nature biomedical engineering

ARTICLES

https://doi.org/10.1038/s41551-020-00666-w

Modelling and prediction of the dynamic responses of large-scale brain networks during direct electrical stimulation

Yuxiao Yang[1,7], Shaoyu Qiao[2,7], Omid G. Sani[1], J. Isaac Sedillo[2], Breonna Ferrentino[2], Bijan Pesaran[2,3,4] and Maryam M. Shanechi[1,5,6] ✉

Direct electrical stimulation can modulate the activity of brain networks for the treatment of several neurological and neuropsychiatric disorders and for restoring lost function. However, precise neuromodulation in an individual requires the accurate modelling and prediction of the effects of stimulation on the activity of their large-scale brain networks. Here, we report the development of dynamic input–output models that predict multiregional dynamics of brain networks in response to temporally varying patterns of ongoing microstimulation. In experiments with two awake rhesus macaques, we show that the activities of brain networks are modulated by changes in both stimulation amplitude and frequency, that they exhibit damping and oscillatory response dynamics, and that variabilities in prediction accuracy and in estimated response strength across brain regions can be explained by an at-rest functional connectivity measure computed without stimulation. Input–output models of brain dynamics may enable precise neuromodulation for the treatment of disease and facilitate the investigation of the functional organization of large-scale brain networks.

Direct electrical stimulation of the brain is a technique for modulating brain activity that can help treat a variety of brain dysfunctions and facilitate brain functions[1–3]. For example, deep brain stimulation (DBS) is effective in neurological disorders[4] such as Parkinson's disease[5] and epilepsy[6], and holds promise for neuropsychiatric disorders such as chronic pain[7], treatment-resistant depression[8] and obsessive–compulsive disorder[9]. Direct electrical stimulation also has the potential to modulate brain functions such as learning[10], and for use in investigating their neural substrates, for example, in speech production[11] and sensory processing[12].

Although the mechanism of action by which direct electrical stimulation alters brain activity is still unknown[4], studies have shown that stimulation alters the activity of multiple brain regions (both local and long range[4,13–17]) distributed across large-scale brain networks. This network-level stimulation effect has been observed with various signal modalities such as local field potential (LFP)[16], electrocorticogram (ECoG)[13,17], functional magnetic resonance imaging (fMRI)[15] and diffusion tensor imaging (DTI)[14]. These observations highlight the essential need for modelling the effect of stimulation on large-scale multiregional brain network activity, which has largely not been possible to date. Such modelling is especially important when the temporal pattern of stimulation needs to change in real time and when the activity of multiple brain regions needs to be monitored. For example, closed-loop DBS therapies for neurological and neuropsychiatric disorders[1–3,18–21] aim to change the stimulation pattern (for example, the frequency and amplitude of a stimulation pulse train) in real time on the basis of feedback of changes in brain activity. In addition, neural feedback may need to be provided from multiple brain regions[1–3,21–23], for example, in neuropsychiatric disorders that involve a large-scale multiregional brain network whose functional organization is not well understood[24–26]. Despite its importance across a wide range of applications, establishing the ability to predict how ongoing stimulation (input) drives the time evolution (that is, dynamics) of large-scale multiregional brain network activity (output) remains elusive[1,18].

Computational modelling studies to date have largely focused on building biophysical models of spiking neurons. Biophysical models can provide valuable insights into the mechanisms of action of stimulation—for example, in explaining population-level disease-specific observations especially for Parkinson's disease[27–31] and epilepsy[32,33]—and guide the design of open-loop stimulation patterns using numerical simulations[34,35]. However, biophysical models are typically for disease-specific brain regions, require some knowledge of their functional organization (for example, the cortical-basal-ganglia network in Parkinson's disease[27–29,31]) and involve a large number of nonlinear model parameters that can be challenging to fit to experimental data from an individual[33]. Thus, biophysical models are difficult to generalize to modelling how stimulation drives large-scale multiregional brain network dynamics in an individual, especially in neuropsychiatric disorders where the disease-relevant brain networks are not well characterized[24–26].

An alternative approach to biophysical models is data-driven modelling, as suggested by computer simulations[18,36,37]. However, previous data-driven studies of the brain[38–42] have not aimed at modelling the dynamic response of large-scale multiregional brain networks to ongoing stimulation. Some studies have built models of brain structural connectivity using diffusion-weighted imaging

[1]Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA. [2]Center for Neural Science, New York University, New York, NY, USA. [3]Neuroscience Institute, New York University Langone Health, New York, NY, USA. [4]Department of Neurology, New York University Langone Health, New York, NY, USA. [5]Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA. [6]Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA. [7]These authors contributed equally: Yuxiao Yang, Shaoyu Qiao. ✉e-mail: shanechi@usc.edu

(DWI) data and then correlated these structural models with the amount of static change in functional connectivity[41] or brain activity[42] observed after stimulation ends. Thus, modelling the brain activity or its dynamics during ongoing stimulation was not the goal of these studies. Other studies have taken important steps by fitting phenomenological nonlinear models of activity during stimulation using data, but have focused on columnar responses within a particular brain region[38] or on individual neurons[39] rather than on multiregional brain networks. Thus, there is a need to develop data-driven input–output (IO) models that resolve the two challenges of (1) predicting the large-scale dynamic neural response to ongoing stimulation and (2) doing so for large-scale multiregional brain networks.

Here, with the goal of developing an enabling technology towards precise modulation of brain functions and dysfunctions, we establish the ability to predict the dynamic response of large-scale multiregional brain networks to ongoing temporally varying stimulation in two awake rhesus macaque monkeys. We achieve this prediction by introducing and developing dynamic IO models for brain network activity using machine learning techniques. We use a customized semi-chronic microdrive system to deliver continuous temporally varying microstimulation while simultaneously recording across a large-scale multiregional brain network. We measure the brain network response using LFP power features. We take the model input as the stimulation amplitude and frequency, which are changed in real time. To obtain data for accurate machine learning, we design and implement a stimulation waveform[18] whose amplitude and frequency change stochastically in time to effectively excite the brain network activity and apply it in vivo in the primate brain. We find that our models can predict the dynamics of brain network response. The dynamic structure of the IO models and modelling the changes in both stimulation amplitude and frequency are also essential for prediction. Further, the brain network response exhibits complex damping and oscillatory dynamics. Finally, the variability in both model prediction accuracy and estimated response strength across different brain regions within the network can be explained by a control-theoretic at-rest functional connectivity measure that is computed using our models fitted to at-rest activity. By predicting the dynamic real-time effect of stimulation, our dynamic IO modelling technology can help facilitate the design of accurate closed-loop neuromodulation systems for treatment of a wide range of neurological and neuropsychiatric disorders, for modulation of brain functions and for probing the brain network organization.

Results

Modelling framework, neural recordings and stochastic stimulation input. We developed a data-driven dynamic modelling and machine learning approach to model the dynamic response of brain network activity—termed brain network dynamics in short—to microstimulation pulse trains with temporally varying amplitude and frequency in two macaque monkeys (monkey A and monkey M) (Fig. 1). As output, we computed the LFP power feature time series from multiple brain regions after rejecting stimulation artefacts (Methods, Supplementary Figs. 1 and 2, and Supplementary Note 1) and at four frequency bands—1–8 Hz (delta + theta), 8–12 Hz (alpha), 12–30 Hz (beta), 30–50 Hz (low gamma)—because of their relevance to brain functions[1,13,22,43] and dysfunctions[1,2,44,45] (Methods and Supplementary Note 2). We refer to each LFP power feature as a network node. As input, we took the amplitude and frequency of the stimulation pulse train, given that they are key factors that influence the stimulation effect[1,27–29]. In each experiment, we performed continuous bipolar microstimulation at a given site, chosen from orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), amygdala (AMG) or superior parietal lobule (SPL), while simultaneously recording LFP activity across multiple brain regions spanning the prefrontal cortices, the motor cortices, the parietal cortices, as well as the striatum, pallidum and AMG (Methods and Supplementary Tables 1 and 2). During the experiments, monkeys were awake and did not actively participate in behavioural tasks.

Accurate IO modelling and machine learning requires an appropriate IO model structure and informative IO datasets to fit the models[18]. We hypothesized that capturing the effect of ongoing stimulation would require modelling how stimulation drives brain network dynamics. We thus built a dynamic linear state-space model (LSSM) structure (Methods and Supplementary Fig. 3). This model describes the effect of stimulation in terms of a latent state whose time variations drive the brain network dynamics. Obtaining informative IO datasets requires delivering an input stimulation waveform that can sufficiently excite brain network activity. Intuitively, this can be done by designing a waveform that is white spectrum in the input space (in amplitude and frequency). On the basis of previous theoretical work[18], we designed and delivered a multilevel noise (MN)-modulated stimulation pulse train to conduct IO modelling. The MN-modulated pulse train stochastically switched the amplitude and frequency in time between multiple discrete levels and thus was white in input space (Methods, Fig. 1a, Supplementary Fig. 4 and Supplementary Notes 3 and 4). Using the IO dataset, we fitted and evaluated the IO models using machine learning techniques and with cross-validation.

Experimental design and IO model evaluation. Overall brain network dynamics (that is, the measured output LFP power feature time series) consist of input-driven dynamics that are due only to the stimulation input as well as intrinsic dynamics that do not depend on the stimulation input (that is, are input-irrelevant) (Methods and Supplementary Note 5). A strong test of an IO model assesses its ability to forward predict the input-driven dynamics in response to a temporally varying stimulation input. Forward prediction is a challenging test because it quantifies how well the IO model can predict the input-driven dynamics using only the past stimulation inputs and without any knowledge of the past measured brain network activity; thus, this test predicts the current value of the LFP power features using only the history of input stimulation, without reference to past measured LFP power features and with zero initialization (Methods and Supplementary Note 6).

To evaluate the accuracy of forward prediction, we need to dissociate the input-driven part of the overall brain network dynamics being measured, which is difficult. To do so, we designed a multi-trial experiment that repeats the same stochastic MN-modulated pulse train on each trial (Fig. 1b, Supplementary Fig. 5 and Supplementary Notes 6 and 7). Given the same input in each trial, single-trial input-driven dynamics are the same in all trials, whereas intrinsic dynamics—which are input-irrelevant—change from trial to trial. Thus, averaging the measured overall brain network dynamics across trials can dissociate the ground-truth input-driven dynamics by reducing intrinsic dynamics while keeping input-driven dynamics the same as single-trial input-driven dynamics. We emphasize that our models are fitted with single-trial data without any averaging and averaging is only done to dissociate the ground truth of single-trial input-driven dynamics from measured brain activity and for assessing the fitted models (Methods).

