Cortical network structure mediates response to stimulation: an optogenetic study in non-human primates

Julien Bloch[1,6,7], Alexander Greaves-Tunnell[2], Eric Shea-Brown[3, 6, 7], Zaid Harchaoui[2], Ali Shojaie[4], and Azadeh Yazdan-Shahmorad[1, 5, 6, 7, 8]

[1]Department of Bioengineering, University of Washington, Seattle
[2]Department of Statistics, University of Washington, Seattle
[3]Department of Applied Mathematics, University of Washington, Seattle
[4]Department of Biostatistics, University of Washington, Seattle
[5]Department of Electrical and Computer Engineering, University of Washington, Seattle
[6]Center for Neurotechnology, University of Washington, Seattle
[7]Computational Neuroscience Center, University of Washington, Seattle
[8]Washington National Primate Research Center, University of Washington, Seattle

Contact information:
JB: julienb@uw.edu
AGT: alecgt@uw.edu
ESB: etsb@uw.edu
ZH: zaid@uw.edu
AS: ashojaie@uw.edu
AYS: azadehy@uw.edu

May 5, 2021

Abstract

Brain stimulation interventions are used for the treatment of neurological disorders such as Parkinson's and depression. Despite some success, the effects of stimulation are poorly understood, as evidenced by unknown mechanisms governing therapeutic outcomes, widespread side-effects, and inconsistent results in clinical translation. This incomplete understanding may originate from incomplete treatment of the effects of stimulation, as most studies limit their scale and subject of analysis to local effects of various stimulation protocols. Indeed, in a recent paper we demonstrated that neural stimulation impacts the brain at the network scale; here, we leverage advances in neural interfaces and interpretable machine learning to reveal that the cortical network is the primary controlling factor of the response to stimulation. Using optogenetic stimulation and micro-electrocorticography (µECoG) recording in two awake rhesus macaques, we induce and record widespread functional connectivity changes over large-scale sensorimotor cortical networks. We then use a novel nonparametric modeling framework to predict these changes from the stimulation protocol and from characteristics of the underlying cortical network. We observe that the stimulation protocol only explains a small portion of the network response to stimulation while the network structure characteristics explain much more of the response, implicating the underlying network as the primary mediator of the response to stimulation. We extract the relationships linking the stimulation and network characteristics to the functional connectivity changes and observe that the mappings diverge over frequency bands and successive stimulations. Finally, we uncover shared processes governing real-time and longer-term effects of stimulation, demonstrating that all neural stimulation interventions must consider both timescales for their effects to be fully understood. The insights and developments generated in this work represent a paradigm shift for the field of neural stimulation.

Introduction

Brain stimulation has demonstrated strong potential for treating neurological disorders. Stimulation interventions have been approved to treat Parkinson's, depression, epilepsy, and dystonia, among other disorders [26]. Stimulation for the treatment of disorders can be broadly categorized into two approaches: modulation of real-time neural activity, and modification of neural connections. Interventions targeting real-time activity are already in clinical use, with deep brain stimulators (DBS) alone implanted in over one hundred thousand individuals to date [34]. Interventions targeting neural connectivity have encountered more difficulties in their transition to clinical practice, as clinical trials [32] and *in vivo* studies [20, 23, 36, 39] have reported inconsistent results. Very little is known about the underlying mechanisms governing the therapeutic outcomes of either approach.

Prior research has primarily analyzed the effects of various stimulation protocols and has largely ignored the role of the pre-existing brain network. For example, many targeted neural connectivity studies seek to induce various changes in neural connectivity by simply adjusting the delay in stimulation between two neural sites. This focus on the stimulation delay emerged from the discovery of Spike-Timing Dependent Plasticity (STDP), a phenomenon in which change in anatomical connectivity between two connected neurons is a function of the delay between their activation times [4–6]. While STDP was initially observed between monosynaptically connected pairs of neurons *in vitro*, later studies attempted to recreate the phenomenon *in vivo*. Some reported successful induction of targeted connectivity changes, but others reported inconsistent effects, most notably off-target (unstimulated) connectivity changes and a lack of response to the protocols [20, 23, 36, 39]. The difficulty in translation from *in vitro* models to *in vivo* models corresponds to a stark contrast in experimental settings: in real brains, connections between neurons are mediated by a highly interconnected network.

Due to this network, stimulation not only induces local activity but also activity at connected regions. Accordingly, to understand the effects of stimulation on neural activity and connectivity changes, brain stimulation must be studied at the network level. In studies of large-scale brain networks, the degree of connectivity between brain regions is often quantified as functional connectivity [31]. Functional connectivity refers to the level of coordination of activity between brain regions and as such is particularly suited for analysis of the effects of stimulation. In addition, functional connectivity has demonstrated clinical importance as aberrant functional connectivity has been implicated in neural disorders such as schizophrenia [17], depression [29], and Alzheimer's disease [18]. Despite the need for network-level studies of the effects of stimulation and the promise of functional connectivity approaches for such studies, there is a lack of analysis of the effects of stimulation on network-level functional connectivity.

To determine whether neural stimulation modifies functional connectivity at the network-level, we recently conducted an experiment in which we stimulated non-human primate cortex and recorded the changes in functional connectivity over a large-scale ($\sim 1\text{cm}^2$) cortical network [38]. To capture the effect of stimulation on both real-time modulation of neural activity and longer-term modification of neural connectivity, we calculated changes in network functional connectivity during stimulation as well as changes in resting-state functional connectivity following stimulation. We found that stimulation did impact neural activity and connectivity at a network scale, as functional connectivity changes were not limited to the stimulation sites but rather were expressed over the entire recorded network. In addition, functional connectivity changes observed during stimulation were highly correlated to those observed after the termination of stimulation. These results hint at an underappreciated involvement of the cortical network in mediating neural activity modulation and neural connectivity modification in response to stimulation, and

indicate that network-level analyses are necessary for a complete understanding of the effects of neural stimulation.

Here, we investigate the precise role of the cortical network in mediating the response to stimulation using the data collected in the previously detailed study [38]. We record from a large-scale ($\sim 1\text{cm}^2$) area of non-human primate primary sensorimotor cortex and stimulate using optogenetics, a light-based technique, enabling us to simultaneously record and stimulate without artifact. We estimate functional connectivity over the cortical network by use of the coherence metric calculated over four standard frequency bands. We study functional connectivity changes (FCC) in the context of modulation of real-time neural activity and modification of resting-state neural connectivity by computing the stimulated-state functional connectivity changes (SS-FCC) and resting-state functional connectivity changes (RS-FCC), respectively. We use an interpretable hierarchical additive modeling framework [40] to predict held-out test data for these FCC. We first quantify the degree to which the stimulation protocol controls FCC by using the stimulation protocol as input for the predictive model and SS-FCC and RS-FCC as separate outputs. We observe that while the protocol has some predictive power for both types of FCC, it is insufficient for accurate prediction. Next, we add information of cortical network structure to the model by calculating graph characteristics of the baseline functional connectivity network and incorporating them in our prediction. We observe that the inclusion of network structure information drastically improves prediction, implicating the cortical network in mediating both SS-FCC and RS-FCC. Finally, we extract the mappings from our protocol and network features to FCC and observe that they are modified by continuous stimulation. By leveraging a cutting-edge neural interface and statistical model to conduct a network-level analysis of the effects of neural stimulation, we elucidate the role of the pre-existing cortical network in mediating the response to stimulation, a key step in informing future neural stimulation interventions.

Results

Large-scale stimulation and recording of non-human primate sensorimotor cortex

We performed a series of experimental sessions consisting of stimulating and recording rhesus macaque cortex. To stimulate neurons without causing a recording artifact we used optogenetic stimulation. Optogenetic stimulation refers to the delivery of light to neurons which have been rendered light-sensitive by viral-mediated expression of opsins, or light-sensitive proteins, in their cell membranes. To record over a large-scale ($\sim 1\text{cm}^2$) area of the cortex while allowing light penetration for optogenetic activation, we used a transparent micro-electrocorticography (µECoG) array which recorded local-field potentials (LFP) from the surface of the primary sensorimotor cortex [33].

For each experimental session we randomly chose two locations corresponding to two electrodes of the µECoG array to be stimulated. The stimulation consisted of alternating 5 ms laser pulses delivered to the stimulation sites, temporally offset by a session-specific delay (Fig. 1) and repeated every 200 ms for 10 minutes. The session-specific delays were randomly chosen at the beginning of each experimental session to be 10 ms, 30 ms, or 100 ms. Locations of stimulation, the delay between the stimulation of the first and second sites, and the location and orientation of the µECoG array placement varied between experimental sessions but were consistent over individual sessions.

The set of experimental sessions spanned multiple days and two rhesus macaques. We recorded neural activity during 10-minute stimulation periods, which we term "stimulation blocks," as well as during 5-minute periods before and after each stimulation period, which we term "resting-state blocks." Each experimental session consisted of alternating resting-state blocks and stimulation blocks, with 6 total resting-state blocks and 5 total stimulation blocks.

From the neural data recordings we calculated the FCC for each pairwise connection between recorded electrodes of an experimental session. We quantified FCC with the coherence metric for the Theta (4-7 Hz), Beta (12-30 Hz), Gamma (30-70 Hz), and High-gamma (70-200 Hz) bands. We calculated the change from a resting-state block to the following stimulation block which we refer to as SS-FCC, and calculated the change from a resting-state block to the following resting-state block which we refer to as RS-FCC.

Stimulation delay is a poor predictor of functional connectivity change between stimulation sites

Studies have shown that change in anatomical connectivity between two monosynaptically connected neurons is a function of activation times, or stimulation delay, between the neurons [4–6], a phenomenon termed STDP. As STDP describes structural changes in connectivity but not real-time modulation of neural activity, we evaluated how well stimulation delay could predict RS-FCC between the sites of stimulation.

The standard function relating stimulation delay to corresponding connectivity change according to the STDP hypothesis is an exponential delay function [7, 8]. However, here we recognize that our experimental setting bears little resemblance to the controlled monosynaptic studies yielding such a curve. As such, we did not enforce an exponential delay function but rather tested whether varying the delay had an effect on RS-FCC by fitting a linear model with delay as categorical input and RS-FCC as continuous output (Fig. 2a). Since STDP is a process governing pairwise connectivity change, we modeled only RS-FCC calculated between stimulation sites, and did not model RS-FCC nor connectivity between unstimulated sites. We observed that delay was a significant predictor of RS-FCC in Theta and

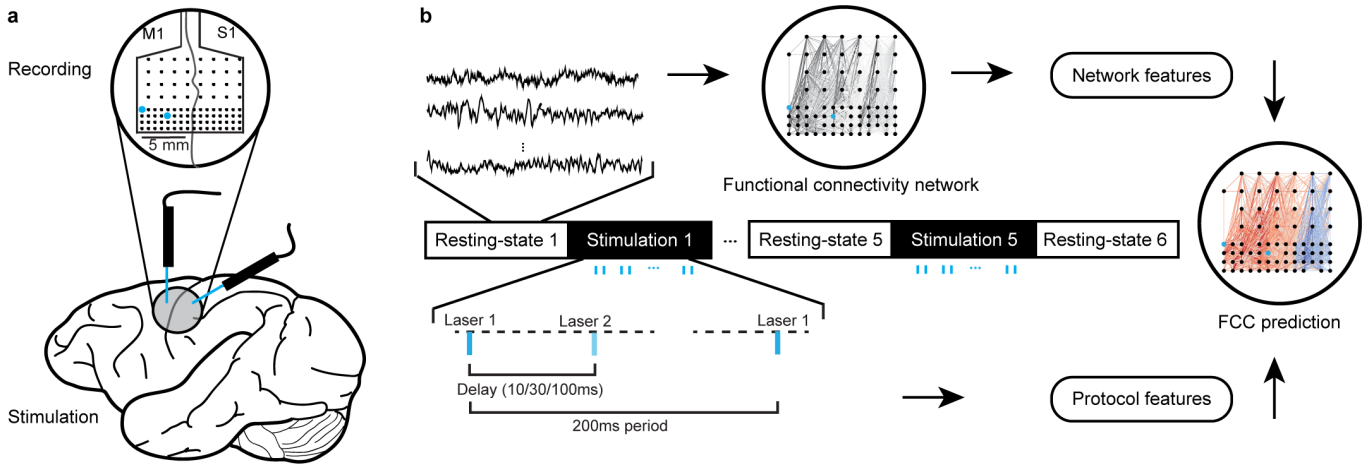

Figure 1: Schematic of the experimental setup and basic modeling pipeline. **a**, Depiction of the transparent 96-electrode µECoG array which we used to record from macaque primary sensorimotor cortex, and the approximate location of the array on the cortex. Optogenetic stimulation was applied via laser light. **b**, We recorded traces of neural LFP during the resting-state and stimulation blocks making up our experimental sessions. During the stimulation blocks, stimulation was applied in a paired pulse protocol whereby the stimulation at the second site followed the first by a session-specific delay. Network coherences were constructed for each individual recording block by calculating pairwise LFP coherences over the network. Network coherence from resting-state recordings were parameterized to generate a set of network features. Details of the protocol was parameterized to generate a set of protocol features. The sets of network and protocol features were fed to a nonparametric model to predict network-wide FCC.

