. This can also be true for the neurostimulation device. For example, with electric current stimulation, we can inject multiple signals using multiple electrodes. This can be simply solved by feeding a single input to each input channel and summing each output to a single output channel. However, when we separate the different channels, we can have more control over each individual output channels. When we have multiple inputs and output, the plant is then a Multiple-Inputs Multiple-Outputs (MIMO) system. Everything developed in this paper is generalizable to MIMO systems, with one caveat: when solving Eq. (6), a unique solution only exists if the system has as more outputs than it has inputs. The user can always ensure this, by using as many neurostimulation input channels than there are EEG output channels. In this generalized setup, we can also define the filter  $\mathcal{H}$  to apply different modifications to each output channel.

## Neurostimulation effects on larger time scales

Our method relies only on the short term dynamics of the brain, using signal feedback and delay compensation to produce an adaptive stimulation current and obtain the desired EEG frequency distribution. However, more traditional neurostimulation techniques rely on the long term dynamics of neural plasticity, which is not modeled in the brain models we use in this paper. Long term brain adaptation to neurostimulation could cause the EEG frequency distribution to diverge from the desired frequency distribution after several minutes of stimulation. This effect could be compensated either by reiterating the model identification step and performing neurostimulation again, or by adjusting the weight of the filter  $\mathcal{H}$  according to the observed changes in real-time. Incorporating the effect of neural plasticity in the brain models would allow our method to produce predictable and durable modification to the EEG frequency distribution, even after we stop the stimulation.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author Contributions

TW, AH and MD contributed to the development of the methods presented in this study. TW produced the source code used for the simulations. JR and AH wrote the introduction section. TW and AH wrote the other sections of the manuscript. All the authors read and approved the submitted version.

## Funding

This research was funded by Inria in the “Action Exploratoire” project *A/D Drugs*.

## Data Availability Statement

The source code used for the simulation results can be freely accessed here: <https://github.com/Thomas-Wahl/neurocodec>

## References

- [1] Yves Edel and François Caroli. Histoire de l'électrochoc : des traitements électriques à la convulsivothérapie en psychiatrie. *Bulletin d'histoire de l'électricité*, 9:87–114, 1987.
- [2] Zhe Sage Chen, Prathamesh (Param) Kulkarni, Isaac R. Galatzer-Levy, Benedetta Bigio, Carla Nasca, and Yu Zhang. Modern views of machine learning for precision psychiatry. *Patterns*, 3(11):100602, 2022.
- [3] FT Sun and MJ Morrell. Closed-loop neurostimulation: the clinical experience. *Neurotherapeutics*, 11(3):553–563, 2014.
- [4] E Basar. Brain oscillations in neuropsychiatric disease. *Dialogues Clin Neurosci.*, 15(13):291–300, 2013.
- [5] Joshua J Schulman, Robert Cancro, Sandlin Lowe, Feng Lu, Kerry D Walton, and Rodolfo R Llinás. Imaging of thalamocortical dysrhythmia in neuropsychiatry. *Front. Hum. Neurosci.*, 5:69, 2011.
- [6] Andrea A. Kühn, Florian Kempf, Christof Brücke, Louise Gaynor Doyle, Irene Martinez-Torres, Alek Pogosyan, Thomas Trottenberg, Andreas Kupsch, Gerd-Helge Schneider, Marwan I. Hariz, Wim Vandenberghe, Bart Nuttin, and Peter Brown. High-frequency stimulation of the subthalamic nucleus suppresses oscillatory  $\beta$ activity in patients with parkinson's disease in parallel with improvement in motor performance. *Journal of Neuroscience*, 28(24):6165–6173, 2008.
- [7] JE Fleming, E Dunn, and MM Lowery. Simulation of closed-loop deep brain stimulation control schemes for suppression of pathological beta oscillations in parkinson's disease. *Front. Neurosci.*, 14:166, 2020.
- [8] MK Hosain, A Kouzani, and S Tye. Closed loop deep brain stimulation: an evolving technology. *Australas Phys Eng Sci Med.*, 37(4):619–634, 2014.
- [9] J Brunelin, M Mondino, L Gassab, F Haesebaert, L Gaha, MF Suaud-Chagny, M Saoud, A Mechri, and E Poulet. Examining transcranial direct-current stimulation (tDCS) as a treatment for hallucinations in schizophrenia. *Am J Psychiatry.*, 169(7):719–724, 2012.
- [10] CJ Stagg, A Antal, and MA Nitsche. Physiology of transcranial direct current stimulation. *J ECT.*, 34(3):144–152, 2018.
- [11] V Nejati, MA Salehinejad, MA Nitsche, A Najian, and AH Javadi. Transcranial direct current stimulation improves executive dysfunctions in adhd: Implications for inhibitory control, interference control, working memory, and cognitive flexibility. *J. Atten. Disord.*, 24(13):1928–1943, 2020.
- [12] D Bennabi and E Haffen. Transcranial direct current stimulation (tdcs): A promising treatment for major depressive disorder? *Brain Sci.*, 8(5):81, 2018.
- [13] Khaled Nasr, David Haslacher, Eran Dayan, Nitzan Censor, Leonardo G. Cohen, and Surjo R. Soekadar. Breaking the boundaries of interacting with the human brain using adaptive closed-loop stimulation. *Progress in Neurobiology*, 216:102311, 2022.