We evaluated the IO models for forward prediction using a fourfold cross-validation (Methods, Fig. 1c, Supplementary Fig. 5 and Supplementary Notes 6 and 8). In each cross-validation fold, in every trial, we left out the same quarter (for example, the first quarter) of IO data as the test set, obtaining one test set for each trial (Supplementary Fig. 5b,d). We then took out the rest of the IO data from each trial (for example, the last three quarters) and concatenated these single-trial data across trials to construct the training set (Supplementary Fig. 5b,c). We first fitted the IO model using the single-trial data in the training set without any averaging

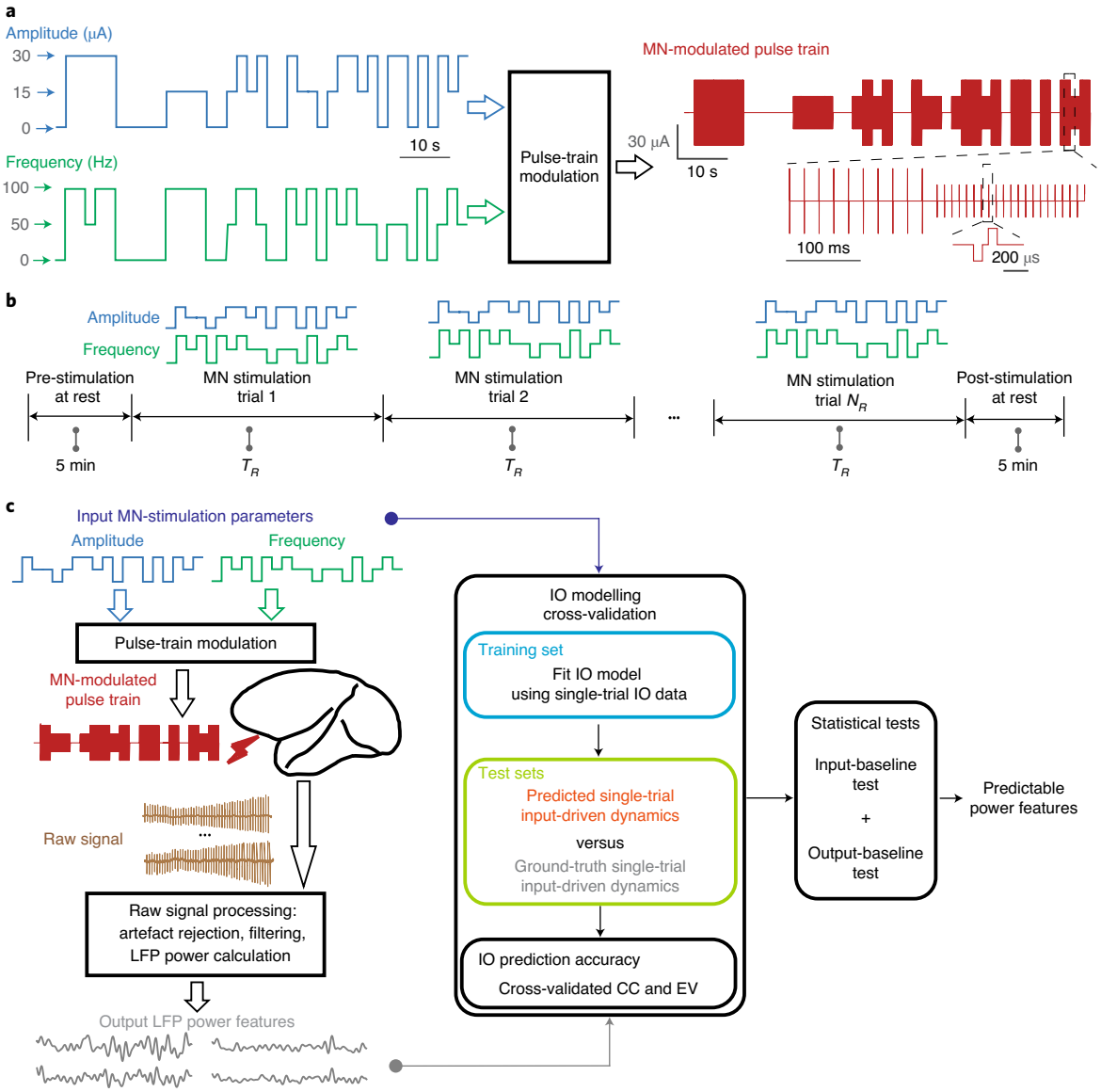

Fig. 1 | Input design, stimulation experiments and IO modelling framework. a, An example MN-modulated microstimulation pulse train was delivered to the cathode lead in bipolar stimulation. The stimulation amplitude and frequency time series were independently generated by stochastically changing between multiple levels (left). These time series were then used to modulate the amplitude and frequency of a biphasic charge-balanced pulse train (right, expanded section of the MN-modulated pulse train). **b**, Stimulation was performed in a multi-trial experiment. In one experiment, the same stochastic MN-modulated pulse train was delivered repeatedly in multiple trials (N_R trials, each with duration T_R) and LFP signals were simultaneously recorded during stimulation and pre- and post-stimulation. **c**, The MN-modulated microstimulation pulse train was delivered to the brain through implanted electrodes (red lightning symbol) while raw neural signals (brown) were simultaneously recorded. In IO modelling, we took the input as stimulation amplitude and frequency and the output as the recorded LFP power features after artefact rejection (Supplementary Figs. 1 and 2) (left). Cross-validation was used to fit and evaluate the IO models (Supplementary Fig. 5) (middle). Two statistical tests (input-baseline test and output-baseline test) were used to evaluate the significance of the forward prediction of single-trial input-driven dynamics in cross-validation (Supplementary Fig. 6) (right).

(Supplementary Fig. 5c). Then in the test sets, we performed the following (Methods and Supplementary Fig. 5d). (1) We used the fitted IO model and knowledge of the stimulation input in each test set to forward predict the LFP power features; this provided the test set's predicted single-trial input-driven dynamics, which was the same in all trials (equivalently in all test sets of a given cross-validation fold) as the input in every trial was identical by design. (2) We compared the prediction and ground truth of single-trial input-driven dynamics by computing the linear correlation coefficient (CC) between them. As described above, the ground truth was found by averaging the measured LFP power features across test sets (equivalently across trials; Supplementary Fig. 5d). We quantified the IO prediction accuracy of each LFP power feature by its CC in forward prediction. We emphasize that the stochastic inputs tested in different cross-validation folds within an experiment or in different experiments were independent (Supplementary Fig. 5b), allowing us to test the generalizability of the IO model. We also computed the explained variance (EV) of our dynamic IO models in forward prediction (Methods). Finally, since the MN pattern is stochastic, for each cross-validation fold, the test data had an input pattern independent of the input pattern in the train data, thus making test and training IO data independent and removing the confound of overfitting to input patterns (Supplementary Fig. 5b and Supplementary Note 9).

To evaluate the statistical significance of IO model predictions and control for the effects of stimulation-artefact rejection, we applied the same modelling to artificially generated IO datasets that kept the output the same but randomly generated the input (input-baseline test; Methods, Fig. 1c and Supplementary Fig. 6a) as well as to artificially generated IO datasets that kept the input the same but replaced the output with pre-stimulation at-rest LFP activity (output-baseline test; Methods, Fig. 1c and Supplementary Fig. 6b). In the output-baseline test, the same stimulation-artefact rejection algorithm was applied to at-rest LFP activity to control for the effect of the stimulation-artefact rejection algorithm on the prediction performance (Methods, Supplementary Note 1 and Discussion). We define the input-baseline P value as the probability that IO prediction accuracy from the input-baseline is larger than that from the actual IO datasets, and similarly define the output-baseline P value (Methods and Supplementary Note 10). We define the LFP power features whose false-discovery rate (FDR)-corrected (across all modelled LFP power features in this IO dataset) $P<0.05$ in both the input-baseline and output-baseline tests as predictable power features (Methods).

Beyond forward prediction as the main measure, we also used the dynamic IO models to predict single-trial overall brain network dynamics during stimulation. To do this, we need to use both the past inputs and the past measured LFP power features to predict the current LFP power features; this approach is termed one-step-ahead prediction (Methods and Supplementary Note 6). We applied this approach to IO datasets using the same cross-validation procedure described above as well as to at-rest LFP datasets without stimulation using a fourfold cross-validation (Supplementary Note 11). Since one-step-ahead prediction assesses overall brain network dynamics, the ground truth is simply the measured LFP power feature in each trial and no averaging is needed to obtain the ground truth in this case.

We conducted 16 MN-stimulation experiments across two monkeys (see Supplementary Table 3 for details of each experiment). In each experiment, we generated an MN-modulated pulse train lasting for a period ranging from 60 s to 270 s. We delivered the generated pulse train for multiple trials (ranging from 10 to 30 trials; Methods and Fig. 1b). We also recorded 5 min of pre-stimulation and 5 min of post-stimulation at-rest LFP signals from up to 208 (monkey A) and 165 (monkey M) channels, respectively. The total duration of each experiment ranged from 10 min to 120 min.

Dynamic IO models accurately predict brain network dynamics in response to stimulation. Across 16 IO datasets, we found that the dynamic IO models accurately predicted the single-trial input-driven dynamics of the brain network in response to stimulation (Figs. 2 and 3 and Supplementary Fig. 7). Figure 2a shows an example LFP power feature recorded from superior frontal gyrus (SFG) in response to OFC stimulation. The LFP power feature's predicted single-trial input-driven dynamics closely followed their ground truth in cross-validation, resulting in a high IO prediction accuracy of 0.65 (Fig. 2a). This prediction was significant both in the input-baseline test and output-baseline test (FDR-corrected $P<10^{-30}$ for both tests; Fig. 2a). Across 16 datasets, the predictable power features (Fig. 2b) had an IO prediction accuracy of 0.46 ± 0.006 (mean ± s.e.m.) in monkey A (Fig. 2c) and a similar IO prediction accuracy of 0.41 ± 0.021 in monkey M (Fig. 2d). Overall IO prediction accuracy across both monkeys was 0.45 ± 0.006 (or equivalently an EV of $21.98\%\pm0.01\%$). Our evaluation of the IO model prediction was robust to the choice of the performance measure as there was a significant positive correlation between the CC and EV (Spearman's rank correlation coefficient $\rho=0.34$, Spearman's $P=8.61\times10^{-8}$).

The predictable channels (defined as LFP channels that had at least one predictable power feature) were distributed across multiple brain regions (Fig. 3). Among all brain regions recorded from both monkeys, 72.09% of them had channels that showed predictable responses, suggesting a large-scale multiregional brain network response to stimulation (Supplementary Table 2). We also found that the IO model predictions were specific to the recorded channel site (FDR-corrected feature permutation-baseline test $P<0.05$; details in Supplementary Fig. 8). In addition, since the same stimulation-artefact rejection algorithm was applied in the output-baseline test, the significance in the output-baseline test also ruled out that prediction was due to the stimulation-artefact rejection algorithm (Discussion).