Gamma bands, but not for Beta or High-gamma (p: Theta 0.016, Beta 0.125, Gamma 0.002, High-gamma 0.239; one-sided F-test with 165 observations, 2 degrees of freedom). For all frequencies we observed that delay had low explanatory power (R^2: Theta 0.050, Beta 0.025, Gamma 0.075, High-gamma 0.018).

Pairwise stimulation drives functional connectivity changes in the stimulated-state and resting-state across the entire recorded network

Previous studies investigating targeted stimulation-induced connectivity change have primarily analyzed the connectivity between the sites of stimulation. However, we recently showed in a previous publication that stimulation induces FCC which extend far beyond the stimulation sites to the surrounding network [38]. The network-wide impact of stimulation is illustrated in Fig. 2b, which plots example SS-FCC over our recording array. The connectivity change between stimulation sites (bolded) is simply one change among many in the recorded network.

We calculated FCC distributions over all networks from all experiments and observed that distributions of SS-FCC had positive mean, with similar variances over all frequency bands except Theta which had higher variance (Fig. 2c). SS-FCC was highly correlated to RS-FCC in all frequency bands, as the Pearson's correlation coefficient was above 0.6 in all frequency bands (Fig. 2d). The means of RS-FCC distributions were less offset from 0 than the means of SS-FCC distributions of the same frequency bands, and variances were similar between RS-FCC and SS-FCC distributions of the same frequency bands. These widespread FCC raise the question of explaining and predicting the stimulation-induced FCC across the entirety of the recorded network.

Nonlinear modeling to predict stimulation-induced network-wide functional connectivity changes

Motivated by the observation of network-wide FCC, we pursued an accurate and interpretable model to predict FCC not only between stimulation sites but over the whole network. While linear modeling is the standard method for statistically rigorous interpretable models, the assumption of linearity is scientifically unreasonable for statistical relationships among complex experimental measurements such as neural data. We relax this assumption through the use of a nonparametric hierarchical additive model which can identify complex nonlinear input-output relationships [40]. As this model retains the structure of summation over individual features familiar from linear modeling, it permits feature-wise investigation of the results (Fig. 3c-f). In addition, careful penalization of the estimation objective (Eq. 18) ensures that the model retains important theoretical guarantees and is encouraged to learn simple, smooth components. In the following sections we use such nonparametric hierarchical models to arrive at our results. These models outperformed linear models uniformly across frequency bands and prediction tasks (Fig. S1).

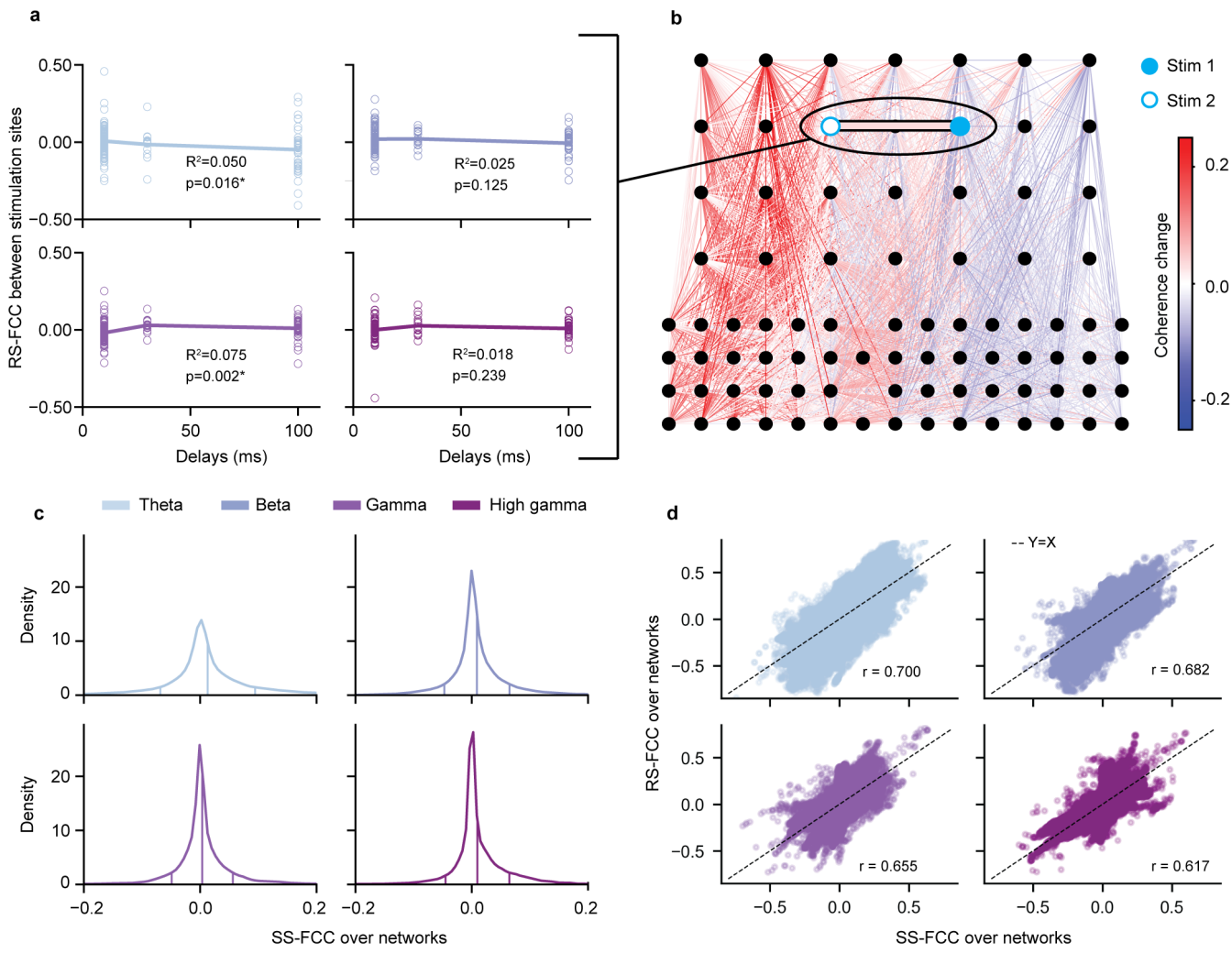

Figure 2: **a**, Delay-based prediction of RS-FCC between stimulation sites has poor explanatory power, and does not achieve statistical significance in the Beta and High-gamma. **b**, Visualization of the RS-FCC of a single example session over the µECoG array. Black circles correspond to electrodes while the edges indicate the change in coherence relative to the preceding baseline period. Optogenetic stimulation sites are indicated by light blue circles. The bolded line connecting the stimulation sites indicates the coherence change between the two sites. **c**, Distribution of the SS-FCC in each frequency band over all sessions. Internal lines indicate the mean, and one standard deviation below and above the mean. **d**, Scatter plot of SS-FCC versus RS-FCC showing high correlation.

The stimulation protocol is a poor predictor of network-wide functional connectivity changes

To test the degree to which the factors of the stimulation protocol controlled the network-wide FCC, we trained models to predict SS-FCC and RS-FCC from the stimulation protocol. We constructed a set of 8 features consisting of: stimulation delay, a categorical identifier of the time ordering of the stimulation block, the distance between a given electrode pair, the distances from the sites of stimulation to the closer or further electrode of an electrode pair, anatomical region of electrodes (Fig. 3a) (see Methods for more details of how they are calculated). Throughout the paper we refer to this set of predictive features as "protocol features." We also include a categorical animal subject indicator in the model in order to capture differences in mean between the two subjects. With this feature set we trained models with output of pairwise FCC between all recording electrodes of an experiment, over each block of all experiments. We fit individual models to predict SS-FCC or RS-FCC over individual frequency bands. We held out 30% of the data to evaluate the predictive performance of the model on unseen data.

We observed that the protocol-based model of FCC had low predictive power for both SS-FCC (R^2: Theta 0.045, Beta 0.026, Gamma 0.057, High-gamma 0.078) and RS-FCC (R^2: Theta 0.028, Beta 0.014, Gamma 0.020, High-gamma 0.014). These results indicate that while factors related to the protocol alone explain some of the variation in the observed coherence change, they are insufficient to generate practically relevant predictions.