- [14] Walter Paulus. Transcranial electrical stimulation (tes – tdc; trns, tacs) methods. *Neuropsychological Rehabilitation*, 21(5):602–617, 2011. PMID: 21819181.
- [15] J Prosky, J Cagle, KK Sellers, R Gilron, C de Hemptinne, A Schmitgen, PA Starr, EF Chang, and P Shirvalkar. Practical closed-loop strategies for deep brain stimulation: Lessons from chronic pain. *Front. Neurosci.*, 15:762097, 2021.
- [16] S Stanslaski, H Farooqi, DE Sanabria, and TI Netoff. Fully closed loop test environment for adaptive implantable neural stimulators using computational models. *J. Med. Device*, 16(3):034501, 2022.
- [17] A Hartshorn and B Jobst. Responsive brain stimulation in epilepsy. *Ther Adv Chronic Dis.*, 9(7):135–142, 2018.
- [18] Aino E. Tervo, Jaakko O. Nieminen, Pantelis Lioumis, Johanna Metsomaa, Victor H. Souza, Heikki Sinisalo, Matti Stenroos, Jukka Sarvas, and Risto J. Ilmoniemi. Closed-loop optimization of transcranial magnetic stimulation with electroencephalography feedback. *Brain Stimulation*, 15(2):523–531, 2022.
- [19] KW Scangos, AN Khambhati, PM Daly, GS Makhoul, LP Sugrue, H Zamanian, TX Liu, VR Rao, KK Sellers, HE Dawes, PA Starr, AD Krystal, and EF Chang. Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nat. Med.*, 27:1696–1700, 2021.
- [20] T Haeusermann, CR Lechner, KC Fong, AB Sideman, A Jaworska, W Chiong, and D Dohan. Closed-loop neuromodulation and self-perception in clinical treatment of refractory epilepsy. *AJOB Neuroscience*, 14(1):32–44, 2023.
- [21] J Guerrero Moreno, CE Biazoli Jr., A Fontes Baptista, and L Remoaldo Trambaiolli. Closed-loop neurostimulation for affective symptoms and disorders: An overview. *Biol. Psychol.*, 161:108081, 2021.
- [22] Michael F Bolus, Adam A Willats, Clarissa J Whitmire, Christopher J Rozell, and Garrett B Stanley. Design strategies for dynamic closed-loop optogenetic neurocontrol in vivo. *Journal of neural engineering*, 15(2):026011, 2018.
- [23] Fei Su, Karthik Kumaravelu, Jiang Wang, and Warren M Grill. Model-based evaluation of closed-loop deep brain stimulation controller to adapt to dynamic changes in reference signal. *Frontiers in neuroscience*, 13:956, 2019.
- [24] M Brandon Westover, Seong-Eun Kim, ShiNung Ching, Patrick L Purdon, and Emery N Brown. Robust control of burst suppression for medical coma. *Journal of neural engineering*, 12(4):046004, 2015.
- [25] Yulin Zhu, Jiang Wang, Huiyan Li, Chen Liu, and Warren M Grill. Adaptive parameter modulation of deep brain stimulation based on improved supervisory algorithm. *Frontiers in neuroscience*, 15:750806, 2021.
- [26] Michael F Bolus, Adam A Willats, Christopher J Rozell, and Garrett B Stanley. State-space optimal feedback control of optogenetically driven neural activity. *Journal of neural engineering*, 18(3):036006, 2021.
- [27] Yuxiao Yang, Allison T Connolly, and Maryam M Shanechi. A control-theoretic system identification framework and a real-time closed-loop clinical simulation testbed for electrical brain stimulation. *Journal of neural engineering*, 15(6):066007, 2018.
- [28] Yuxiao Yang, Justin T Lee, Jennifer A Guidera, Ksenia Y Vlasov, JunZhu Pei, Emery N Brown, Ken Solt, and Maryam M Shanechi. Developing a personalized closed-loop controller of medically-induced coma in a rodent model. *Journal of neural engineering*, 16(3):036022, 2019.