We also found that brain network dynamics responded to real-time changes in both stimulation amplitude and frequency. We separately modelled the effect of only one of these stimulation parameters (either amplitude or frequency) by repeating the model fitting and cross-validation procedure while assuming that the other stimulation parameter was always zero (Supplementary Note 12). Compared with modelling both stimulation parameters, the IO prediction accuracy when modelling only the stimulation amplitude was significantly smaller among predictable power features (0.33 ± 0.01 versus 0.45 ± 0.006, two-sided Wilcoxon signed-rank test, $P=1.52\times10^{-15}$; Supplementary Fig. 9a,b) and similar results held when modelling only the stimulation frequency (0.28 ± 0.01 versus 0.45 ± 0.006, two-sided Wilcoxon signed-rank test, $P=8.41\times10^{-28}$, Supplementary Fig. 9a,c). These results show that including both stimulation amplitude and frequency as input improved the prediction.

In addition, we found that predictable power features were distributed across all four frequency bands—that is, 1–8 Hz, 8–12 Hz, 12–30 Hz and 30–50 Hz. Further, after controlling for the larger effect of stimulation artefacts on the high-gamma band of 70–100 Hz (Supplementary Fig. 2 and Supplementary Note 1), the IO model also predicted the single-trial input-driven dynamics in this band with an IO prediction accuracy of 0.49 ± 0.03, which was not significantly different from that of the other four bands (Kruskal–Wallis test, $P=0.06$; Supplementary Fig. 10). IO modelling was robust to the choice of method used for computing the LFP power features at these frequency bands and different methods did not change the IO prediction accuracy (Supplementary Fig. 11 and Supplementary Note 2).

Beyond predicting single-trial input-driven dynamics, the fitted IO models also predicted the single-trial overall brain network dynamics. The cross-validated CC between the one-step-ahead prediction and the ground truth of the single-trial overall dynamics was 0.76 ± 0.02 (mean ± s.e.m.) for predictable power features (output-permutation test, $P<10^{-16}$; Methods and Supplementary Fig. 12). As expected, one-step-ahead prediction had higher accuracy than forward prediction (0.76 ± 0.02 versus 0.45 ± 0.006, two-sided Wilcoxon signed-rank test, $P<10^{-16}$) because the former used both the past inputs and the past measured LFP power features for prediction, whereas the latter could only use the past inputs (Methods and Supplementary Note 6). This result demonstrates how forward prediction is a more challenging assessment of the IO model because it cannot use the past measured LFP power features—which change from trial to trial—and needs to predict the input-driven dynamics purely based on the past inputs.

Finally, our IO models were tested in experiments performed on different days that spanned seven months and did not involve instructing the monkeys to perform a task to earn rewards (Methods). The fact that the IO models predicted the brain network response in both monkeys and in all experimental sessions suggests that the model prediction is robust to changes in brain or animal state. Together, these results show that the IO modelling framework can robustly predict the brain network dynamics in response to ongoing temporally varying stimulation.

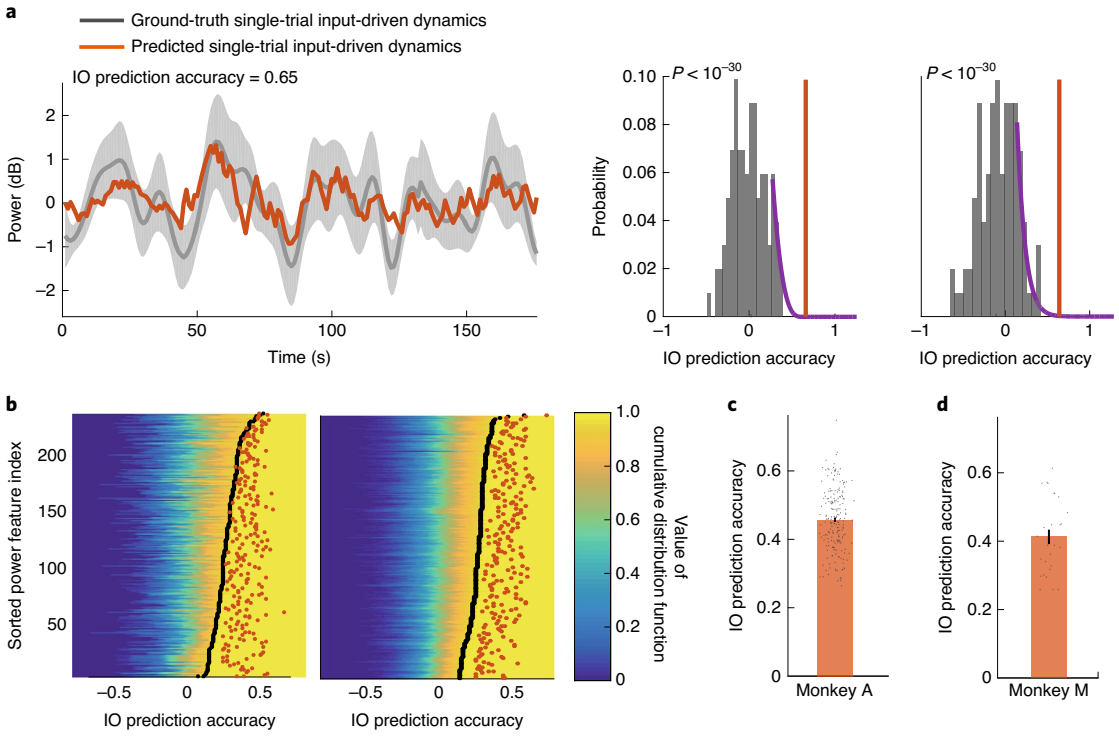

Fig. 2 | Dynamic IO models accurately predict brain network dynamics in response to stimulation. a, An example forward-prediction trace recorded from SFG in response to OFC stimulation from monkey A is shown. Forward prediction provides the predicted single-trial input-driven dynamics, which closely followed its ground truth in cross-validation (left), resulting in a high IO prediction accuracy. The grey shaded area provides the s.e.m. in the ground truth, which was computed from the measured LFP power features across trials. Histograms (dark grey) of the IO prediction accuracy in input-baseline (middle) and output-baseline (right) tests. These panels also show the fitted generalized Pareto distribution (GPD) to the tails of input-baseline and output-baseline distributions (purple, Supplementary Note 10) and the actual IO prediction accuracy (red vertical line). The prediction was significant according to both the input-baseline and output-baseline tests, with *P* values indicated on the top left. **b**, Input-baseline (left) and output-baseline (right) tests for predictable power features are visualized. Each row corresponds to one predictable power feature; for each of these power features, the cumulative distribution function of the input-baseline and output-baseline is shown by the colour map. The IO prediction accuracy resulting in a value of 0.95 in the cumulative distribution function (corresponding to the 0.05 significance level) is shown as a black dot for each input-baseline and output-baseline corresponding to one power feature (one row). The actual IO prediction accuracy value for each power feature is shown by the red dot, which was larger than the value associated with the black dot (that is, larger than the value associated with the significance level). The LFP power features (that is, rows) were sorted by the black dots for their cumulative distribution functions. **c**, Distribution of the significant IO prediction accuracy is shown for monkey A. The bar represents the mean and the black error bar represents s.e.m. Raw IO prediction accuracies are shown with dots ($N=206$ independent samples—that is, LFP power features). **d**, Same as **c**, but for monkey M. Note monkey M had fewer stimulation sessions, resulting in fewer dots (Supplementary Table 3, $N=27$ independent samples—that is, LFP power features).

The dynamic structure of the IO model is essential for accurate prediction. We found that the dynamic structure of the IO model was essential for accurate prediction of the brain network response and that this response exhibited complex damping and oscillatory dynamics. To investigate the dynamical characteristics, we examined in greater detail predictable power features that exhibited a response that could be studied and fitted special cases of the dynamic IO model that lacked various dynamical characteristics to these responses for comparison as follows.

First, we showed that the brain response depended on the history of stimulation, and thus was dynamic. To do so, we compared the dynamic IO model with a static linear regression model (Fig. 4a,b). The static regression modelled the output LFP power features at each time as a linear function of the input stimulation parameters at that time without history dependence (Methods). We repeated the same cross-validation procedure for the regression model. Among the predictable power features, the IO prediction accuracy from the regression model was significantly smaller than that from the dynamic IO model (0.22 ± 0.02 versus 0.45 ± 0.006, two-sided Wilcoxon signed-rank test, $P=2.08\times10^{-27}$; Fig. 4a,b). This result suggests that the LFP power features responded dynamically to the real-time changes of stimulation amplitude and frequency and depended on their history.

We also found that the dynamic IO model did not simply smooth the response (Fig. 4c,d,f and Supplementary Fig. 13). In our dynamic IO model, the dynamics of the brain network response are characterized by the eigenvalues of the state transition matrix, which describes how the latent state evolves in time as a weighted function of the past inputs (Fig. 4c; Methods, Supplementary Fig. 13 and Supplementary Note 13). Within the same cross-validation procedure, we built a special case of the dynamic IO model with eigenvalues fixed at 1, which characterized the effect of input stimulation at each time as simply a smoothed average of past input values with equal weights (Methods and Supplementary Note 13). The IO prediction accuracy of this smoothing model was significantly smaller than that of the dynamic IO model (0.20 ± 0.02 versus 0.45 ± 0.006, two-sided Wilcoxon signed-rank test, $P=2.42\times10^{-32}$; Fig. 4d,f and Supplementary Fig. 13).

Moreover, we found that the brain network response to stimulation exhibited oscillatory dynamics, which correspond to the weights of past inputs exhibiting oscillations into the past (Fig. 4c,e,f and Supplementary Fig. 13). Complex conjugate pairs of eigenvalues