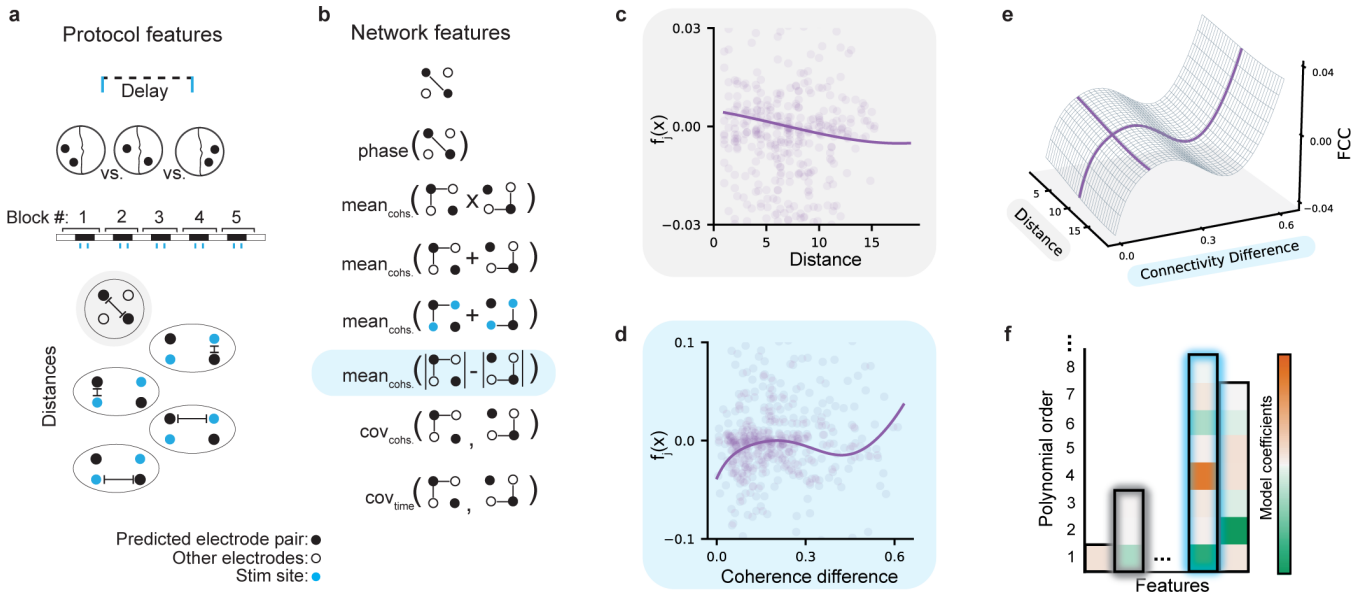

Figure 3: Feature generation and predictive modeling. **a**, Visualization of the protocol features. From top to bottom, the features are: stimulation delay, anatomical region, block number, distance, stim 1 distance to closer, stim 2 distance to closer, stim 1 distance to further, stim 2 distance to further. **b**, Visualization of the network characteristics which the network features quantify. From top to bottom, the visualizations correspond to: initial coherence, phase, length 2 path strength, coherence with network, coherence with stim sites, coherence difference, electrode covariance, and time covariance. **c-d** The nonparametric model can identify feature mappings of a variety of complexities, such as (**c**) near-linear fits such as that found for the distance feature, or (**d**) higher-order polynomials such as that found for the connectivity difference feature. $f_j(x)$ indicates the contribution of a feature to the predicted FCC. Throughout the figure, the gray shadowing indicates the distance feature, while the light blue shadowing indicates the connectivity difference feature. **e**, An example of how the feature mappings are additively combined. **f**, Individual feature mappings are represented in a basis of increasingly complex nonlinear functions. The order of this representation is selected automatically by a penalized estimation procedure, with the y-axis of this figure indicating the polynomial order. The order selection for the distance feature and connectivity difference feature are shadowed in gray and light blue, respectively.

Network structure determines stimulated-state and resting-state functional connectivity changes

Our observations that stimulation induces a change in connectivity over an entire network, in tandem with the poor predictive power of the protocol-based model, lead us to hypothesize that the effects of stimulation may be mediated by the existing cortical network. We tested this hypothesis by constructing a set of 8 features extracted from the baseline functional connectivity network. These features are: initial coherence, which is the pairwise coherence between an electrode pair; phase, which is the absolute value of difference between electrode spectral phases; length 2 path strength, which is the average strength of a connection between an electrode pair passing through another recording site; coherence with network, which is the average coherence between a given electrode pair and the rest of the network; coherence with stim sites, which is the average coherence between a given electrode pair and the stimulation sites; coherence difference, which is the average absolute difference in connection strength between each element of the electrode pair and the recorded network; electrode covariance, which is the average covariance of an electrode pair's coherence to the other electrodes in the network over time; and time covariance, which is the average covariance across time series between a given electrode pair's coherence and the others in the network (Fig. 3b) (see Methods for details of how these are calculated). These features are primarily derived from graph theoretic measures [10, 11, 13, 14] which have been used with success in past computational neuroscience analyses [22, 24, 28, 35]. Throughout this paper we refer to this set of features as "network features."

We observed that the model predictions on the test set was dramatically improved compared to the protocol-only model for each frequency band for both SS-FCC (R^2: Theta 0.240, Beta 0.130, Gamma 0.221, High-gamma 0.323) (Fig. 4) and RS-FCC (R^2: Theta 0.218, Beta 0.087, Gamma 0.138, High-gamma 0.218) (Fig. 6a). This improvement in performance across all frequency bands and for both stimulated-state and resting-state contexts suggests that properties of the local network at baseline may be highly informative with respect to prediction of

stimulation-induced changes.

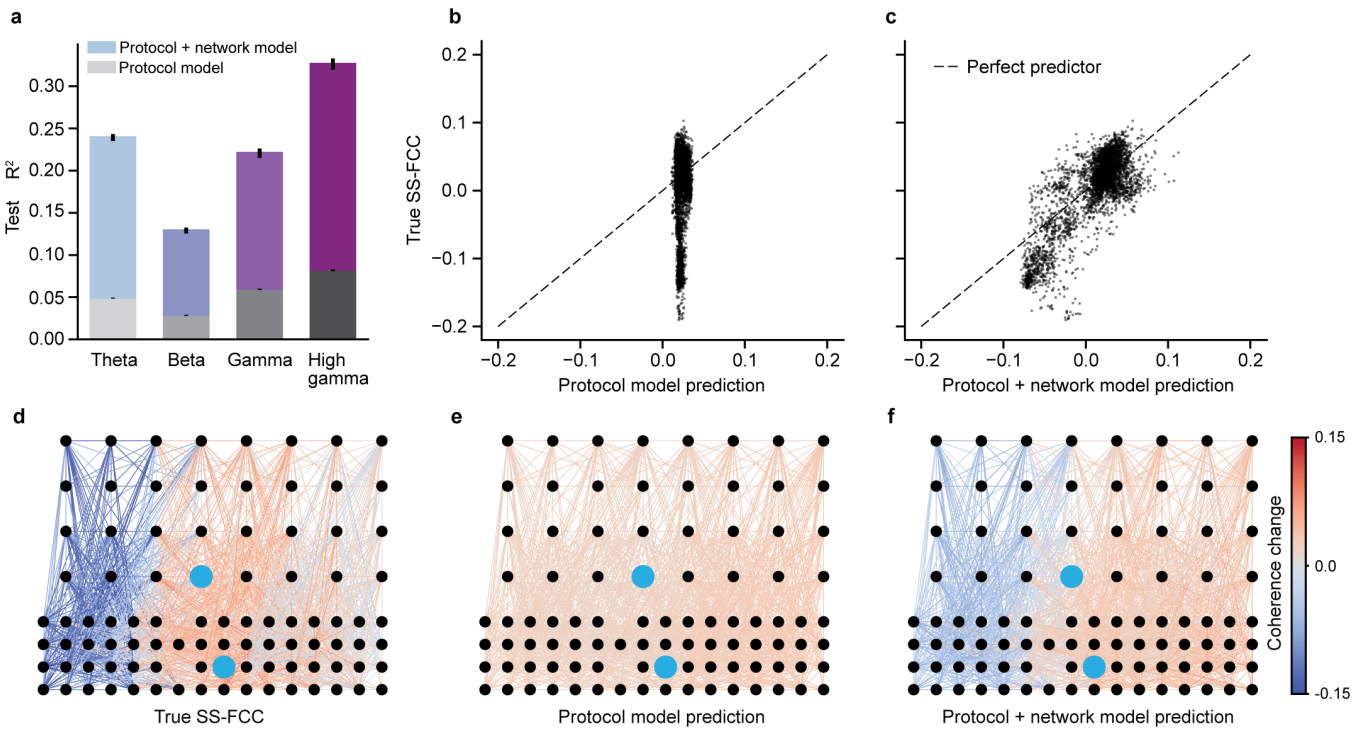

Figure 4: A model trained to predict stimulation-induced FCC over a cortical network based on protocol features alone is outperformed by one which incorporates information about existing network connectivity. Results in this figure correspond to prediction quality for SS-FCC; results are similar but slightly worse for RS-FCC. **a**, Bar graph indicating r-squared accuracy on held-out data. Grayscale bars indicate accuracy of models trained on only protocol features and colored bars indicate accuracy of models trained on both protocol and network features. Error bars indicate standard deviation. **b-f**, Results of prediction of a representative experimental session from the protocol versus the protocol-and-network models. **b**, Scatter plot of the protocol model's predictions of network SS-FCC versus the actual changes observed during that experimental session. **c**, Scatter plot of the protocol-and-network model's predictions of SS-FCC versus the actual changes observed during that experimental session. **d**, Example SS-FCC over a cortical network, visualized over the µECoG array used for recording. Black circles indicate electrodes of the array, while lines indicate coherence change between electrode pairs. The observed coherence changes correspond to the y-axis of **b** and **c**. **e**, The protocol model's SS-FCC prediction, corresponding to the x-axis of **b**. **f**, The protocol-and-network model's SS-FCC prediction, corresponding to the x-axis of **c**.

Individual network features are more important than individual stimulation protocol features for predicting functional connectivity changes

We quantified the relative importance of each feature through the decrease in accuracy of prediction on the test set when that feature was excluded from the data. The statistical uncertainty associated with the importance was quantified through a resampling procedure [19] (see Methods for details).

Using this procedure, we show the importances of each feature per frequency band in Fig. 5a for SS-FCC and in (Fig. S2) for RS-FCC. Four of the five most important features for SS-FCC prediction and each of the five most important features for RS-FCC prediction were network features. Features appearing in the top five of both contexts were time covariance, coherence difference, and initial coherence.

Feature mappings display diverging trends across frequency bands

We decomposed the nonlinear model into feature mappings representing the estimated mapping between each individual feature and the change in coherence. The statistical uncertainty associated with the estimated nonlinear mapping from the feature to the response was quantified through a resampling procedure [19] and representative feature mappings were obtained through point-wise interpolation of the resampled feature mappings (see Methods for details).

We plot the SS-FCC feature mappings of some important features as well as others of interest in Fig. 5b-f. We observed that feature mappings were generally more complicated and nonlinear for network features than for protocol features such as distance, for which FCC decreased nearly linearly in the high-confidence regions of all frequency