- [29] Hao Fang and Yuxiao Yang. Predictive neuromodulation of cingulo-frontal neural dynamics in major depressive disorder using a brain-computer interface system: A simulation study. *Frontiers in Computational Neuroscience*, 2023.
- [30] Hao Fang and Yuxiao Yang. Designing and validating a robust adaptive neuromodulation algorithm for closed-loop control of brain states. *Journal of Neural Engineering*, 19(3):036018, 2022.
- [31] S Martin, I Iturrate, R Chavarriaga, R Leeb, A Sobolewski, AM Li, J Zaldivar, I Peciu-Florianu, E Pralong, M Castro-Jiménez, D Benninger, F Vingerhoets, RT Knight, J Bloch, and J del R. Millán. Differential contributions of subthalamic beta rhythms and 1/f broadband activity to motor symptoms in parkinson’s disease. *npj Parkinson’s Disease*, 4:32, 2018.
- [32] Fleur M Howells, Hendrik S Temmingh, Jennifer H Hsieh, Andrea V van Dijen, David S Baldwin, and Dan J Stein. Electroencephalographic delta/alpha frequency activity differentiates psychotic disorders: a study of schizophrenia, bipolar disorder and methamphetamine-induced psychotic disorder. *Translational psychiatry*, 8(1):1–11, 2018.
- [33] Hirano Y, Oribe N, Kanba S, Onitsuka T, Nestor PG, and Spencer KM. Spontaneous gamma activity in schizophrenia. *JAMA Psychiatry*, 72(8):813–821, 2015.
- [34] PE Holtzheimer and HS Mayberg. Deep brain stimulation for psychiatric disorders. *Annu Rev Neurosci.*, 34:2890307, 2011.
- [35] Gregor Leicht, Sebastian Vauth, Nenad Polomac, Christina Andreou, Jonas Rauh, Marius Mußmann, Anne Karow, and Christoph Mulert. EEG-Informed fMRI Reveals a Disturbed Gamma-Band-Specific Network in Subjects at High Risk for Psychosis. *Schizophrenia Bulletin*, 42(1):239–249, 07 2015.
- [36] Zhongming Liu, Cristina Rios, Nanyin Zhang, Lin Yang, Wei Chen, and Bin He. Linear and nonlinear relationships between visual stimuli, eeg and bold fmri signals. *Neuroimage*, 50(3):1054–1066, 2010.
- [37] D Popivanov, J Dushanova, A Mineva, and I Krekule. Detection of successive changes in dynamics of eeg time series: linear and nonlinear approach. In *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 1590–1591. IEEE, 1996.
- [38] Seul Ah Kim and ShiNung Ching. Quasilinearization-based controllability analysis of neuronal rate networks. In *2016 American Control Conference (ACC)*, pages 7371–7376. IEEE, 2016.
- [39] Joao P Hespanha. *Linear systems theory*. Princeton university press, 2018.
- [40] A Khintchine. Korrelationstheorie der stationären stochastischen prozesse. *Mathematische Annalen*, 109(1):604–615, 1934.
- [41] CW Gardiner. *Handbook of Stochastic Methods*. Springer, Berlin, 2004.
- [42] L De Tommasi, B Gustavsen, and Tom Dhaene. Robust transfer function identification via an enhanced magnitude vector fitting algorithm. *IET control theory & applications*, 4(7):1169–1178, 2010.
- [43] Otto JM Smith. A controller to overcome dead time. *ISA J.*, 6:28–33, 1959.
- [44] Manfred Morari and Evangelhos Zafiriou. *Robust process control*. Morari, 1989.
- [45] Karl J Åström. *Introduction to stochastic control theory*. Courier Corporation, 2012.
- [46] Axel Hutt. The anesthetic propofol shifts the frequency of maximum spectral power in eeg during general anesthesia: analytical insights from a linear model. *Frontiers in Computational Neuroscience*, 7:2, 2013.
- [47] Charles Van Loan. Computing integrals involving the matrix exponential. *IEEE transactions on automatic control*, 23(3):395–404, 1978.