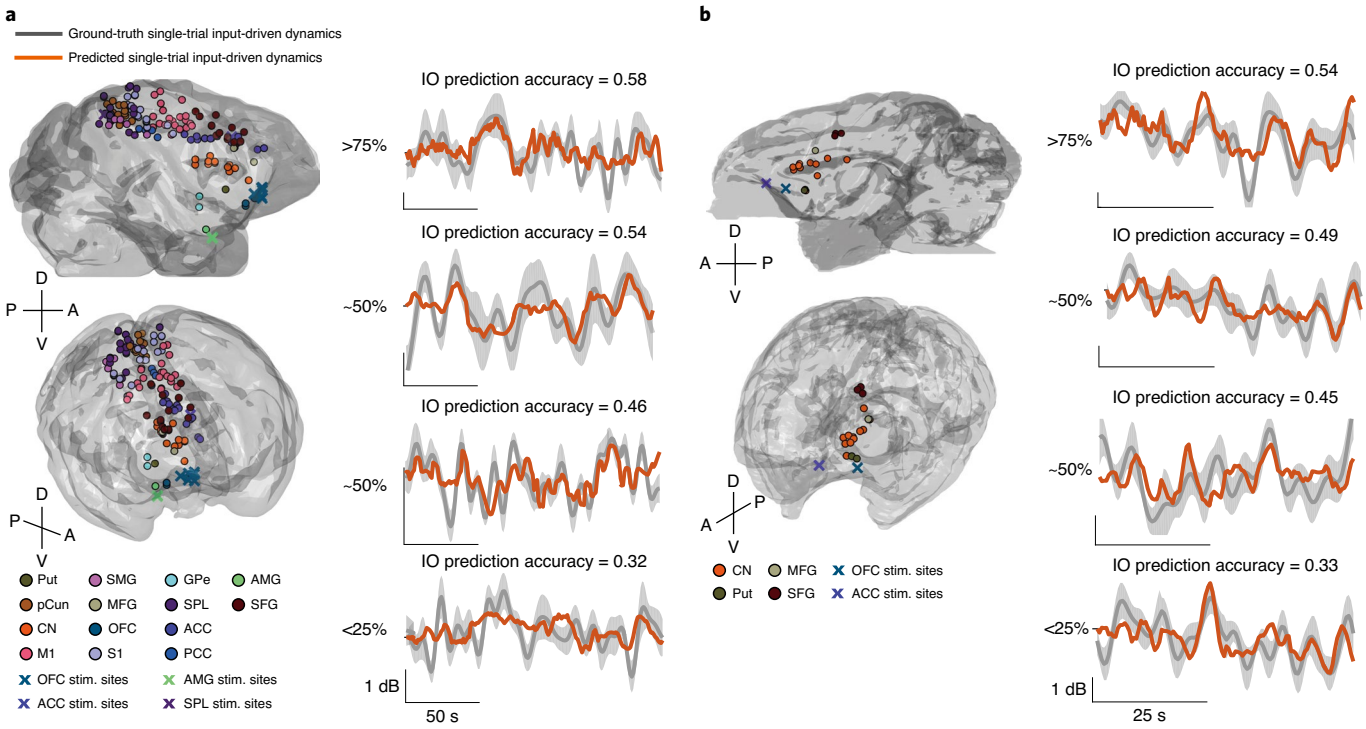

Fig. 3 | Dynamic IO models predict the response to stimulation across multiple brain regions. a, All predictable channels for monkey A are shown on the 3D-reconstructed monkey brains from magnetic resonance imaging (MRI) data (data are pooled across all IO datasets for monkey A). The anatomical region of the predictable channels and stimulation (stim.) sites are colour coded. Anterior (A), posterior (P), dorsal (D) and ventral (V) directions are indicated. Brain regions that had at least two predictable channels are shown. Example forward-prediction traces of power features across multiple brain regions are shown to the right of the 3D brain plots with similar convention as in Fig. 2a. The grey solid trace represents the ground-truth single-trial input-driven dynamics and the grey shaded area provides the s.e.m. in the ground truth. The orange solid trace represents the predicted single-trial input-driven dynamics. The IO prediction accuracies in the four rows are within >75% (top), ~50% (middle two) and <25% (bottom) quantiles of the distribution of IO prediction accuracy. More example forward-prediction traces are shown in Supplementary Fig. 7. **b**, Same as **a** but for monkey M. CN, caudate nucleus; GPe, external globus pallidus; M1, precentral gyrus; MFG, middle frontal gyrus; PCC, posterior cingulate cortex; pCun, precuneus; Put, putamen; S1, postcentral gyrus; SMG, supramarginal gyrus. Brain regions are further described and illustrated in Supplementary Tables 1 and 2.

and negative real eigenvalues represent oscillatory dynamics (Fig. 4c, Methods, Supplementary Fig. 13 and Supplementary Note 13). First, we found that more than half (53%) of the eigenvalues across all fitted dynamic IO models represented oscillatory dynamics (Supplementary Fig. 13). Second, we repeated the same cross-validated dynamic IO modelling, but this time constrained the state transition matrix to only have positive real eigenvalues. Thus, this analysis made the fitted dynamic IO model unable to model any oscillatory dynamics (Methods and Supplementary Note 13). We term this model the non-oscillatory model. We found that the IO prediction accuracy of this non-oscillatory model was again significantly smaller than that of the dynamic IO models (0.29 ± 0.01 versus 0.45 ± 0.006, two-sided Wilcoxon signed-rank test, $P = 4.48 \times 10^{-29}$; Fig. 4e,f and Supplementary Fig. 13), confirming the existence of oscillatory IO dynamics in the brain network response. Also, as expected, the IO prediction accuracy using the non-oscillatory model was significantly larger than the smoothing model (0.29 ± 0.01 versus 0.20 ± 0.02, two-sided Wilcoxon signed-rank test, $P = 1.77 \times 10^{-6}$; Fig. 4f), because the smoothing model is a special case of the non-oscillatory model in which the past input weights do not need to be equal and can decay (Supplementary Fig. 13).

As a control, we found that the conclusions of all the above three comparisons held even when considering all LFP power features (whether predictable or not) (two-sided Wilcoxon signed-rank test, $P < 10^{-9}$ for all comparisons). Moreover, beyond forward prediction of single-trial input-driven dynamics, these conclusions also held for one-step-ahead prediction of single-trial overall brain network dynamics, with the dynamic IO model still outperforming the other three models for one-step-ahead prediction (two-sided Wilcoxon signed-rank test, $P < 10^{-16}$ for all comparisons). Finally, we performed a time-scale analysis of the dynamic effect of stimulation (Supplementary Notes 14 and 15). We found that mainly the slower time-scale content of the temporally varying stimulation amplitude and frequency drove the brain network dynamics (Supplementary Figs. 14–17).

At-rest functional controllability explains the variability in the IO prediction accuracy across different network nodes. The IO prediction accuracy varied across different network nodes (Figs. 2b–d and 3). We thus investigated whether this variability could be explained using at-rest data even without stimulation. In particular, we hypothesized that the IO prediction accuracy at different network nodes should depend on their at-rest functional connectivity to the stimulation node—defined as the four LFP power features at the stimulation site—and thus this variability should be explainable using at-rest brain network activity.

To test this hypothesis, we calculated a control-theoretic connectivity measure, termed controllability[46], from the stimulation node to each network node using at-rest LFP power features (Methods, Fig. 5a and Supplementary Notes 16 and 17). Controllability measures have shown promise in studying structural connectivity in the brain based on tractography derived from DTI/DWI data[41,47–50] (Discussion). Our goal was instead to model electrophysiological

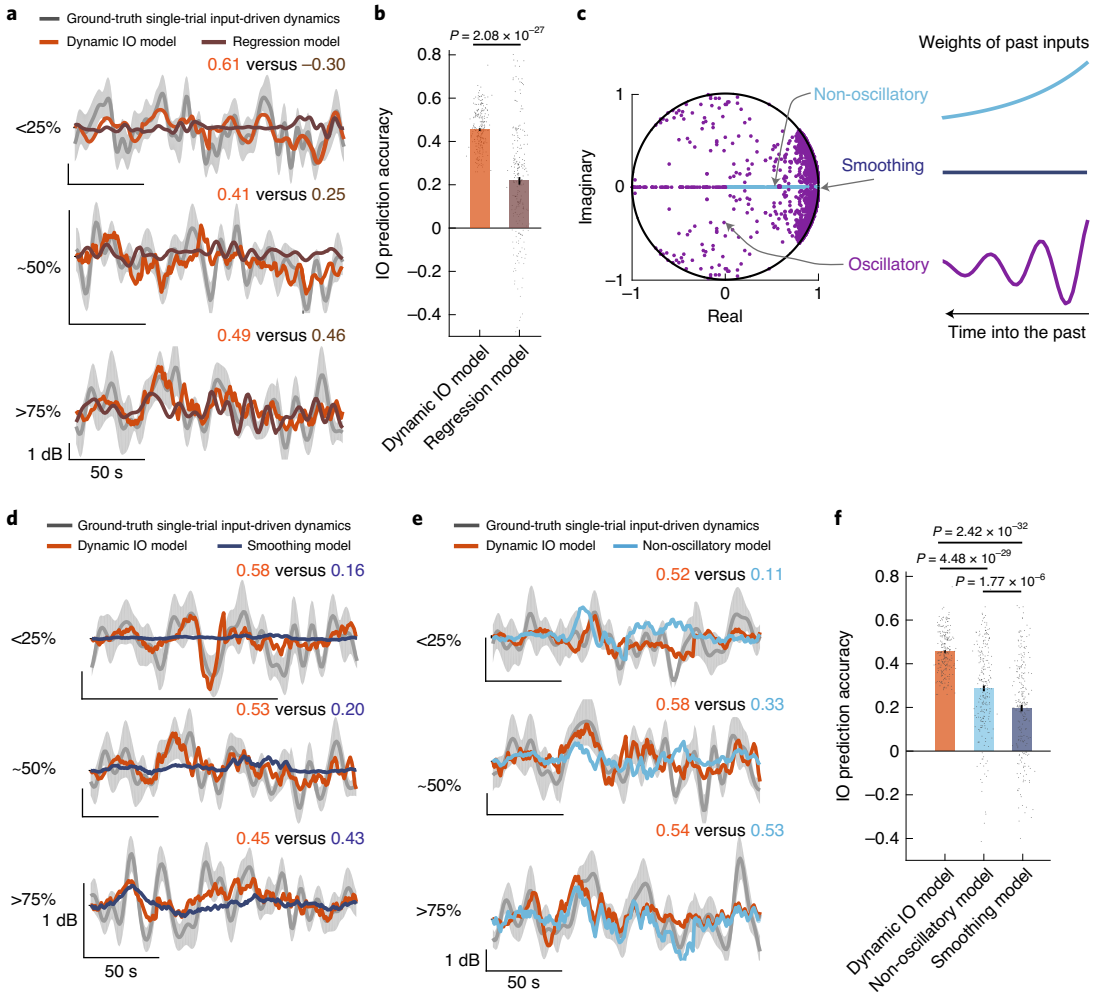

Fig. 4 | The dynamic structure of the IO model is essential for accurate prediction. a, Forward-prediction traces of three example LFP power features using the dynamic IO model and the regression model are shown. The regression model's IO prediction accuracies of the three examples are within <25% (top), ~50% (middle) and >75% (bottom) quantiles of the distribution of the regression model's IO prediction accuracies. The >75% example (bottom) is chosen as an example where the IO prediction accuracy of the regression model is similar to that of the dynamic IO model. The grey shaded area provides the s.e.m. in the ground truth. The IO prediction accuracy comparison is indicated on the top right of each example. **b**, The IO prediction accuracy of the dynamic IO model was significantly larger than the regression model. The bar represents the mean and the black error bar represents the s.e.m. Raw IO prediction accuracies are shown with dots ($N = 233$ independent samples—that is, LFP power features—for both bars). Two-sided Wilcoxon signed-rank test P value is shown. **c**, The eigenvalues of the state transition matrix across all fitted LSSMs from the 16 IO datasets are shown in the complex domain (left) (details in Supplementary Fig. 13 and Supplementary Note 13). Eigenvalues on the positive real axis are coloured light blue and those elsewhere are coloured purple. All eigenvalues were within the unit circle (black circle), representing stable dynamics of the brain network response to stimulation. Example weights of past inputs associated with different types of eigenvalue are shown (right). Eigenvalue at 1 represents smoothing dynamics without any damping (right middle). Eigenvalue on the positive real axis but smaller than 1 represents non-oscillatory dynamics with exponential damping (right top). Complex conjugate eigenvalues represent oscillatory dynamics with exponential damping (right bottom). **d**, Forward-prediction traces of three example LFP power features using the dynamic IO model and the smoothing model being compared to. Figure convention is the same as **a**, with examples chosen to show a range of IO prediction accuracy of the comparison model within <25%, ~50% and >75% quantiles of its IO prediction accuracy distribution. **e**, Same as **d**, with the comparison model being the non-oscillatory model. **f**, Bar plot summarizing the statistics of the IO prediction accuracy of the dynamic IO model, the smoothing model and the non-oscillatory model. The bar represents the mean and the black error bar represents the s.e.m. Raw IO prediction accuracies are shown with dots ($N = 233$ independent samples—that is, LFP power features—for all bars). Two-sided Wilcoxon signed-rank test P values are shown.

brain network activity that measures network function[51]. We thus hypothesized that applying a controllability measure to LFP activity should provide a measure of functional network connectivity that quantifies how easy it is to control the activity of each network node from the stimulation node. We fitted a LSSM to at-rest network LFP power features by taking these features directly as the state variable (Methods). We then used this LSSM to calculate the functional controllability of each network node from the stimulation node (Methods and Supplementary Note 17). For each network node, we examined the relationship between its at-rest functional controllability and its IO prediction accuracy computed during stimulation by our dynamic IO model (Fig. 5a).