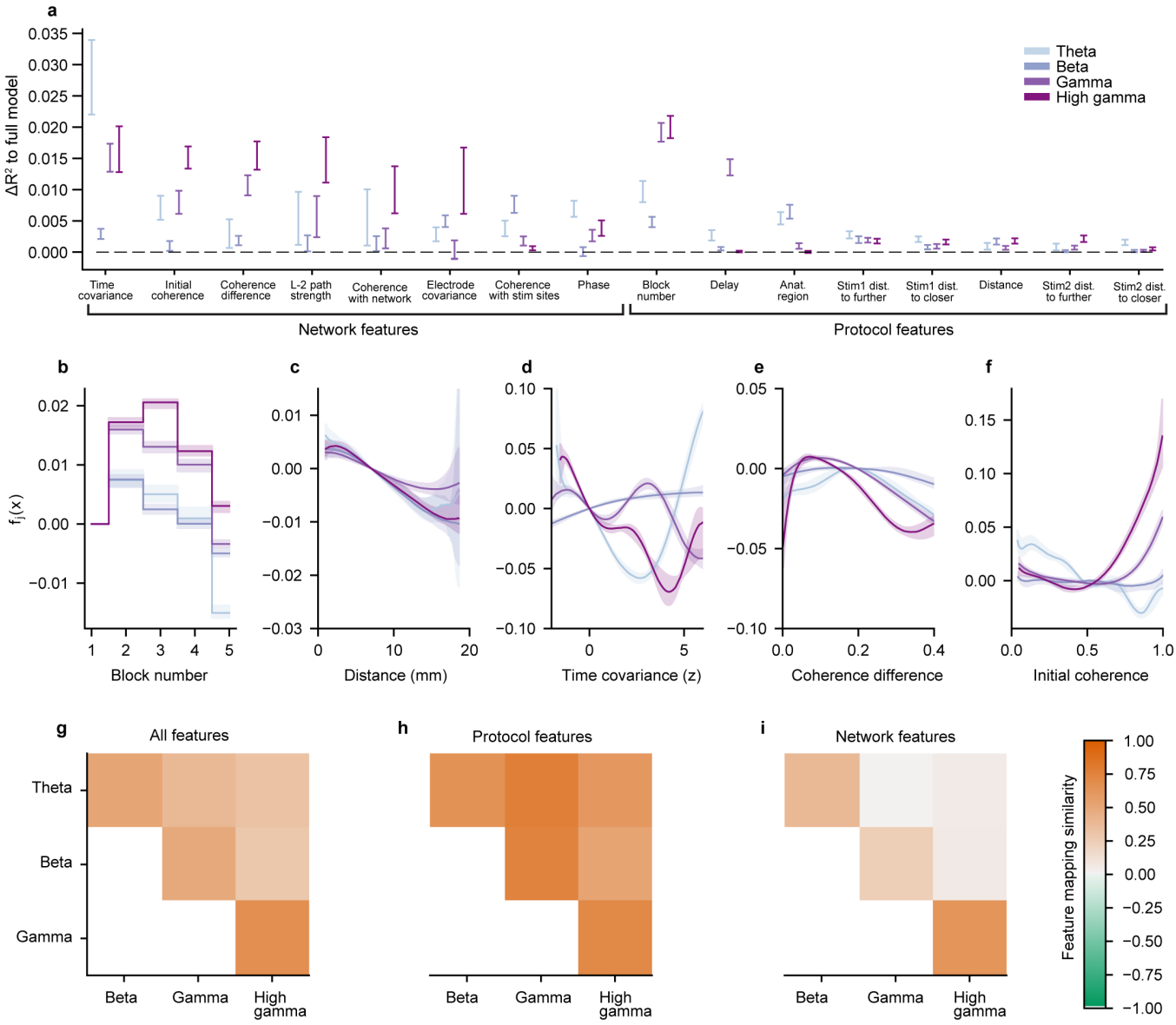

Figure 5: Feature-wise analysis of the nonlinear additive model. Results in this figure correspond to the SS-FCC model; results were similar for the RS-FCC model. **a**, Feature importances as measured by the decrease in model accuracy when the feature was omitted. Features were partitioned into the protocol and network categories, then sorted by average importance over all frequency bands. Line ranges indicate 95% confidence intervals. **b-f**, Feature mappings for the **b** time-ordered block number, **c** electrode pair distance, **d** time covariance (z-scored), **e** coherence difference, and **f** initial coherence features. $f_j(x)$ indicates the contribution of a feature to the predicted FCC. Shading indicates 95% confidence intervals. **g-i**, Average cosine similarity of the component functions between each pair of frequency bands, over **g** all features, **h** protocol features, and **i** graph features.

band models (Fig. 5c). We also observed that in the higher frequency bands, electrodes which had a higher initial coherence tended to strongly increase their coherence when stimulated (Fig. 5f). Additionally, we found a small effect of increasing SS-FCC for the first two or three temporal blocks depending on frequency band, followed by decreasing SS-FCC in later blocks (Fig. 5b). We evaluate the temporal characteristics of the FCC in more detail in a later section.

We summarized the similarity of all features, protocol features, and network features by calculating the cosine similarity of the feature mappings (Fig. 5g-i). Generally, we observed that the mappings from protocol features to SS-FCC were fairly consistent across frequency bands, while the mappings from network features to SS-FCC varied sharply between frequency bands. Only the network feature mappings of Gamma and High-gamma were similar.

Feature mappings are similar for stimulated-state and resting-state functional connectivity changes

We calculated the similarity of feature mappings between SS-FCC and RS-FCC and observed that they exhibited high similarity in all frequency bands (Fig. 6b). Notably, the groups with the highest similarity between SS-FCC and RS-FCC feature mappings were the network features in the Theta and High-gamma bands. We plot representative SS-FCC and RS-FCC feature mappings of five features from the Theta and High-gamma models in Fig. 6c-d.

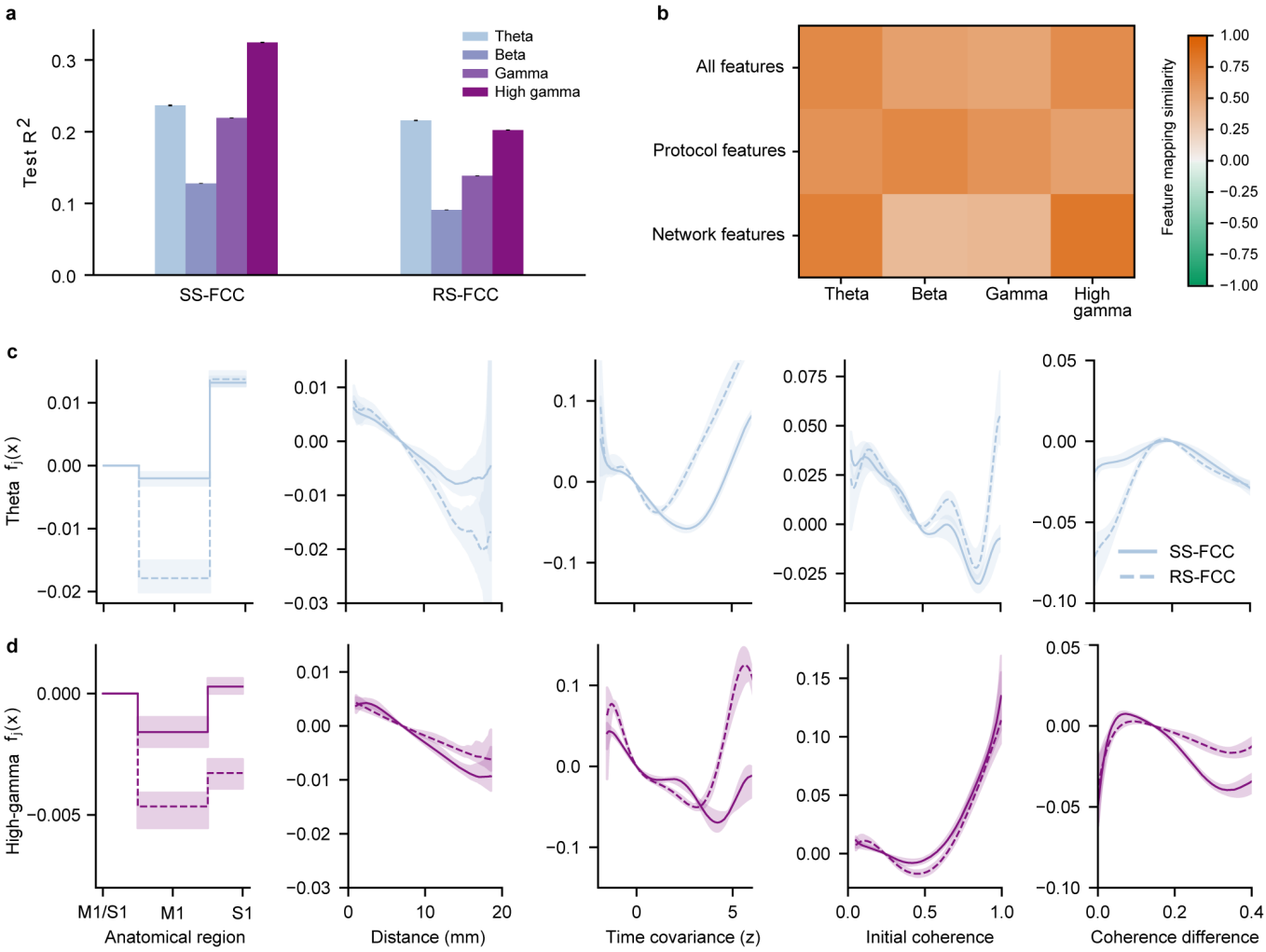

Figure 6: **a**, Comparison of the predictive accuracy for nonlinear additive models of SS-FCC prediction versus RS-FCC prediction. Error bars indicate standard deviation. **b**, Heatmap of average cosine similarity between feature mappings of SS-FCC and RS-FCC over frequencies. **c-d**, Comparison of feature mappings for SS-FCC and RS-FCC models for the **c** Theta band and **d** High-gamma band. The features from left to right are anatomical region of electrodes, electrode pair distance, time covariance (z-scored), initial coherence, and coherence difference. Shading indicates 95% confidence intervals.

Repeated stimulation modifies mappings from features to functional connectivity changes

The modeling approach employed thus far constrained the learned feature mappings to be constant across all temporal blocks of an experimental session. The influence of the block's time location was represented by a learned global shift in the FCC between all electrode pairs measured in that block. We hypothesized that successive stimulation blocks within an experiment had a more complex effect on FCC than a simple average shift, and rather modified the actual feature mappings over the time course of stimulation. We tested this hypothesis by allowing the nonlinear relationship between each feature and the response to vary across the five stimulation blocks by adding an interaction term between the block number feature and all other features. Rather than averaging over potential heterogeneity in the network dynamics resulting from the accumulated effects of stimulation, this approach allowed for the investigation of a time-course of smoothly varying predictors of coherence changes.

In line with the hypothesis that repeated stimulation indeed produces such heterogeneity across blocks within a session, we found a substantial improvement in prediction quality with the time-varying model for both SS-FCC (R^2 Theta 0.426, Beta 0.247, Gamma 0.415, High-gamma 0.527) (Fig. 7a) and RS-FCC (R^2 Theta 0.410, Beta 0.251,

Gamma 0.359, High-gamma 0.436) over the static model. We also calculated the cosine similarities of feature mappings for each block of the time-varying SS-FCC model and found that successive stimulation results in the mappings diverging from each other, as their similarity decreased over subsequent blocks (Fig. 7d). These results indicate that the relationship between features and connectivity change evolved dynamically over the course of repeated stimulation.