- [48] Joséphine Riedinger and Axel Hutt. Mathematical model insights into eeg origin under transcranial direct current stimulation (tdcs) in the context of psychosis. *Journal of Clinical Medicine*, 11(7):1845, 2022.
- [49] M. Hashemi, A. Hutt, and J. Sleigh. How the cortico-thalamic feedback affects the EEG power spectrum over frontal and occipital regions during propofol-induced anaesthetic sedation. *J. Comput. Neurosci.*, 39(1):155, 2015.
- [50] P.L. Nunez and R. Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, New York - Oxford, 2006.
- [51] PD Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodogram. *Trans. Audio Electroacoustics*, AU-15(2):70–73, 1967.
- [52] Adam O Hebb, Jun Jason Zhang, Mohammad H Mahoor, Christos Tsiokos, Charles Matlack, Howard Jay Chizeck, and Nader Pouratian. Creating the feedback loop: closed-loop neurostimulation. *Neurosurgery Clinics*, 25(1):187–204, 2014.
- [53] Erfan Nozari, Maxwell A Bertolero, Jennifer Stiso, Lorenzo Caciagli, Eli J Cornblath, Xiaosong He, Arun S Mahadevan, George J Pappas, and Dani Smith Bassett. Is the brain macroscopically linear? a system identification of resting state dynamics. *arXiv preprint arXiv:2012.12351*, 2020.
- [54] Julien Modolo, Alexandre Legros, Alex W Thomas, and Anne Beuter. Model-driven therapeutic treatment of neurological disorders: reshaping brain rhythms with neuromodulation. *Interface focus*, 1(1):61–74, 2011.
- [55] Tim Wagner, Uri Eden, Jarrett Rushmore, Christopher J Russo, Laura Dipietro, Felipe Fregni, Stephen Simon, Stephen Rotman, Naomi B Pitskel, Ciro Ramos-Estebanez, et al. Impact of brain tissue filtering on neurostimulation fields: a modeling study. *Neuroimage*, 85:1048–1057, 2014.
- [56] Seyyed Amir Hosseini, Mohammadreza Toulabi, Ahmad Salehi Dobakhshari, Alireza Ashouri-Zadeh, and Ali Mohammad Ranjbar. Delay compensation of demand response and adaptive disturbance rejection applied to power system frequency control. *IEEE Transactions on Power Systems*, 35(3):2037–2046, 2019.
- [57] Liwen Guo, Frank Cardullo, Jake Houck, Lon Kelley, and Thomas Wolters. New predictive filters for compensating the transport delay on a flight simulator. In *AIAA Modeling and Simulation Technologies Conference and Exhibit*, page 5441. American Institute of Aeronautics and Astronautics, 2004.
- [58] Gregory S Ledva, Evangelos Vrettos, Silvia Mastellone, Göran Andersson, and Johanna L Mathieu. Managing communication delays and model error in demand response for frequency regulation. *IEEE Transactions on Power Systems*, 33(2):1299–1308, 2017.
- [59] Sahin Sönmez and Saffet Ayasun. Stability region in the parameter space of pi controller for a single-area load frequency control system with time delay. *IEEE Transactions on Power Systems*, 31(1):829–830, 2015.
- [60] Symon K Podilchak, Brian M Frank, Al P Freundorfer, and Yahia MM Antar. High speed metamaterial-inspired negative group delay circuits in cmos for delay equalization. In *2009 2nd Microsystems and Nanoelectronics Research Conference*, pages 9–12. IEEE, 2009.
- [61] NS Bukhman and SV Bukhman. On the negative delay time of a narrow-band signal as it passes through the resonant filter of absorption. *Radiophysics and Quantum Electronics*, 47(1):68–76, 2004.
- [62] Henning U Voss. A universal negative group delay filter for the prediction of band-limited signals. *arXiv preprint arXiv:1706.07326*, 2017.
- [63] Leonid Mirkin and Zalman J. Palmor. *Control Issues in Systems with Loop Delays*, chapter 59, pages 627–648. Birkhäuser Boston, Boston, MA, 2005.