As a stimulation site, we found that OFC had significantly higher at-rest functional controllability to the network nodes compared with the other stimulation sites tested. The percentage of network nodes with significant at-rest functional controllability from OFC

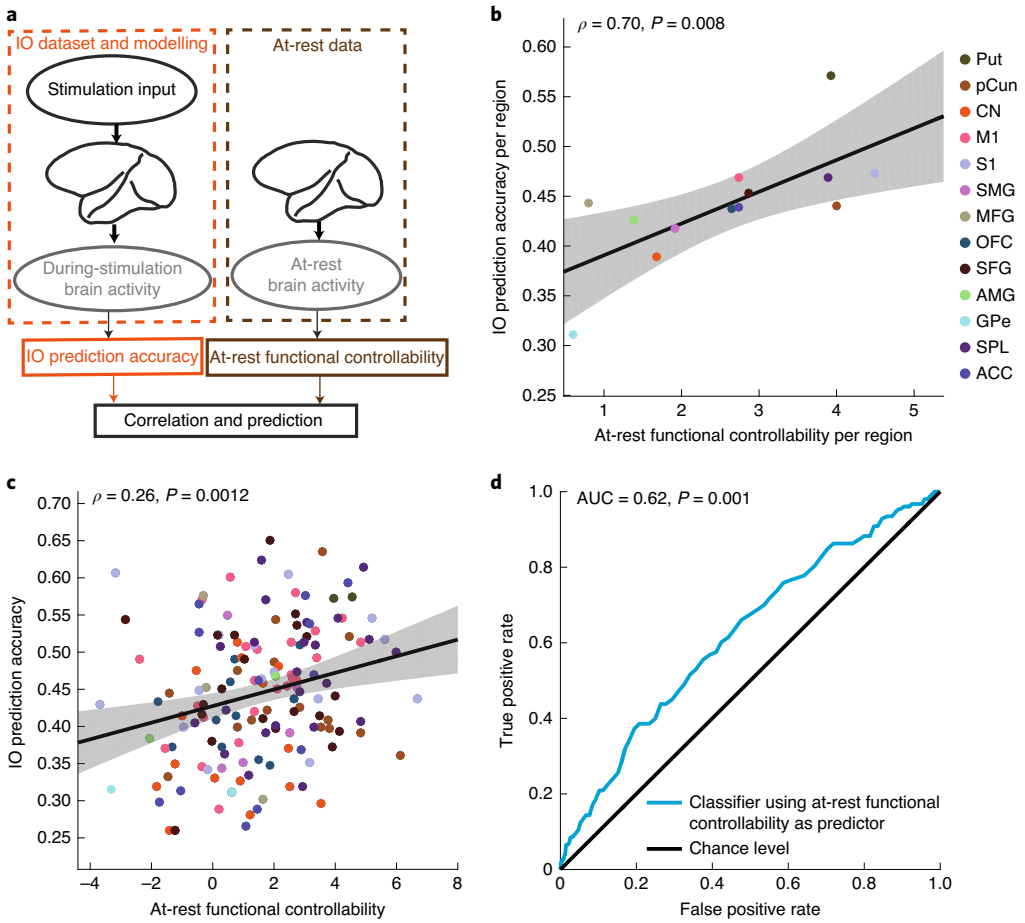

Fig. 5 | At-rest functional controllability explains the variability in the IO prediction accuracy at different network nodes. a, The IO prediction accuracy was calculated using IO data during OFC stimulation (left). At-rest functional controllability was calculated from at-rest LFP data without stimulation and provided our measure of functional connectivity from the stimulation node to each network node (right). **b**, At-rest functional controllability of each brain region significantly correlated with the IO prediction accuracy at that region ($N=13$ independent samples—that is, brain regions). The least-squares fitted linear line is shown as solid line and its 95% confidence bound is shown as the shaded area. The CC ρ and Pearson's P value are shown on the top left. **c**, Among the predictable power features, at-rest functional controllability is significantly correlated with the IO prediction accuracy. Dots present raw power feature data ($N=153$ independent samples—that is, predictable power features) and are colour coded according to the brain region from which they are recorded, similar to **b**. Other figure conventions are the same as in **b**. **d**, Among all network nodes (all power features), ROC of using at-rest functional controllability to classify whether or not each network node (each power feature) would exhibit predictable responses to stimulation is shown. ROC was significantly above the 45° line, which represents chance level. The permutation test P value is shown on the top left (Methods).

was much higher compared with the other stimulation sites (OFC: 17.32% versus ACC: 4.78%, AMG: 1.00%, SPL: 0.40%; Methods). Moreover, the z-scored at-rest functional controllability from OFC (4.92 ± 0.49, mean ± s.e.m.) was much larger than chance level (chance-level z-score was 0 ± 1, see Methods) and significantly higher than the other three stimulation sites (OFC: 4.92 ± 0.49 versus ACC: 2.14 ± 0.49, AMG: 1.16 ± 0.25, SPL: 1.79 ± 0.28, pairwise Kruskal–Wallis test, FDR-corrected $P<10^{-15}$ for all paired comparisons with OFC). Thus, to obtain reliable at-rest functional controllability values for testing our hypothesis, we focused our subsequent analyses on the ten datasets with OFC stimulation (Supplementary Table 3).

In our OFC-stimulation datasets, we found that at-rest functional controllability explained the variability in the IO prediction accuracy across brain regions, which was computed by our dynamic IO model. First, the average at-rest functional controllability at each brain region had a strong positive correlation with the average IO prediction accuracy at that region (Fig. 5b; CC $\rho=0.70$, Pearson's $P=0.008$). This linear correlation accounted for 49% of the variability in IO prediction accuracy. Second, a linear regression model significantly predicted the value of average IO prediction accuracy from average at-rest functional controllability within cross-validation (Methods; permutation test $P=0.004$).

Even at the level of single network nodes (that is, LFP power features), at-rest functional controllability explained the variability in the IO prediction accuracy as computed by the dynamic IO model. First, among predictable power features, at-rest functional controllability significantly correlated with the IO prediction accuracy (Fig. 5c; CC $\rho=0.26$, Pearson's $P=0.0012$). Second, a linear regression model significantly predicted the value of IO prediction accuracy for each network node from its at-rest functional controllability in cross-validation (Methods; permutation test $P=0.001$).

In these two analyses, we focused on predictable power features because the IO prediction accuracy of other (unpredictable) power features did not pass the chance level (input-baseline and output-baseline tests) and would just inject noise into these correlation analyses. Nevertheless, even when including all modelled LFP power features regardless of their predictability in the analysis and despite the injected noise, at-rest functional controllability still significantly correlated with the IO prediction accuracy (CC $\rho=0.12$, Pearson's $P=2.8\times10^{-9}$; Supplementary Fig. 18). Further, when

considering all LFP power features, at-rest functional controllability of the predictable power features was significantly larger than that of the unpredictable power features (1.52 ± 0.17 versus 0.61 ± 0.04, Kruskal–Wallis test, $P = 3.72 \times 10^{-7}$). Consistently, among all LFP power features, we found that at-rest functional controllability of a power feature can determine whether or not that power feature would have predictable responses to stimulation. To make this determination, we compared the at-rest functional controllability to a threshold (Methods). We found that the receiver operating characteristic (ROC) area under the curve (AUC) for this threshold-based test was 0.62 and significantly greater than chance (permutation test $P = 0.001$; Fig. 5d and Methods).

Finally, the above results also serve as two additional control analyses (Discussion). First, these results provide an independent validation of the dynamic IO modelling framework. This is because these results show the consistency of the IO prediction accuracy calculated by the dynamic IO model from data during stimulation with at-rest functional controllability calculated from data without applying any stimulation. Second, this consistency with at-rest data that has no stimulation further mitigates concern due to confound of stimulation artefacts on the dynamic IO modelling results.

At-rest functional controllability explains the variability in the estimated response strength. We found that at-rest functional controllability also explained the variability in the estimated strength of the input-driven dynamics, which we term estimated response strength. We obtained the estimated response strength for each network node—that is, LFP power feature—as the average energy of the temporal variations in the node's predicted single-trial input-driven dynamics, which was computed by our IO model (Methods and Supplementary Fig. 19a). The estimated response strengths of the predictable power features for OFC were significantly higher than for the other stimulation sites (pairwise Kruskal–Wallis test, FDR-corrected $P < 0.05$ for all paired comparisons with OFC; Supplementary Fig. 20). Given the larger response strength values and that the functional controllability values were also larger for OFC stimulation (see previous section), we first focused on OFC-stimulation datasets for our correlation analyses. We found that at-rest functional controllability from the stimulation node was positively correlated with the estimated response strength both across brain regions and across network nodes (Pearson's $P < 0.05$ for both cases; Supplementary Fig. 19b,c). Also, both across brain regions and across network nodes, a linear regression model could significantly predict the estimated response strength from the at-rest functional controllability within cross-validation (permutation test $P < 0.05$ for both cases).

Further, beyond OFC stimulation, across all four stimulation sites, the estimated strength of the overall network response to a stimulation site correlated significantly with the overall at-rest functional controllability from that site to the rest of the network (CC $\rho = 0.95$, Pearson's $P = 0.045$; Supplementary Fig. 20b) but not with the overall physical distance from that site to the rest of the network (Pearson's $P = 0.38$; Supplementary Fig. 20c). We found the above overall network value of the controllability, physical distance and estimated response strength by averaging their values across all predictable power features in the network. Together, these results suggest that the LFP power features with stronger at-rest functional controllability from the stimulation site had stronger dynamic responses as estimated by the fitted IO model.