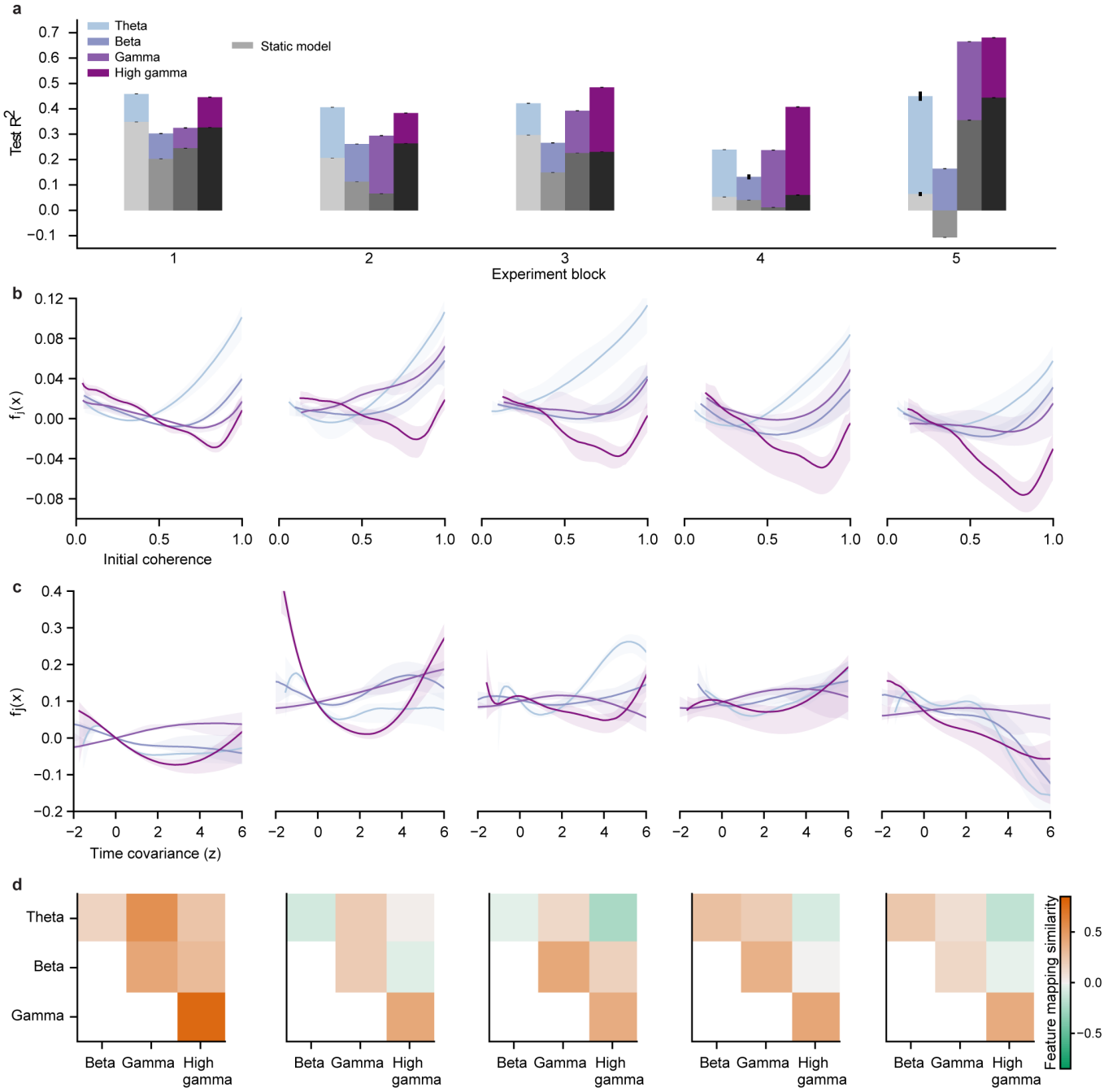

Figure 7: Feature mappings mediating FCC over a network are dependent on the stimulation history. Results in this figure correspond to prediction of SS-FCC; results are similar for prediction of RS-FCC. **a**, Bar graphs indicating r-squared accuracy on held-out data of the time-varying and static model predictions. Greyscale bars indicate the static model while colored bars indicate the time-varying model. The time-varying models outperformed the static model over all frequencies and experiment blocks. Error bars indicate 95% confidence intervals. **b-c**, Example feature mappings of two network features, **b** initial coherence and **c** time covariance (z-scored) used in the time-varying model. We observed that the feature mappings change over successive experiment blocks. Shading indicates 95% confidence intervals. **d**, Heatmaps of similarities of feature mappings between models trained for different frequency bands. Similarity was quantified by feature-averaged cosine similarity of the feature mappings. We observed that while feature mappings were similar at the onset of the stimulation, they diverged over the time-course of stimulation.

Discussion

Here for the first time we investigate the role of the cortical network in mediating stimulation-induced FCC. We leverage advances in neural interface technology and statistical modeling to induce and model large-scale SS-FCC and RS-FCC. Previous analyses of the effects of neural stimulation were limited to stimulation parameters; here, we augment the scope by additionally considering the effects of the pre-existing network structure and by analyzing FCC over a large-scale ($\sim 1\text{cm}^2$) network. We show that the effects of stimulation on both real-time neural activity modulation and longer-term connectivity modification are widespread over the network, and that the underlying network structure mediates these effects. Our findings may explain inconsistent results historically reported *in vivo* and in clinical trials and pave the way for increased understanding and development of neural stimulation interventions.

The influence of the stimulation protocol

While brain stimulation can be delivered to a local area, the highly-interconnected nature of the brain means that any local region tightly modulates and is modulated by other regions. Any analysis of the effects of stimulation which only considers the stimulation protocol thus prioritizes the direct, local effect of stimulation over the mediation of the response by the brain network.

For some applications of neural stimulation, especially those which target real-time modulation of neural activity, the network-agnostic, protocol-focused approach has demonstrated strong results. For example, DBS devices are able to ameliorate symptoms of Parkinson's and essential tremor by stimulating with a pre-set stimulation protocol [26]. However, these interventions cause many reported side effects [43], and the mechanisms underlying the side affects and governing their therapeutic outcomes are not understood.

For neural stimulation interventions targeting connectivity change *in vivo*, the protocol-focused approach has been found lacking. The dominant framework for stimulation-induced targeted connectivity change was developed *in vitro*, a setting in which the cortical network is largely nonexistent and the stimulation protocol can be extremely precise. Attempts to translate the *in vitro* findings to *in vivo* have been largely inconsistent, with some reporting promising results [16] and others reporting effects such as off-target (unstimulated) connectivity changes and a lack of response to the protocols [20, 23, 36, 39].

In this study we observed that, in accordance with the working hypothesis of previous studies, the stimulation protocol does influence functional connectivity change. However, further analysis indicated that this influence is minor, and that a protocol-focused analysis of stimulation only explains a small fraction of the effects.

Network-level analysis of neural stimulation

We treat the cortical network as both the mediator and target of the effects of stimulation. Our focus on the network structure is facilitated by our use of novel neurotechnologies for stimulation and recording. We used a large-scale µECoG array to record from a large-scale cortical network, rather than a small random set of individual electrodes. This differed from past studies in which the inclusion of a small set of unstimulated neural sites had primarily been relegated to that of a statistical control [16, 20, 23, 36]. In addition, we used convection-enhanced delivery of the optogenetic virus to obtain widespread opsin expression [33, 42], which allowed us to stimulate many distinct sites of the sensorimotor cortex and to record neural activity during stimulation without artifact.

Some recent studies have investigated the effects of stimulation at a network level, but limited their treatment of the network in several ways. Namely, they did not consider connectivity between unstimulated regions, and did not consider network-level effects but rather analyzed pairwise connectivity independently [21, 37, 41, 44]. Here, we estimated a complete network between all recording sites by calculating their pairwise coherences. We used the coherence metric which is highly correlated to other functional connectivity metrics such as evoked response [38, 41]. In addition, unlike metrics such as evoked response, coherence does not rely on stimulation to estimate functional connectivity. This had the dual advantage of allowing us to infer connectivity during resting-state blocks and to estimate connectivity between unstimulated sites. We further used graph-theory-inspired metrics to incorporate characteristics of the whole coherence network in our edge-wise FCC predictions, instead of treating individual connections independently from each other.

The cortical network as the mediator and target of stimulated-state and resting-state functional connectivity changes

Both real-time and long-term modulation of brain networks are relevant for neural stimulation-based medical interventions. Here, we induced, estimated, and modeled network-level FCC which we observed during stimulation as well as after stimulation. The high correlation in FCC and high similarity in feature mappings between the two contexts hint at shared processes underlying SS-FCC and RS-FCC. Practically, this also means that observing stimulation-induced FCC in real-time may lead to more accurate longer-term FCC.

Interventions targeting real-time modulation often do not consider the network effects nor the longer-term effects of stimulation. Here we show that the effects of neural stimulation are not simply localized but spread out over a large

network, and persist following stimulation. These observations could provide a mechanism for the yet unexplained side effects of real-time interventions such as DBS, which include mania, psychosis, and depression [43]. Fortunately, we show that the secondary effects of network-level and persisting FCC can be predicted by taking into account characteristics of the underlying cortical network.

Our observation that the underlying network mediates RS-FCC demonstrates that interventions targeting RS-FCC must take into account the underlying network for accurate induction of connectivity changes. The historical neglect of the underlying network structure in mediating RS-FCC could provide an explanation for reports of off-target connectivity changes and inconsistent connectivity changes between stimulation sites. Stimulation-induced RS-FCC must be understood and modeled over a whole network in order to obtain a complete picture of the effects of stimulation on neural connectivity modification. As here the type of connectivity changes we study are functional connectivity changes, we point out that we cannot pinpoint the nature of these changes as either solely anatomical or activity-based, as functional connectivity is related to both [15, 27].

Our observation that SS-FCC was better predicted than RS-FCC is in line with the greater difficulties reported by neural connectivity modification approaches versus real-time modification approaches. In our experiments, this may be explained by several differences. During stimulation, the stimulation-driven changes may be larger than those observed during resting-state, corresponding to a higher signal-to-noise ratio of FCC and rendering them easier to predict. Additionally, the time between a baseline resting-state block and its corresponding stimulation block is shorter than the time between the following resting-state block, and this shorter time period of FCC may render the FCC easier to predict.

By demonstrating that incorporating characteristics of the underlying network is necessary for accurate prediction of SS-FCC and RS-FCC, we reveal that the structure of the network mediates both real-time and longer-term changes in network-level activity. As SS-FCC and RS-FCC are summarizing metrics of network-level real-time neural activity modulation and longer-term neural connectivity modification, these results indicate that both approaches must tailor their stimulation protocol to the underlying network for effective outcomes.

The evolving response to successive neural stimulation

We observed that feature mappings describing the response to stimulation change in response to repeated stimulation. This result indicates that stimulation interventions likely need to be adapted over time, as repeatedly stimulating will not repeatedly induce the same response nor even be governed by the same network rules mediating the response. This is consistent with the increased success of some closed-loop neural stimulation interventions over open-loop [25], as it acknowledges that the effects of stimulation can be made more accurate when the stimulation is informed by recent neural activity. We additionally show, however, that stimulation should be adapted to characteristics of the whole network, not only to local activity.

Interpretable predictive modeling beyond the linear paradigm

In this study we capitalized on our network-level treatment of FCC by using a data-driven nonlinear modeling approach. In doing so, we were able to not only develop an accurate predictive model but observe feature mappings displaying the relationships between the protocol and network features and FCC. These feature mappings have the dual purpose of elucidating underlying mechanisms of network-level response to stimulation, and informing future stimulation interventions.

Approaches to generate insights about relationships between input and target variables often employ simplistic linear models, while approaches to generate maximally predictive models often employ models which are difficult to interpret, termed "black-box" models. Linearity is a common assumption which prioritizes interpretability over accurate estimation . Enforcing linearity is not data-driven and therefore may not capture the complexity of the true input-output relationships. At the other extreme, prediction-focused black-box models can learn complex input-output mappings at the expense of interpretative ability such as observation of the mappings. The unique modeling approach we employ in our study displays advantages of each as it can identify complex mappings between the features and FCC while allowing the features mappings to be observed.

We found that eliminating the artificial and restrictive assumptions of the linear model yielded dramatically improved predictive performance. Particularly, we showed that as the complexity of our input features increased in complexity from simple stimulation parameters to the actual structure of the cortical network, their feature mappings exhibited strong nonlinearities and thus would not be able to be fully appreciated by a strictly linear approach.

Finally, in light of recent unsuccessful attempts to model the effects of neural stimulation with nonlinear models [44], our data-driven method for inferring feature mappings of variable complexity represents a timely contribution for principled and accurate modeling of the effects of brain stimulation.

Neuroscientific and clinical utility of our insights

While past studies have focused on the effects of stimulation protocol, here we have shown that the cortical network structure mediates the response to stimulation, both for real-time and longer-term contexts. Our findings, only

made possible by novel advances in neural interfaces and statistical modeling, represent an accurate methodology for predicting FCC and a paradigm shift for understanding the effects of neural stimulation. Our observations and methods can be used to shed light on clinical neural stimulation technologies which are used in practice but not well understood, such as deep brain stimulation. Additionally, they can be used to improve and develop novel neural stimulation interventions for neural disorders.

Methods

Two adult male rhesus macaques (monkey G: 8 years old, 17.5 Kg; monkey J: 7 years old, 16.5 Kg) were used in this study. All procedures were performed under the approval of the University of California, San Francisco Institutional Animal Care and Use Committee and were compliant with the Guide for the Care and Use of Laboratory Animals.