- [64] MH Ko. Safety of transcranial direct current stimulation in neurorehabilitation. *Brain Neurorehabil.*, 14(1):e9, 2021.
- [65] Hideyuki Matsumoto and Yoshikazu Ugawa. Adverse events of tdc and tacs: A review. *Clinical Neurophysiology Practice*, 2:19–25, 2017.
- [66] V Moliadze, D Atalay, A. Antal, and W Paulus. Close to threshold transcranial electrical stimulation preferentially activates inhibitory networks before switching to excitation with higher intensities. *Brain Stimul.*, 5:505–511, 2012.
- [67] W Potok, O van der Groen, M Bächinger, D Edwards, and N Wenderoth. Transcranial random noise stimulation modulates neural processing of sensory and motor circuits, from potential cellular mechanisms to behavior: A scoping review. *eNeuro*, 9(1):ENEURO.0248–21.2021, 2022.
- [68] Jan Schier. Estimation of transport delay using parallel recursive modified gram–schmidt algorithm. *International journal of adaptive control and signal processing*, 11(5):431–442, 1997.
- [69] NA Dudarenko, NA Polinova, and AV Ushakov. Fundamental matrix of linear continuous system in the problem of estimating its transport delay. *Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki*, 14(5), 2014.
- [70] Iñaki Iturrate, Michael Pereira, and José del R Millán. Closed-loop electrical neurostimulation: challenges and opportunities. *Current Opinion in Biomedical Engineering*, 8:28–37, 2018.
- [71] A. Hutt and P. beim Graben. Sequences by metastable attractors: interweaving dynamical systems and experimental data. *Front. Appl. DYn. Syst. Stat.*, 3:11, 2017.

## Cortico-thalamic brain model details

The differential equation system (13) develops as

$$\begin{aligned}
\tau_e \frac{dV_e(t)}{dt} &= -V_e(t) + F_e T_c (V_e(t) - V_i(t)) + F_{ct} T_{th} (V_{th,e}(t) - V_{th,i}(t)) + F_{ccx} S_e(v(t)) + \mu_e + I_e + \xi_e(t) + b_1 u(t) \\
\tau_i \frac{dV_i(t)}{dt} &= -V_i(t) + F_i T_c (V_e(t) - V_i(t)) + \mu_i + I_i + \xi_i(t) + b_2 u(t) \\
\tau_{th,e} \frac{dV_{th,e}(t)}{dt} &= -V_{th,e}(t) + F_{tc} T_c (V_e(t - \tau) - V_i(t - \tau)) + \mu_{th,e} + \xi_{th,e}(t) \\
\tau_{th,i} \frac{dV_{th,i}(t)}{dt} &= -V_{th,i}(t) + F_{tr} T_{ret} (V_{ret}(t)) + \mu_{th,i} + \xi_{th,i}(t) \\
\tau_{ret} \frac{dV_{ret}(t)}{dt} &= -V_{ret}(t) + F_{rt} T_{th} (V_{th,e}(t) - V_{th,i}(t)) + F_{rc} T_c (V_e(t - \tau) - V_i(t - \tau)) + \mu_{ret} + \xi_{ret}(t) \\
\tau_{ce} \frac{dv(t)}{dt} &= -v(t) + F_{cx} S_e(v(t)) - M_{cx} S_i(w(t)) + M_{cx,th} T_{th} (V_{th,e}(t - \tau) - V_{th,i}(t - \tau)) + \mu_{ce} + I_{ce} + \xi_{ce}(t) + b_3 u(t) \\
\tau_{ci} \frac{dw(t)}{dt} &= -w(t) - F_{cx} S_i(w(t)) + M_{cx} S_e(v(t)) + \mu_{ci} + I_{ci} + \xi_{ci}(t) + b_4 u(t)
\end{aligned} \tag{17}$$

where the transfer functions are defined as

$$\begin{aligned}
T_m(x) &= \frac{1}{2} \left( 1 - \operatorname{erf} \left( -\frac{x}{\sqrt{2}\sigma_m} \right) \right) \\
S_m(x) &= \frac{1}{2} \left( 1 - \operatorname{erf} \left( -\frac{x}{\sqrt{2}\sigma_{cm}} \right) \right).
\end{aligned} \tag{18}$$