Nonlinear dynamic IO modelling does not outperform the linear dynamic IO models. To examine the effect of the linear form in IO modelling, we next compared the IO LSSM with a general family of nonlinear dynamic IO models and found that nonlinear models did not improve the dynamic prediction (Fig. 6a,b). We used nonlinear autoregressive with exogenous term (NLARX) models[52] within the same cross-validation procedure (Methods). Among predictable power features, IO prediction accuracy using LSSM was significantly larger than that using NLARX (0.45 ± 0.006 versus 0.34 ± 0.01, two-sided Wilcoxon signed-rank test, $P = 5.13 \times 10^{-23;}$ Fig. 6b). Also, when considering all modelled LFP power features (whether predictable or not), IO prediction accuracy using LSSM was still significantly larger than that using NLARX (two-sided Wilcoxon signed-rank test, $P = 1.10 \times 10^{-9}$). Moreover, beyond forward prediction of single-trial input-driven dynamics, for one-step-ahead prediction of the single-trial overall brain network dynamics, the LSSM again outperformed the NLARX model (two-sided Wilcoxon signed-rank test, $P < 10^{-16}$). The NLARX did not perform as well in cross-validation likely because, due to its nonlinear form, it needed to fit more parameters compared with LSSM, which made it prone to overfitting (Discussion). Indeed, even though we picked the number of parameters in LSSM and NLARX using the same optimal procedure and to maximize cross-validated IO prediction accuracy in the training data (Methods), the number of LSSM parameters per power feature was significantly smaller and only 18.84% of those in NLARX (10.97 ± 0.31 versus 58.25 ± 4.49, two-sided Wilcoxon signed-rank test, $P = 5.58 \times 10^{-40}$; Fig. 6c). Thus, while the NLARX prediction also captured the main shape of the ground-truth input-driven dynamics, it could for example lead to spurious high frequency jumps (Fig. 6a).

Controlling for the effect of stimulation-induced adaptation on IO modelling. Brain dynamics can adapt to continuous electrical stimulation[53–55]. We thus investigated stimulation-induced adaptation in brain network dynamics and their potential influence on our IO modelling. To do so, we performed additional, constant-stimulation experiments on the same day as the MN-stimulation experiment by delivering 10 min of stimulation with a constant amplitude and frequency (Supplementary Table 4). Overall, using the three analyses below, we observed that adaptation is present but the degree of adaptation is relatively small; after explicitly removing the adaptation effects, the IO prediction accuracy did not differ significantly from that in the main analyses, showing that our conclusions were not affected by adaptation.

First, we compared the functional brain network topology before and after stimulation using pre-stimulation and post-stimulation at-rest data, respectively (Supplementary Fig. 21a and Supplementary Note 18). We defined the topology as four frequency-specific coherence networks[51] at the same four frequency bands used in the main analyses. We quantified the change of each network using multiple standard network topology measures in graph theory[41,56,57] (Supplementary Note 18). Overall, for both the MN-stimulation and constant-stimulation datasets, the functional brain network topology changed after stimulation ended compared with before stimulation, but the amount of this change was relatively small (Supplementary Fig. 21b,c); in particular, the average change of coherence network topology measures (maximum range: 0 to 1) was less than 0.02 at each frequency band (details in Supplementary Note 18).

Second, consistent with the relatively small network adaptation effects, we observed only relatively small adaptation effects in LFP power feature dynamics during stimulation. To assess adaptation across trials, we computed the autocorrelations of LFP power features at delays longer than the duration of a single trial (Supplementary Fig. 22a and Supplementary Note 19). We found that for the MN-stimulation datasets, stimulation-induced changes in auto-correlation coefficient (maximum range: 0 to 1) were relatively small, with a median less than 0.015 (Supplementary Fig. 22b and details in Supplementary Note 19). The values of the stimulation-induced changes in auto-correlation coefficient in the constant-stimulation datasets were similar to those in the MN-stimulation datasets, again supporting our findings (two-sided

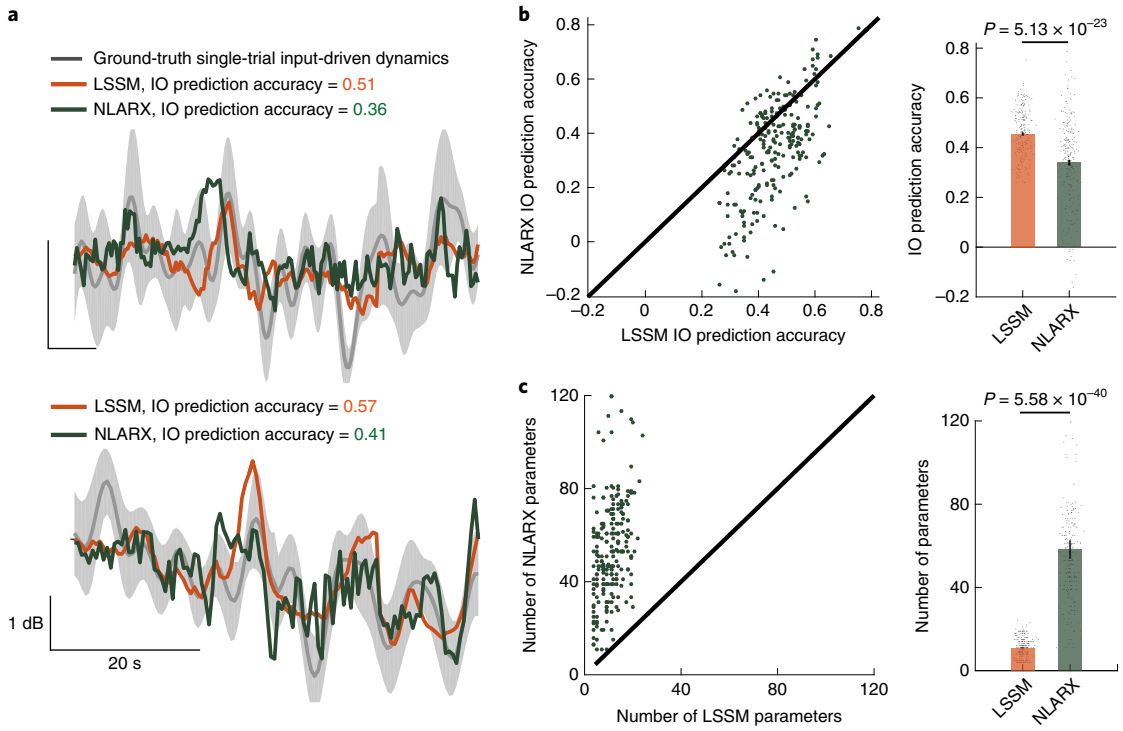

Fig. 6 | Nonlinear dynamic IO modelling does not outperform the linear dynamic IO models. a, Forward-prediction traces of example LFP power features using the nonlinear IO model NLARX are shown for monkey A (top) and monkey M (bottom). The grey shaded area provides the s.e.m. in the ground truth. **b**, Among predictable power features, the IO prediction accuracy of LSSM was significantly larger than NLARX. Left: each coloured dot shows the IO prediction accuracy of NLARX versus that of LSSM for one predictable power feature, with the 45° line representing equality. The dots are largely below the 45° line, showing the benefit of LSSM. Right: the bar represents the mean and the black error bar represents s.e.m. Raw IO prediction accuracies are shown with dots ($N = 233$ independent samples—that is, LFP power features—for both bars). Two-sided Wilcoxon signed-rank test P value is also shown. **c**, Among predictable power features, the number of parameters in LSSM was significantly smaller than NLARX. Left: each coloured dot shows the number of parameters of NLARX versus that of LSSM for modelling one predictable power feature, with the 45° line representing equality. The dots are largely above the 45° line, showing that LSSM used fewer parameters. Right: the bar represents the mean and the black error bar represents s.e.m. Raw numbers of model parameters are shown with dots ($N = 233$ independent samples—that is, LFP power features—for both bars). Two-sided Wilcoxon signed-rank test P value is also shown.

Wilcoxon sum-rank test, $P = 0.78$; Supplementary Fig. 22b and details in Supplementary Note 19).

Third, we hypothesized that our IO prediction accuracy was probably unaffected by adaptation effects given (1) the relatively small stimulation-induced changes in auto-correlation coefficient and (2) the fact that the input remains independent in training and test sets in our cross-validation design, even in the presence of adaptation (Supplementary Fig. 5b and Supplementary Note 9). To test this hypothesis, before IO modelling, we first high-pass filtered the LFP power features at a frequency equal to the inverse of the duration of a single trial. This filtering suppressed any autocorrelations beyond the duration of a trial. We then repeated the IO modelling procedure. We found that the IO prediction accuracy did not change significantly compared with our main analyses (two-sided Wilcoxon signed-rank test, $P = 0.41$; Supplementary Fig. 22c and Supplementary Note 19), consistent with our hypothesis.

The LSSM could also predict the intrinsic dynamics of at-rest LFP power features. Beyond IO modelling, we found that the same LSSM framework could also predict the intrinsic dynamics of at-rest LFP power features measured across the entire network without stimulation (data and method details in Supplementary Note 11). In our at-rest LFP datasets and for each LFP power feature, we built a special case of LSSM, termed at-rest LSSM, which did not have the stimulation input term. Across all at-rest LFP power features, the cross-validated CC between the one-step-ahead-predicted and true LFP power feature was 0.83 ± 0.006 (output-permutation test, $P < 10^{-16}$; Supplementary Fig. 23), suggesting the ability to predict intrinsic dynamics. Also, the state transition matrix of the at-rest LSSM was similar to that of the IO LSSM during stimulation (Supplementary Fig. 24), suggesting that during-stimulation intrinsic dynamics are constrained by at-rest intrinsic dynamics.

The fitted dynamic IO models enable closed-loop control of a simulated internal brain state. An important application of the dynamic IO models would be to enable model-based closed-loop control of an internal brain state such as mood or pain[1,3,18,21,22] (Discussion). We extensively analysed the ability of the fitted dynamic IO models from our data for closed-loop control using a comprehensive numerical simulation study (Figs. 7 and 8) and found that they can enable closed-loop control as detailed below.