Large-scale neural interface

Stimulation and recording of the cortex was achieved using our large-scale optogenetic interface consisting of a semi-transparent micro-electrode array, semi-transparent artificial dura, titanium implant, widespread optogenetic expression, and laser-delivered optical stimulation [33]. Neurons in the primary sensorimotor cortex were optogenetically photosensitized via viral-mediated expression of the C1V1 opsin. We injected a viral cocktail of AAV5-CamKIIa-C1V1(E122T/E162T)-TS-eYFP-WPRE-hGH (2.5×10^{12} virus molecules/ml; Penn Vector Core, University of Pennsylvania, PA, USA, Addgene number: 35499) in the primary sensory (S1) and primary motor (M1) cortices of the left hemisphere of the two rhesus macaques using the convection-enhanced delivery technique as described in [33]. We infused 200 µL of the cocktail was over four sites in monkey G and 250 µL over five sites in monkey J. We then verified expression of the opsin by fluorescent imaging of the eYFP marker.

The chronic neural interface was implanted by performing a 25mm craniotomy over the primary sensorimotor cortices of the left hemisphere of two rhesus macaques, replacing the dura mater beneath the craniotomy with a chronic transparent artificial dura, and attaching a titanium cap over the craniotomy. During experiments we recorded neural activity from a custom micro-electrocorticography array of 96 electrodes covering 92mm^2 of the cortex. The µECoG arrays consisted of platinum-gold-platinum electrodes and traces encapsulated in Parylene-C [30]. The µECoG array was placed on the surface of the brain at the beginning of each experimental session and removed at the termination of each session.

Neural data was recorded by sampling local-field potentials at 24 kHz from the µECoG array using a Tucker-Davis Technologies system (FL, USA). We stimulated the cortex by delivering light via a fiber optic (core/cladding diameter: 62.5/125 µm, Fiber Systems, TX, USA) connected to a 488 nm laser (PhoxX 488-60, Omicron-Laserage, Germany) positioned above the array such that the tip of the fiber-optic cable touched the array.

Structure of experimental sessions

We performed a total of 36 experimental sessions over multiple days and two subjects. The experimental sessions consisted of recording while no stimulation took place, which we termed a "resting-state block," and recording during stimulation, which we termed a "stimulation block." Each resting-state block was 5 minutes long and each stimulation block was 10 minutes long. Each experimental session consisted of alternating resting-state blocks and stimulation blocks, with 6 total resting-state blocks and 5 total stimulation blocks, as a resting-state block was completed before the first and after the final stimulation block. The µECoG array was replaced on the cortex for each experimental session. Additionally, each recording block of an experimental session had the same stimulation protocol.

We stimulated two locations per experimental session which varied between experimental sessions but remained fixed throughout an experimental session. The locations were randomly chosen across the array. We imaged the stimulation locations on the array and subsequently identified the closest recording electrode to each stimulation location.

We applied stimulation using a paired-pulse protocol of 5 ms pulses separated by a session-specific delay (Fig. 1). The paired-pulses were repeated every 200 ms for the entirety of each stimulation block. The session-specific delays were randomly chosen at the beginning of each experimental session to be 10 ms, 30 ms, or 100 ms; once chosen they remained fixed for an experimental session.

Signal preprocessing

The broadband recorded surface potentials (sampled at 24 kHz) were visually inspected and electrodes which were faulty or displayed photoelectric artifact were removed from further analysis. The total amount of data removed constitutes no more than 15% of the raw time series. We then downsampled the data to 1 kHz after applying a low-pass Chebychev filter for anti-aliasing.

Signal processing to obtain a time-varying coherence network

Within each stimulation block and each resting-state block, the recorded LFP were partitioned into non-overlapping 20-second windows. All quantities computed from the raw LFP time series were computed on a per-window basis, then summarized across all windows in the block.

We computed an estimate of the coherence between all non-faulty electrodes for each 20-second window in every resting-state block and stimulation block. We denote by E the number of non-faulty electrodes and denote the multivariate LFP time series for a given window by the two-dimensional array $\boldsymbol{x}_{1:T} \in \mathbb{R}^{E\times T}$, which consists of the time-ordered concatenation of the LFP recordings $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{Et})^T$ for each electrode at time t. We estimated the multivariate spectral density function $S(\lambda) \in \mathbb{C}^{E\times E}$ for the time series $\boldsymbol{x}_{1:T}$ by Welch's method [1]. The coherence between electrodes i and j at frequency λ is denoted as $C_{ij}(\lambda)$ and computed from the estimated spectral density via $C_{ij}(\lambda) = |S_{ij}(\lambda)|^2/S_{ii}(\lambda)S_{jj}(\lambda)$. We computed the coherence at 403 linearly spaced frequencies in the interval $(0, 200)$ Hz.

Summaries over four frequency bands, 4-7 Hz (Theta), 12-30 Hz (Beta), 30-70 Hz (Gamma), and 70-199 Hz (High Gamma), were computed by averaging the coherence estimates within each band. Thus, for each session, the raw LFP time series was transformed into a sequence of matrix-valued quantities representing the average band-limited coherence in each 20-second interval. Since the coherence is symmetric and the diagonal conveys no information on pairwise behavior of the LFP electrode signals, we retained only values in the upper triangle above the diagonal for modeling and prediction.

We defined SS-FCC as the difference between the mean coherence of an electrode pair calculated during a stimulation block and the preceding resting-state block. We defined RS-FCC as the difference between the mean coherence of an electrode pair calculated during a resting-state block and the preceding resting-state block. We use FCC as a general term to refer to either SS-FCC or RS-FCC. Let $C_{ij}^{(w)}(\lambda_k)$ be the coherence between electrodes i and j at frequency λ_k in the 20-second window w, as estimated by the procedure described above. The band-limited coherence was obtained by averaging over the indices $k \in K_b$ corresponding to frequency band b: $C_{bij}^{(w)} = |K_b|^{-1}\sum_{k\in K_b} C_{bij}^{(w)}(\lambda_k)$. Within an experimental session, we denote by W_{R_ℓ}, $\ell \in \{1,2,3,4,5,6\}$ and W_{S_ℓ}, $\ell \in \{1,2,3,4,5\}$, the collections of 20-second windows belonging to resting-state block ℓ or stimulation block ℓ, respectively. Note that the additional resting-state block corresponds to a final recording block after the 5th stimulation block. The electrode coherences were summarized for each of these blocks by their mean values:

$$\tilde{C}_{bij}^{(R_\ell)} = \frac{1}{|W_{B_\ell}|}\sum_{w\in W_{B_\ell}} C_{bij}^{(w)} \tag{1}$$

$$\tilde{C}_{bij}^{(S_\ell)} = \frac{1}{|W_{S_\ell}|}\sum_{w\in W_{S_\ell}} C_{bij}^{(w)} \tag{2}$$

The SS-FCC was then computed as

$$y_{bij\ell} = \tilde{C}_{bij}^{(S_\ell)} - \tilde{C}_{bij}^{(R_\ell)} \tag{3}$$

and the RS-FCC was computed as

$$y'_{bij\ell} = \tilde{C}_{bij}^{(R_{\ell+1})} - \tilde{C}_{bij}^{(R_\ell)}. \tag{4}$$

Nonlinear modeling of stimulation-induced coherence change

We used a nonlinear additive model to explain the observed FCC in terms of features derived from both the experimental protocol and the information on functional connectivity available in the resting-state block prior to stimulation. In this framework, a single observation was denoted by the pair $(y_n, \boldsymbol{x}_n)$, $n = 1, \ldots, N$, with $y_n \in \mathbb{R}$ an FCC measurement and $\boldsymbol{x}_n \in \mathbb{R}^p$ the corresponding features. In switching to the single index n we indicate that the data was aggregated over all unique electrode pairs and all experimental blocks. The dependence on the band b was subsequently omitted for notational simplicity; the same analysis was repeated for the data in each frequency band. The FCC was modeled as

$$y_n = \beta_0 + \sum_{j=1}^{p} f_j(x_{nj}) + \varepsilon_n. \tag{5}$$

The model consists of an intercept β_0, nonlinear functions f_j controlling the impact of each feature $j = 1, \ldots, p$ on the response, and an error term ε_n.

The utility of the model derives from the nonlinearity of the feature mappings and the additive procedure through which they are combined to yield a prediction for the FCC. The additive structure of the model allows for the visualization and interpretation of individual feature-response relationships. For each feature $j = 1, \ldots, p$, this is represented by the feature mapping f_j. Moreover, the predictive impact of feature j can be measured by the decrease in prediction error on a test set relative to a model in which the feature is omitted (or equivalently, a full model is estimated while enforcing $f_j = 0$). The same analysis can be applied to groups of features. While additivity is essential for model investigation, the nonlinearity of the feature mappings enables the identification of more complex relationships than can be expressed in a linear model. Linearity is a strong modeling assumption that is both difficult to justify scientifically and, as demonstrated in the results, harmful in terms of predictive accuracy.

Feature representation of processed data

A single observation of the FC corresponds to a unique pair of electrodes and a specific recording block. The corresponding features were constructed to satisfy two objectives: first, that the subsequent analysis separates the influence of the protocol parameters from that of the network structure of the functional connectivity; second, that all information in the features was available prior to stimulation.

To satisfy the first objective, we partitioned the features into two groups: *protocol* features, which summarize aspects of the experimental setting and protocol; and *network* features, which summarize information in the electrode coherences during baseline recording. The designation "network features" derives from consideration of the estimated band-limited coherence as the adjacency matrix of an undirected, edge-weighted graph. The protocol features describe the key parameters of the experimental framework [38]. The network features were intended to serve as simple but informative summaries of the connectivity information available in recordings of baseline activity prior to stimulation. They correspond to summary statistics computed over spatial or temporal ranges, basic quantities pertaining to the estimated spectrum [3], or network features that have found previous application in graph analysis [2, 11]. Finally, a subject-level indicator was included to adjust for potential global differences in FCC between the two macaque subjects. The adjustment is included in all models, as the subject is neither an aspect of the stimulation protocol nor a property of the baseline connectivity network.

Protocol features

The protocol features were comprised of the categorical measurements `Anatomical region`, `Delay`, and `Block number`, as well as the real-valued features `Distance`, `Stim1 distance to closer`, `Stim1 distance to further`, `Stim2 distance to closer`, and `Stim2 distance to further`. `Anatomical region` indicates whether the two electrodes corresponding to a given measurement are both in region M1, both in region S1, or if one is in each region. `Delay` encodes three levels of time-delay in the pulse of the paired laser stimulation: 10, 30, or 100 ms. While the delay parameter in the stimulation protocol is real-valued, in the context of our data it is only measured at three distinct settings. We therefore chose to estimate an effect for each setting individually rather than to estimate a nonlinear function of the delay given observations at only three points. `Block number` indicates the time-order position of the experimental block in which the observation was recorded. In the "block analysis" regression design, we removed this feature and instead allowed all feature mappings to vary with the block number, thus investigating the predictive impact of allowing the entire model to evolve dynamically over the discrete time-stages of an experimental session. `Distance` measures the distance between the two electrodes. `Stim1 distance to closer`, `Stim1 distance to further`, `Stim2 distance to closer`, and `Stim2 distance to further`, encode distances between the stimulation sites and the closer and further electrodes of the electrode pair being predicted.

Network features

The network features were computed from the LFP time series recorded during the resting-state block preceding the stimulation for which FCC are being predicted. With the exception of the `Phase` feature, all of these features were derived from the tensor obtained by time-concatenation of the coherence matrices across all windows in the resting-state block.