The  $\xi_x$  terms represent the driving noises, which are uncorrelated Gaussian noise defined as

$$\langle \xi_x(t) \rangle = 0 \quad , \quad \langle \xi_x(t) \xi_y(t') \rangle = \frac{Q_x}{N} \delta_{xy} \delta(t - t'),$$

with  $x = e, i, (th, e), (th, i), ret, ce, ci$ . The variances in Eq. (18) are defined as

$$\begin{aligned}
\sigma_c^2 &= \frac{Q_e}{\tau_e} + \frac{Q_i}{\tau_i}, & \sigma_{th}^2 &= \frac{Q_{th,e}}{\tau_{th,e}} + \frac{Q_{th,i}}{\tau_{th,i}}, & \sigma_{ret}^2 &= \frac{Q_{ret}}{\tau_{ret}} \\
\sigma_{ce}^2 &= \frac{Q_{ce}}{\tau_{ce}}, & \sigma_{ci}^2 &= \frac{Q_{ci}}{\tau_{ci}}
\end{aligned}$$

All the parameters are given in Table 3.

parameter	description	value
$\tau_e$	exc. decay time (infragranular)	10 ms
$\tau_i$	inh. decay time (infragranular)	50 ms
$\tau_{th,e}$	exc. decay time (relay)	5 ms
$\tau_{th,i}$	inh. decay time (relay)	30 ms
$\tau_{ret}$	exc. decay time (reticular)	8 ms
$\tau_{ce}$	exc. decay time (supragranular)	5 ms
$\tau_{ci}$	inh. decay time (supragranular)	20 ms
$\tau$	cortico-thalamic propagation delay	40 ms
$F_e$	exc. synaptic strength	1.0
$F_i$	inh. synaptic strength	2.0
$F_{ct}$	synaptic strength (relay $\rightarrow$ cortex)	1.2
$F_{tc}$	synaptic strength (cortex $\rightarrow$ relay)	1.0
$F_{tr}$	synaptic strength (reticular $\rightarrow$ relay)	1.0
$F_{rt}$	synaptic strength (relay $\rightarrow$ reticular)	0.3
$F_{rc}$	synaptic strength (cortex $\rightarrow$ reticular)	0.6
$F_{cx}$	synaptic strength (exc. $\rightarrow$ exc.)	2.18
$M_{cx}$	synaptic strength (inh. $\rightarrow$ exc.)	3.88
$F_{ccx}$	synaptic strength (supragranular $\rightarrow$ infragranular)	0.05
$F_{cx,th}$	synaptic strength (thalamic relay $\rightarrow$ supragranular)	0.1
$\mu_e$	exc. noise input (infragranular)	0.1
$\mu_i$	inh. noise input (infragranular)	0.0
$\mu_{th,e}$	exc. noise input (relay)	1.3
$\mu_{th,i}$	inh. noise input (relay)	1.0
$\mu_{ret}$	exc. noise input (reticular)	0.0
$\mu_{ce}$	exc. noise input (supragranular)	0.05
$\mu_{ci}$	inh. noise input (supragranular)	0.05
$I_e$	exc. resting input (infragranular)	2.7
$I_i$	inh. resting input (infragranular)	1.7
$I_{ce}$	exc. resting input (supragranular)	1.1
$I_{ci}$	inh. resting input (supragranular)	0.4
$Q_e$	exc. input noise variance (infragranular) (pathological)	$3 \times 10^{-5}$
$Q'_e$	exc. input noise variance (infragranular) (healthy)	$5 \times 10^{-5}$
$Q_i$	inh. input noise variance (infragranular)	0.001
$Q_{th,e}$	exc. input noise variance (relay) (pathological)	$2.5 \times 10^{-6}$
$Q'_{th,e}$	exc. input noise variance (relay) (healthy)	$1.2 \times 10^{-5}$
$Q_{th,i}$	inh. input noise variance (relay)	$12.6 \times 10^{-6}$
$Q_{ret}$	exc. input noise variance (reticular)	$10.9 \times 10^{-6}$
$Q_{ce}$	exc. input noise (supragranular)	$2 \times 10^{-5}$
$Q_{ci}$	inh. input noise (supragranular) (pathological)	$8 \times 10^{-5}$
$Q'_{ci}$	inh. input noise (supragranular) (healthy)	$1 \times 10^{-6}$
$N$	number of neurons	1000
$b_{1,2,3,4}$	input coupling constants	1
$c_1$	observation coefficient (supragranular)	0.3
$c_3$	observation coefficient (infragranular)	1

Table 3: **Parameter set of model (17)**. The choice of parameters is for the most part based on the paper in which it was developed [48].