Guided by previous work in neural decoding of mood state[22], we designed realistic closed-loop simulations (details in Supplementary Note 20). In these simulations, the same fitted IO models from our IO data here were used to build a model-based closed-loop controller, which was required to take a simulated internal brain state (for example, mood state) to a target level (Fig. 8a). To enforce the same IO prediction accuracy in simulations as observed in our IO data, we introduced a mismatch between the simulated brain model and our fitted IO model that was used in the controller, and we also included model noises in the simulated brain (Fig. 8a and Supplementary Note 20). To simulate a worst-case scenario, we

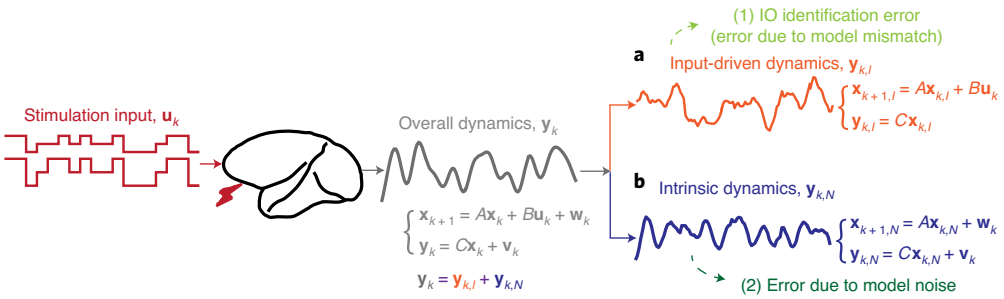

Fig. 7 | The overall brain network dynamics can be decomposed into input-driven dynamics and intrinsic dynamics to explain two possible sources for forward-prediction error. a,**b**, The overall brain network dynamics (middle panel, grey, see equation (1) in Methods) is decomposed into two parts: dynamics that are driven by the stimulation input termed input-driven dynamics (**a**; right, orange; equation (2)) and dynamics that are not driven by the stimulation input termed intrinsic dynamics (**b**; right, blue; equation (3)). Such a decomposition explains the two possible sources of the forward-prediction error for a fitted IO model (Methods): (1) IO identification error or equivalently the error due to a mismatch between the fitted IO model and the actual IO relationship in the brain (right, top, light green) and (2) model noise (right, bottom, dark green). A given value of forward-prediction error can be due to any combination of the above two sources. These two sources of error differently affect the performance of closed-loop controllers that are designed from the fitted IO models. It is the errors due to the first source that would affect closed-loop control performance in the worst way because the feedback controller is directly designed based on the identified IO relationship. Thus, in worst-case closed-loop control simulations, we make the IO identification error (source (1)) as large as the entire amount of forward-prediction error by including a mismatch and in addition we also include model noise (source (2)) in the simulated brain as large as that fitted in the fitted IO model. Best-case model-based closed-loop control occurs when the model noise (source (2)) leads to the forward-prediction error of our fitted IO models. This is because our fitted IO models quantify the noise with the state noise term $\mathbf{w}_k$ and the observation noise term $\mathbf{v}_k$. and thus the controller can account for them. Thus, in best-case closed-loop control simulations, we set the IO identification error (source (1)) to zero while keeping the simulated model noise (source (2)) as large as that fitted in the fitted IO model. Details are presented in Supplementary Note 20.

picked the mismatch large enough to recreate the entire amount of forward-prediction error observed in our IO data; further, on top of this mismatch, we also added stochastic noises to the simulated brain, where the noise covariance was as large as what was fitted in the fitted IO model (Figs. 7 and 8a). We also simulated a best-case scenario in which the forward-prediction error was due to stochastic noise (but there was no mismatch; Figs. 7 and 8a and Supplementary Note 20). The actual performance of the controller would thus be between the worst-case scenario and the best-case scenario. We quantified the control error of a controller as the root mean square error between the controlled internal brain state and its target level. We defined the baseline of this error as the error when no stimulation was applied. Dividing the control error by the baseline root mean square error, we obtained a normalized control error, which was 0 for perfect control and 1 for the baseline of no stimulation.

We found that with the same IO prediction accuracy observed in our data (0.45 ± 0.006 CC or $21.98\% \pm 0.01\%$ EV), the fitted IO models from our data achieved closed-loop control of the simulated internal brain state. On average, the normalized control error using model-based controllers designed from our fitted IO models was between 0.18 and 0.42, significantly smaller than the baseline of 1 (best-case and worst-case performances, respectively; two-sided Wilcoxon signed-rank test, $P < 10^{-30}$ in both cases; Fig. 8b,c). Importantly, the IO model was critical in achieving closed-loop control. We specifically compared performance with a model-free closed-loop controller that did not incorporate an IO model and only used feedback of brain network activity to turn stimulation on when the estimated internal brain state fell below the target level[1–3]. We found that even in their worst-case scenario, the model-based closed-loop controllers significantly outperformed model-free closed-loop controllers, which had on average almost twice the normalized control error (0.42 ± 0.02 versus 0.80 ± 0.05, respectively, two-sided Wilcoxon signed-rank test, $P = 4.1 \times 10^{-24}$; Fig. 8b,c). Finally, by sweeping the forward-prediction error in the simulations, we found that closed-loop control performance significantly improved as the simulated IO prediction accuracy increased (Spearman's rank correlation, Spearman's $P < 10^{-30}$; Fig. 8d), further demonstrating the importance of the IO model in enabling closed-loop control.

Discussion

Here we provide the first demonstration that large-scale multiregional brain network responses to ongoing temporally varying electrical stimulation can be predicted by developing data-driven dynamic IO models. Further, we demonstrate that the variability in estimated response strength and in IO prediction accuracy across different network nodes can be explained by their at-rest functional connectivity to the stimulation node, which is computed as the functional controllability of our models fitted to at-rest activity. Finally, our models reveal and characterize the oscillatory and damping dynamics of the brain network response to stimulation and enable model-based closed-loop control of a simulated internal state (Fig. 8).

Characteristics of the dynamic brain network response to stimulation. We found that the brain network response to stimulation was dynamic and dependent on the history of stimulation. Further, the dynamic structure of the IO model that quantified complex brain network dynamics in response to stimulation—with both oscillatory and damping characteristics—was key to achieving accurate predictions. Damping dynamics indicate that stimulation inputs that are further in the past have a smaller contribution to current brain network activity (Fig. 4c and Supplementary Fig. 13). The damping dynamics that we found during stimulation are in line with previous studies showing that after stimulation ends, the effect on brain activity washes out[13,44]. Oscillatory dynamics indicate that the contribution of past inputs to current brain network activity oscillates at a certain period into the past (Fig. 4c and Supplementary Fig. 13). In the absence of stimulation, brain activity underlying many brain functions such as movements[58–60], speech[61], memory[62,63], attention[64,65], decision making[66,67], learning[68] and sensory processing[69,70] has been shown to involve structured dynamics whose modelling has led both to improved understanding of the underlying neural

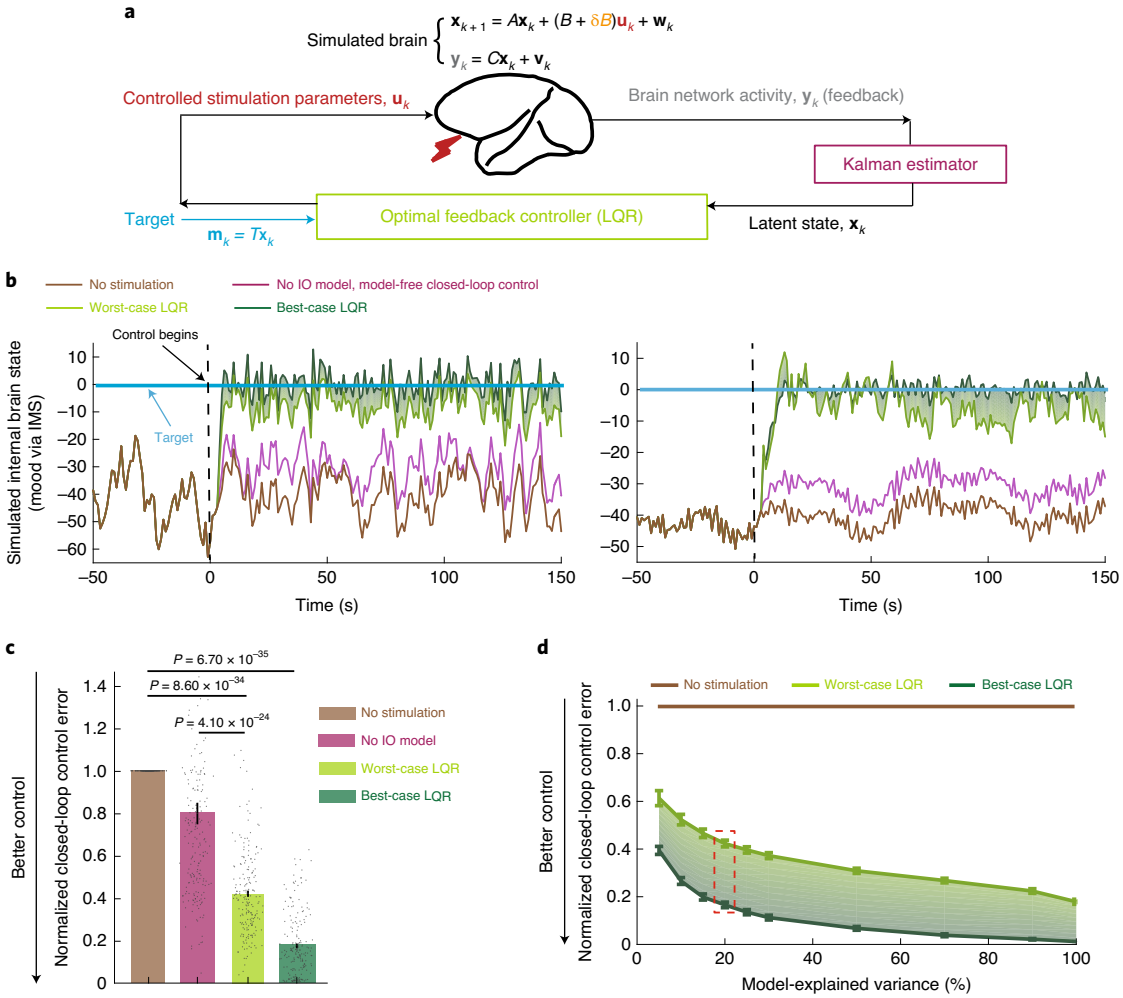

Fig. 8 | The fitted IO models enable closed-loop control of a simulated internal brain state. a, The fitted IO models to our datasets were used to conduct closed-loop simulations. A mismatch δB in the input matrix (orange) was introduced in the simulated brain model (top equation) to produce an IO identification error as large as the entire amount of forward-prediction error observed in our datasets. Model-based closed-loop controllers were designed based on the original fitted IO models (equation (1)) and consisted of a Kalman state estimator and a feedback controller. We used the optimal linear quadratic regulator (LQR) as the feedback controller. Brain network activity was used as feedback and the model-based closed-loop controller identified the stimulation parameters in real time to drive the internal brain state to a particular target (Supplementary Note 20). **b**, Control of a simulated internal brain state using two example fitted IO models from our datasets. As an illustrative example, we take the internal brain state to be a mood state related to depression and anxiety symptoms measured with the validated immediate mood scalar (IMS) questionnaire that has been used in previous work (Supplementary Note 20). The control performance using our fitted IO model is expected to lie within the shaded green area—that is, between the best case and worst case. Model-free closed-loop control (magenta) did not use an IO model but used feedback of brain network activity to turn stimulation on and off. **c**, Bar plot showing the closed-loop control performance across all fitted IO models in our datasets and for different control scenarios. The case with no stimulation provides the baseline performance. Control performance using our fitted IO models is expected to lie between the best case and worst case. The bar represents the mean and the black error bar represents s.e.m. Raw IO prediction accuracies are shown with dots ($N = 233$ independent samples—that is, LFP power features—for all bars). Two-sided Wilcoxon signed-rank test P values are shown. **d**, In simulations, worst-case and best-case model-based closed-loop control performances both changed as a function of the model-explained variance (EV) in forward prediction and the actual performance is expected to lie within the shaded green area. Solid lines represent mean and whiskers represent s.e.m. ($N = 233$ independent samples—that is, LFP power features—for all error bars). Red dashed box represents the performance associated with the EV observed in the actual IO datasets.

mechanisms[60,71–73] and to decoding of movement[1,58,73–78], speech[61] and mood[22]. Similarly, the design of future neuromodulation techniques to control brain function may benefit from the ability to precisely model the dynamic brain network response to stimulation, as we demonstrate here.