We introduce some tensor notation prior to defining these features. The coherence tensor for a given resting-state period is a three-dimensional array, denoted as $\boldsymbol{C} \in \mathbb{R}^{E \times E \times T}$. Two-dimensional slices through the array at time t or electrode i are indicated as $C_{t\cdot\cdot}$ and $C_{\cdot i \cdot}$, respectively. One-dimensional vectors obtained by fixing an electrode i and a time point t are denoted $\vec{c}_{ti\cdot}$; vectors obtained by fixing both electrodes i and j are denoted $\vec{c}_{\cdot ij}$. The scalar coherence value at time t for electrodes i and j is denoted c_{tij}.

The `Initial coherence` indicates the mean coherence between electrodes i and j at baseline:

$$\frac{1}{T}\sum_{t=0}^{T} c_{tij} = \text{mean}(\vec{c}_{\cdot ij}) \tag{6}$$

The `Coherence with network` summarizes the average coherence between electrodes i and j and the remaining electrodes in the array. This corresponds to the normalized sum of vertex strengths of nodes i and j [11]:

$$\frac{1}{T}\sum_{t=0}^{T}\frac{1}{E}\sum_{k=0}^{E}(c_{tik}+c_{tjk}) = \text{mean}(\boldsymbol{C_i}+\boldsymbol{C_j}) \tag{7}$$

The `Coherence difference` measures the mean absolute difference in coherence between electrodes i and j to other electrodes in the network:

$$\frac{1}{T}\sum_{t=0}^{T}\frac{1}{E}\sum_{k=0}^{E}|c_{tik}-c_{tjk}| \tag{8}$$

The `Length 2 path strength` represents the average strength of length-2 paths connecting electrodes i and j. This feature is similar to standard graph clustering coefficient metrics quantifying the total edge-weight of triangle motifs that include a given node [11, 14]. However, it excludes the direct edge between electrodes i and j, which reduces correlation with the initial coherence. Alternatively, the `L=2 path strength` can be considered the unnormalized cosine similarity between the network-coherence vectors of electrodes i and j [2]:

$$\frac{1}{T}\sum_{t=0}^{T}\frac{1}{E}\sum_{k=0}^{E}c_{tik}c_{tjk} = \frac{1}{T}\sum_{t=0}^{T}\frac{1}{E}(\vec{c}_{ti\cdot})\bullet(\vec{c}_{tj\cdot}) \tag{9}$$

The `Coherence with stim sites` is the average coherence between electrodes i and j to the two optogenetic stimulation sites. Below, the indices a and b refer to the electrode sites corresponding to the laser locations for a given experiment:

$$\frac{1}{4T}\sum_{t=0}^{T}(c_{tia}+c_{tib}+c_{tja}+c_{tjb}) \tag{10}$$

The `Electrode covariance` and `Time covariance` capture the variability of the coherence measurements over the electrode array and over the total number of 20-second time windows in a resting-state block, respectively.

The `Electrode covariance` represents a time-average of the covariance between the vectors representing the connectivity of electrodes i and j to the rest of the network:

$$\frac{1}{T}\sum_{t=0}^{T}\text{Cov}(\vec{c}_{ti\cdot},\vec{c}_{tj\cdot}). \tag{11}$$

The `Time covariance` is the average across all other electrodes k of the covariance between the time series of coherence values to electrodes i and j:

$$\frac{1}{E}\sum_{k=0}^{E}\text{Cov}(\vec{c}_{\cdot ik}\vec{c}_{\cdot jk}). \tag{12}$$

Finally, the `Phase` was the only network feature not derived from the coherence tensor, since the coherence only contains magnitude information from the estimated spectral density. Writing the $(i,j)^{th}$ cross-spectral component computed within window w as

$$S_{ij}^{(w)}(\lambda_k) = |S_{ij}^{(w)}(\lambda_k)|e^{i\theta_{ijk}^{(w)}}, \tag{13}$$

and denoting by K_b the set of indices k such that λ_k is in frequency band b, the phase feature θ_{ij} in block ℓ is given by

$$\tilde{\theta}_{ij}^{(B_\ell)} = \frac{1}{|W_{B_\ell}|}\sum_{w\in W_{B_\ell}}\frac{1}{|K_b|}\sum_{k\in K_b}\theta_{ijk}^{(w)} \tag{14}$$

Outlier removal Nonlinear regression methods are sensitive to extreme outliers, which can exert highly disproportionate influence on the model estimate. We wished to achieve robustness against such outliers while minimizing perturbation of the data and subsequent analysis. Therefore we adopted a highly permissive definition, by which an observation was considered an outlier if its absolute deviation from the mean exceeded 20 times the interquartile range along any dimension (i.e. feature). Across all frequency bands and models, the maximum amount of data excluded by this procedure was less than 0.1% of the total number of observations.

Regression designs

We investigated two different configurations of the regression design matrix within our nonlinear modeling framework. For each design, the data were aggregated across all of the experimental sessions for each subject. While significant session heterogeneity suggested that even stronger prediction results were possible for models fit to individual sessions, our objective was to estimate a model that generalized well in the sense of accurate prediction on data from many sessions.

Full data. First we investigated the performance of the nonlinear model on the full data, which included all experimental blocks from all sessions. This yielded $N = 481,505$ observations of $p = 16$ features, which expanded to dimension 21 after dummy-coding of the categorical variables `Delay`, `Region`, and `Block number`, and to dimension 22 after inclusion of the categorical animal subject indicator.

Block interactions. Under the full data design, the time evolution of the response was limited to global shifts corresponding to the categorical feature `Block number`. We subsequently investigated the impact of allowing the shape of the feature mappings to vary with the block number, thus expanding the model's capacity to capture time variation in feature-response relationships over the course of repeated stimulation. This constituted an interaction design, whereby the categorical feature `Block number` was removed, and the remaining features were each augmented with interaction terms for binary indicators that denoted whether the measurement was made in each of blocks 2, 3, 4, and 5. The number of observations remained $N = 481,505$, while the number of features expanded to $p = 94$.

Estimation of the nonlinear additive model

We used a train-test paradigm to assess the predictive performance of the nonlinear model on unseen data. For each frequency band and regression design, the data were split uniformly at random such that 70% of the observations were assigned to the training set and the remaining 30% were assigned to the test set. All model parameters and hyperparameters were selected using data from the training set only. Prior to estimation of the model parameters, the real-valued features were standardized such that they have mean zero and unit standard deviation. Whenever the data were split between train and test sets, the parameters of the linear transformation corresponding to standardization were computed from the training data and applied to the test data.

The nonlinear feature mappings f_j, $j = 1, \ldots, p$, in Eq. (5) were represented in the basis of polynomial functions

$$f_j(x) = \beta_{j1}x + \beta_{j2}x^2 + \cdots + \beta_{jK}x^K. \tag{15}$$

The complete set of parameters to be estimated was thus $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$, where $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jK})^T \in \mathbb{R}^K$. The *order* of the polynomial representation for f_j is defined as the maximum integer value $k \in 1, \ldots, K$ for which β_{jk} is nonzero. The smoothness of f_j is determined by the order k of its polynomial representation and the magnitude of the coefficients $\beta_{j1}, \ldots, \beta_{jk}$. Numerical implementation requires the specification of an upper bound K, corresponding to truncation of the infinite basis at some maximum order. We used $K = 10$ in all experiments; this decision was justified by the fact that the parameter estimates decay either to negligible values or to exactly zero before reaching order $K \leq 10$, as a result of the hierarchical penalization of the model coefficients.

All categorical features were dummy-coded such that the C categories for each feature are represented by $C - 1$ indicator variables. Polynomial expansion of these variables generates identical columns and thus rank-deficiency in the design matrix, which leads to instability in the optimization routine. This problem was avoided by truncating the polynomial expansion of all categorical variables to order 1 rather than order K. There is no resulting loss of generality from this truncation. Rather, this expanded the framework to allow for simultaneous estimation of smooth nonlinear functions of continuous features and discrete functions (that can be visualized for example, as a step function) of categorical features.

Automatic order selection via hierarchically penalized estimation

Given an observation of the features $\boldsymbol{x}$, the model prediction of the SS-FCC or RS-FCC is given by

$$\hat{f}(x) = \beta_0 + \sum_{j=1}^{p} f_j(x_j). \tag{16}$$

We used the standard square loss on the training set D^{train}

$$\mathcal{L}(\beta) = \frac{1}{|D^{\text{train}}|} \sum_{(x,y)\in D^{\text{train}}} \left(y - \hat{f}(x)\right)^2 \tag{17}$$

as the model goodness-of-fit criterion. Direct optimization of model fit on a training set over a sufficiently complex model class is well-known to yield estimates that perform poorly on unseen data, an issue commonly known as overfitting. We avoided this problem by adoption of a penalized estimation framework specifically designed to control the magnitude of the parameter estimates and automatically select the order of each feature mapping in the model. This approach augments the standard square loss with a penalty term $\Omega_j(\boldsymbol{\beta}_j; \alpha, \lambda)$ for each feature j, defined as

$$\Omega_j(\boldsymbol{\beta}_j; \alpha, \lambda) = \alpha\lambda \sum_{k=1}^{K} w_k \left(\sum_{i=1}^{N} \left(\beta_{jk} x_{ij}^k\right)^2 + \cdots + \left(\beta_{jK} x_{ij}^K\right)^2 \right)^{\frac{1}{2}} + (1-\alpha)\lambda \left(\sum_{i=1}^{N} \left(\beta_{j1} x_{ij}\right)^2 + \cdots + \left(\beta_{jK} x_{ij}^K\right)^2 \right)^{\frac{1}{2}}, \tag{18}$$

with hyperparameters $\lambda \in \mathbb{R}_+$ and $\alpha \in [0, 1]$. The the weights $w_k = k^3 - (k-1)^3$ were chosen to satisfy theoretical criteria that guarantee statistical convergence of the estimator over a class of smooth functions [40].

The model parameters were estimated by solving

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}) + \sum_{j=1}^{p} \Omega_j(\boldsymbol{\beta}_j; \alpha, \lambda). \tag{19}$$

We used a block coordinate descent algorithm as implemented in the R package `HierBasis` [40].

The first term in the penalty $\Omega_j(\boldsymbol{\beta}_j; \lambda, \alpha)$ induces hierarchical sparsity in the estimate of $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jK})^T$, while the second term in (18) shrinks the magnitudes of the estimated coefficients towards zero. Hierarchical sparsity guarantees that if an estimated coefficient $\hat{\beta}_{jk} = 0$, then all higher-order coefficients for that feature $\hat{\beta}_{jk'} = 0$, with $k < k' \leq K$; this is equivalent to selecting an order $k-1$ representation for the feature mapping f_j. The selection procedure is thus *automatic* and *data-driven* in that it emerges as a direct mathematical consequence of the penalized estimation framework, which seeks the best model fit to the data under the structural constraints imposed by $\Omega_j(\boldsymbol{\beta})$.

Cross-validation for model hyperparameters

The penalized loss function requires specification of two regularization hyperparameters $\lambda \in \mathbb{R}_+$ and $\alpha \in [0, 1]$, which control the overall regularization strength and tradeoff between terms inducing hierarchical and feature-wise sparsity, respectively. The parameter α was selected over the range from 0 to 1, inclusive, in increments of 0.1. The parameter λ was selected over 100 log-linearly-spaced values between $\lambda_{\max}$ and $\lambda_{\max} \times 10^{-5}$.

The value $\lambda_{\max}$ was selected such that all estimated feature mappings are equal to zero for every value of investigated α; it was obtained by a backtracking line search algorithm [12]. This algorithm requires a gross upper bound λ_0, defined as any value of $\lambda > 0$ such that $\widehat{\boldsymbol{\beta}} = 0$ for all grid values of α. It was found manually; we found that $\lambda_0 = 0.1$ sufficed for all designs and frequency bands.

We selected the pair (α^*, λ^*) by first finding the (α, λ) pair minimizing the average R^2 on the validation set over 5-fold cross-validation on the training data. We selected α^* as the α-coordinate of this pair, and selected λ^* as the largest value of λ such that the mean validation R^2 of (α^*, λ) was within one standard error of the mean validation R^2 at (α^*, λ^*). This follows the "one standard error" strategy for one-dimensional cross-validation [9] and represents a conservative approach to regularization corresponding to our preference for smoother feature mapping estimates.

Performance measure. We quantified model performance by predictive accuracy on the test set, as measured by the coefficient of determination,

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y}_i)^2}, \tag{20}$$

where $\hat{y}_i$ was the model prediction after cross-validation and estimation on the training set and $\bar{y}_i$ was the mean of the regression target on the test set.

Stability, importance, and similarity of learned feature mappings

Resampling for feature stability

We took a resampling approach to assess the stability of the estimated feature mappings of the nonlinear model. Following the subsampling heuristic of [19], we drew a subsample of size $\lfloor N/2 \rfloor$ from the data without replacement and ran the full estimation procedure, including hyperparameter selection, on the subsampled data. The procedure was repeated for 100 independent trials, resulting in 100 estimated functions for each feature. The feature-wise variability in the nonlinear additive feature mappings was then assessed pointwise, by computing upper and lower quantiles for the range of function values at each point in a fine grid.

A predictive measure of feature importance

Traditional tests of statistical significance are not available for the coefficients of the nonlinear additive model as the asymptotic distribution of the penalized estimator is not known. Instead, we assessed importance of a feature or group of features by evaluating the difference in predictive performance on the test set between the full model and a model fit without the parameter or parameter group of interest. Removing the feature or feature group is equivalent to estimating the full model under the constraint that their corresponding parameters are all equal to zero. The comparison can thus be viewed as between the full model and one in which a hypothesis of null response is enforced for the feature or group of interest. When estimating the reduced model, we used the regularization parameters (α^*, λ^*) obtained by cross-validation on the full data.

Quantifying feature similarity across frequency bands

Let $\hat{f}_j^a$ and $\hat{f}_j^b$ be the estimated feature mappings for feature j on the data corresponding to frequency bands a and b, respectively. We computed a quantitative measure of their similarity,

$$s_j^{ab} = \frac{\langle f_j^a, f_j^b \rangle}{\|f_j^a\| \, \|f_j^b\|}. \tag{21}$$

The quantity s_j^{ab} is the cosine similarity of the estimated feature mappings, considered as elements of a common Hilbert space of functions. By definition, $s_j^{ab} \in [-1, 1]$. Similarity increases as s_j^{ab} approaches 1 or -1 (which indicates f_j^a and $-f_j^b$ are perfectly similar) and is minimized at 0.

The inner product $\langle f_j^a, f_j^b \rangle = \int_{\mathcal{X}_j} f_j^a(x) f_j^b(x) dx$ and norm $\|f_j\| = \langle f_j, f_j \rangle$ are given by integrals whose domain $\mathcal{X}_j$ depends on the feature j. For real-valued features, we took $\mathcal{X}_j$ to be the interval $[-5, 5]$, which after standardization corresponds to the range of all observed measurements within 5 standard deviations of the mean. Due to the polynomial representation of the nonlinear feature mappings, these could be computed exactly.

Acknowledgements

We thank Daniel Silversmith for his help with data collection and Philip Sabes for his laboratory in which the data was collected. This work was supported by the Big Data for Genomics and Neuroscience Training Grant (NIH 5T32LM012419, JB and AGT), the Center for Neurotechnology (NSF ERC 1028725, JB and ESB), the Washington National Primate Research Center (NIH P51 OD010425, AYS), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NIH K12HD073945, AYS), NSF grant DMS-2023239 (ZH), NSF grant DMS-1722246 (AS), NSF grant DMS-1915855 (AS), and NIH grant R01GM114029 (AS).

Author Contributions

JB and AGT drafted the manuscript. JB and AYS conceptualized the study. AYS performed the experiments. JB processed the data. AGT performed the nonparametric modeling. All authors revised and edited the manuscript.

Competing Interests

The authors declare no disclosures.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. Welch, P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* **15,** 70–73 (1967).
2. Salton, G. & McGill, M. *Introduction to Modern Information Retrieval* (McGraw-Hill, Inc., Auckland, 1983).
3. Brockwell, P. J., Davis, R. A. & Fienberg, S. E. *Time Series: Theory and Methods* (Springer Science & Business Media, 1991).
4. Bliss, T.V.P. and Collingridge, G. A synaptic model of memory: LTP in the hippocampus. *Nature* **361,** 31–39 (1993).
5. Markram, H., Lübke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275,** 213–215. ISSN: 00368075 (1997).
6. Bi, G.-q. & Poo, M.-m. Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *J. Neurosci* **18,** 1–9. ISSN: 0270-6474 (1998).
7. Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A. & Poo, M. M. A critical window for cooperation and competition among developing retinotectal synapses. *Nature.* ISSN: 00280836 (1998).
8. Song, S., Miller, K. D. & Abbott, L. F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience.* ISSN: 10976256 (2000).
9. Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning* **10** (Springer Series in Statistics New York, 2001).
10. Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science.* ISSN: 00368075 (2002).
11. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 3747–3752 (2004).
12. Boyd, S. & Vandenberghe, L. *Convex Optimization* (Cambridge University Press, 2004).
13. Milo, R. *et al.* Superfamilies of Evolved and Designed Networks. *Science.* ISSN: 00368075 (2004).
14. Onnela, J. P., Saramäki, J., Kertész, J. & Kaski, K. Intensity and coherence of motifs in weighted complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **71** (2005).
15. Waites, A. B., Stanislavsky, A., Abbott, D. F. & Jackson, G. D. Effect of prior cognitive state on resting state networks measured with functional connectivity. *Human Brain Mapping* **24,** 59–68. ISSN: 10659471 (2005).
16. Jackson, A., Mavoori, J. & Fetz, E. E. Long-term motor cortex plasticity induced by an electronic neural implant. *Nature* **444,** 56–60. ISSN: 14764687 (2006).
17. Garrity, A. G. *et al.* Aberrant "default mode" functional connectivity in schizophrenia. *American Journal of Psychiatry.* ISSN: 0002953X (2007).
18. Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M. & Scheltens, P. Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral Cortex* **17,** 92–99. ISSN: 10473211 (2007).
19. Meinshausen, N. & Bühlmann, P. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72,** 417–473 (2010).
20. Rebesco, J. M., Stevenson, I. H., Körding, K. P., Solla, S. A. & Miller, L. E. Rewiring neural interactions by micro-stimulation. *Frontiers in Systems Neuroscience* **4,** 1–15. ISSN: 16625137 (2010).
21. Keller, C. J. *et al.* Intrinsic functional architecture predicts electrically evoked responses in the human brain. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 17234. ISSN: 10916490 (2011).
22. Pernice, V., Staude, B., Cardanobile, S. & Rotter, S. How structure determines correlations in neuronal networks. *PLoS Computational Biology* **7.** ISSN: 1553734X (2011).
23. Rebesco, J. M. & Miller, L. E. Enhanced detection threshold for in vivo cortical stimulation produced by Hebbian conditioning. *Journal of Neural Engineering* **8.** ISSN: 17412560 (2011).
24. Zhao, L., Beverlin, B., Netoff, T. & Nykamp, D. Q. Synchronization from second order network connectivity statistics. *Frontiers in Computational Neuroscience* **5,** 1–16. ISSN: 16625188 (2011).
25. Edwardson, M. A., Lucas, T. H., Carey, J. R. & Fetz, E. E. New modalities of brain stimulation for stroke rehabilitation. *Experimental Brain Research* **224,** 335–358 (2013).

26. Lozano, A. M. & Lipsman, N. Probing and Regulating Dysfunctional Circuits Using Deep Brain Stimulation. *Neuron* **77,** 406–424. ISSN: 08966273 (2013).

27. Wang, Z. *et al.* The Relationship of Anatomical and Functional Connectivity to Resting-State Connectivity in Primate Somatosensory Cortex. *Neuron* **78,** 1116–1126. ISSN: 08966273 (2013).

28. Hu, Y., Trousdale, J., Josić, K. & Shea-Brown, E. Local paths to global coherence: Cutting networks down to size. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics.* ISSN: 15502376 (2014).

29. Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D. & Pizzagalli, D. A. Large-scale network dysfunction in major depressive disorder: A meta-analysis of resting-state functional connectivity. *JAMA Psychiatry* **72,** 603–611. ISSN: 2168622X (2015).

30. Ledochowitsch, P. *et al.* Strategies for optical control and simultaneous electrical readout of extended cortical circuits. *Journal of Neuroscience Methods* **256,** 220–231 (2015).

31. Fornito, A., Zalesky, A. & Bullmore, E. T. *Fundamentals of Brain Network Analysis* ISBN: 9780124079083 (2016).

32. Levy, R. M. *et al.* Epidural Electrical Stimulation for Stroke Rehabilitation. *Neurorehabilitation and Neural Repair* **30,** 107–119. ISSN: 15526844 (2016).

33. Yazdan-Shahmorad, A. *et al.* A Large-Scale Interface for Optogenetic Stimulation and Recording in Nonhuman Primates. *Neuron* **89,** 927–939. ISSN: 10974199 (2016).

34. Hariz, M. My 25 stimulating years with DBS in Parkinson's disease. *Journal of Parkinson's Disease* **7,** S33–S41. ISSN: 1877718X (2017).

35. Ocker, G. K. *et al.* From the statistics of connectivity to the statistics of spike times in neuronal networks. *Current Opinion in Neurobiology.* ISSN: 18736882 (2017).

36. Seeman, S. C., Mogen, B. J., Fetz, E. E. & Perlmutter, S. I. Paired Stimulation for Spike-Timing-Dependent Plasticity in Primate Sensorimotor Cortex. *The Journal of Neuroscience* **37,** 1935–1949. ISSN: 0270-6474 (2017).

37. Keller, C. J. *et al.* Induction and quantification of excitability changes in human cortical networks. *Journal of Neuroscience* **38,** 5384–5398. ISSN: 15292401 (2018).

38. Yazdan-Shahmorad, A., Silversmith, D. B., Kharazia, V. & Sabes, P. N. Targeted cortical reorganization using optogenetics in non-human primates. *Elife* **7,** e31034 (2018).

39. Bloch, J. A. *et al.* Cortical Stimulation Induces Network-Wide Coherence Change in Non-Human Primate Somatosensory Cortex. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS,* 6446–6449. ISSN: 1557170X (2019).

40. Haris, A., Shojaie, A. & Simon, N. Nonparametric regression with adaptive truncation via a convex hierarchical penalty. *Biometrika* **106,** 87–107 (2019).

41. Huang, Y. *et al.* Intracortical Dynamics Underlying Repetitive Stimulation Predicts Changes in Network Connectivity. *The Journal of neuroscience* **39,** 6122–6135. ISSN: 15292401 (2019).

42. Khateeb, K., Griggs, D., Sabes, P. N. & Yazdan-Shahmorad, A. Convection Enhanced Delivery of Optogenetic Adeno-associated Viral Vector to the Cortex of Rhesus Macaque Under Guidance of Online MRI Images. *Journal of visualized experiments : JoVE* **147** (2019).

43. Zarzycki, M. Z. & Domitrz, I. Stimulation-induced side effects after deep brain stimulation- A systematic review. *Acta Neuropsychiatrica* **32,** 57–64. ISSN: 16015215 (2020).

44. Yang, Y. *et al.* Modelling and prediction of the dynamic responses of large-scale brain networks during direct electrical stimulation. *Nature Biomedical Engineering* (2021).