Learning a dynamic IO model. A critical component that allowed us to learn the dynamic IO model was the use of a stochastic MN-modulated pulse train that sufficiently excited brain network activity (that is, was white in amplitude and frequency space)[18]. Some previous stimulation studies have used single stimulation pulses to elicit cortico-cortical evoked potentials[79] with a distinct goal of studying brain connectivity[79]. Here we focused on modelling the brain network dynamics in response to ongoing stimulation pulse trains because pulse-train stimulation has been effective for treating neurological disorders[5,6,44] and holds promise for treating neuropsychiatric disorders[1,7,8]. For example, pulse trains with a fixed amplitude and frequency have been used to examine the changes in brain activity following or during stimulation[13,14,17,44], although without modelling. In contrast to these fixed pulse trains, in our

experiments, the MN-modulated pulse train is designed to have stochastically changing amplitude and frequency with time to sufficiently excite brain activity. We found that real-time changes in both amplitude and frequency modulated the dynamic network response, which could be accurately predicted by the dynamic IO model. Moreover, we found that our fitted dynamic IO models are invertible because they are controllable (Fig. 8 and Supplementary Note 21): given a target internal brain state, the dynamic IO model can analytically compute how stimulation amplitude and frequency should be changed over time to achieve the target. This result can help facilitate closed-loop neuromodulation of internal brain states[1,18].

Another key component of our IO modelling is the LSSM structure. This structure describes brain network dynamics in terms of a latent state that exhibits both intrinsic dynamics as well as dynamics driven by the stimulation input. In many neuromodulation applications, the time available for collecting IO data for machine learning is limited to avoid delaying therapy and to protect the subject[18]. Thus, a critical benefit of the LSSM model structure is its linear form with fewer parameters than a dynamic nonlinear IO model (NLARX[52]) whose performance suffered compared with the LSSM. Nevertheless, an important question for future studies is whether nonlinear models can improve the prediction of brain network response in cases where stimulation time is less constrained.

Dynamic IO modelling and design of closed-loop neuromodulation systems. Closed-loop neuromodulation systems could improve the efficacy of DBS treatments, especially for neuropsychiatric disorders for which open-loop stimulation has been variably efficacious, such as major depression[80,81]. Further, closed-loop neuromodulation can also facilitate brain functions such as memory[82,83]. Dynamic modelling is especially important in such closed-loop scenarios, where stimulation parameters such as frequency and amplitude need to change in real time[1–3,18,21,84,85] unlike open-loop scenarios, which feature a fixed stimulation pattern. Closed-loop changes in stimulation need to be guided (1) by real-time feedback of brain activity and internal brain state such as mood or pain level[1,18,21,22], and (ii) by an IO model to predict how a change in stimulation would alter brain activity and thus the internal brain state[1]. Thus, in these applications, the IO model needs to solve two major challenges. First, it has to predict the neural response during ongoing and temporally varying stimulation rather than after stimulation ends or to fixed stimulation patterns. Second, it must make this prediction across large-scale multiregional brain networks—given their involvement in many neuropsychiatric disorders[1–3,22,24–26,86]—rather than within a particular brain region. We demonstrate such a prediction to provide an enabling technology for facilitating the design of future closed-loop neuromodulation systems for neuropsychiatric and neurological disorders.

To evaluate the fidelity of our IO models for closed-loop control, we tested them in an extensive closed-loop simulation study (Supplementary Note 22 expands further on the model fidelity). These simulations show that the fitted IO models to our data with the exact same IO prediction accuracy observed in our data can enable model-based closed-loop control of simulated internal brain states. While beyond the scope of this work, future experiments performing model-based closed-loop control of large-scale multiregional brain networks encoding various internal brain states such as mood or pain in neuropsychiatric disorders are critical research directions. The dynamic IO models provide a necessary and general enabling technology to help realize such model-based closed-loop controllers for internal brain states across many neuropsychiatric disorders.

Finally, another special application of our dynamic IO model is to predict the steady-state brain network response—that is, the response to stimulation pulse trains at a fixed amplitude and frequency. Current experiments typically perform a grid search over stimulation amplitudes and frequencies to investigate these steady-state responses[13,17]. A grid search may suffer from a large search space and thus be inefficient. A more efficient approach could be to predict the steady-state brain network response by setting the amplitude and frequency input to a fixed value in our dynamic IO model once it is fitted. Future investigations of how well such an approach works may also guide the optimal selection of stimulation parameters in open-loop neuromodulation systems in an efficient way.

Stimulation-induced neural adaptation, state dependencies and non-stationarities. Continuous fixed electrical stimulation patterns can induce neural plasticity and adaptation[14], leading to non-stationary brain network dynamics. In this first demonstration of dynamic IO modelling, we show that a time-invariant IO model can predict the dynamic brain network response to continuous real-time changing stimulation lasting 10 min to 120 min. While our IO modelling does not aim to develop a stimulation paradigm that targets neural plasticity, such a plasticity-based paradigm is a critical complementary approach for developing future therapies. Previous studies have designed open-loop stimulation patterns that target neural plasticity via sufficiently desynchronizing neural populations[53,54] to achieve symptom alleviation in Parkinson's disease[55]. Our approach that directly models the IO dynamics can be combined with the approach that targets neural plasticity via adapting the IO model parameters. Also, in practical applications, non-stationary brain network dynamics may occur due to state dependency, for example, due to a change in psychiatric state[26]. To track the non-stationarity induced by plasticity and state dependency, future work can develop adaptive LSSMs[87] that also include an input to improve IO prediction performance and benefit closed-loop neuromodulation systems that need to operate over long time periods.

Dynamic IO modelling and at-rest functional controllability. Across the large-scale multiregional brain network, at-rest functional controllability explained the variability in both the IO prediction accuracy and the estimated response strength by the IO models during stimulation. This result provides an independent validation of our dynamic IO models with a measure that is independently computed from at-rest data without any stimulation. Theoretically, this controllability measure quantifies the energy needed at a stimulation node to change the neural activity at a network node by a given amount[46] (Supplementary Note 17). High controllability indicates that it is easier to drive a network node from the stimulation node and provides a useful connectivity measure. Previous studies have addressed controllability of structural brain networks on the basis of tractography derived from DTI/DWI data[41,42,47–50] to understand how structural brain networks constrain function and behaviour in human subjects[49,50] or constrain functional connectivity in numerical or computer simulations of stimulation[48]. Controllability of structural brain networks has also been shown to be correlated with the static change in functional connectivity[41] or in brain activity[42] that occurs after stimulation ends. However, it is not clear whether controllability is correlated with the strength of dynamic changes in brain network activity during ongoing stimulation. Moreover, DTI/DWI data provide a measure of anatomical connectivity rather than network function, which can instead be provided by large-scale recordings of LFP activity or more generally other measures of network electrophysiological activity[51]. Here, by introducing the functional controllability measure using at-rest electrophysiological LFP data and relating it to the IO models, we demonstrate that this at-rest functional measure explained the variability in IO prediction accuracy across network nodes. At-rest functional controllability also predicted—in cross-validation—which network nodes

would have a predictable response during stimulation from OFC. Further, among all four stimulation sites, at-rest functional controllability explained which stimulation site (that is OFC) most strongly modulated the estimated network response. By contrast, physical distances between recording and stimulation sites did not provide these explanations, suggesting that the brain network responses reflect multiregional directed functional interactions.

LFP network responses to stimulation can vary for several reasons. LFP network responses depend on (1) the strength of excitatory (for example, AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid)) and inhibitory (for example, GABA (γ-aminobutyric acid)) mediated post-synaptic potentials; and (2) the spatial distribution of excitatory and inhibitory post-synaptic potentials in response to pre-synaptic inputs from the stimulation site[51]. Stimulation may change the relative strength of excitatory and inhibitory post-synaptic potentials[51] differently across responding electrodes. Stimulation may also induce different spatial reorganization of excitatory and inhibitory post-synaptic potentials that changes the strength of effective dipole moment generating the measured LFP activity at different responding electrodes[88]. Each of these possibilities can increase LFP response variability. For a given stimulation site, microstimulation may cause both orthodromic and antidromic stimulation effects[89]. Therefore, the observed variable dynamic LFP responses distributed across various brain regions reflect multiregional directed functional interactions via anatomical pathways that are connected through white matter fibre tracts. Our results thus highlight how LFP power features could reflect modulation of convergence and divergence of cortico–subcortical limbic communication during MN stimulation. These anatomical pathways involve cortico–basal ganglia, cortico–thalamic and basal ganglia–thalamic projections[90] that link the stimulation and responding sites that we have identified (Fig. 3). Future work will need to integrate electrophysiology tools with calcium imaging and other genetic tools to label anatomically identified projection systems and more completely assess the connectivity between the labelled stimulation and responding sites[91,92].

Determining where to stimulate is a major challenge in developing neuromodulation therapies and different stimulation sites have had variable efficacy for example in treating major depression[8]. In this work, to demonstrate our dynamic IO modelling technique as a general enabling technology, we stimulated from four different sites (OFC, ACC, AMG and SPL) and showed that our models can generalize to predicting the brain network response to stimulation at all four sites. The most frequently tested stimulation site in our study was OFC (10 out of the 16 IO datasets), which has recently been suggested as a potential stimulation target for improving mood in human subjects with depression symptoms[13] and has been shown to be important in decoding mood state from human subjects[22]. Future studies focused on neuromodulation in depression can, for example, use the dynamic IO models to help predict the brain network response during closed-loop OFC stimulation. Among our four stimulation sites, OFC stimulation more robustly modulated the network—larger controllability values and response strengths, which could be reliably estimated. Thus, we focused our correlation analyses of functional controllability and performance on the OFC-stimulation datasets. Future work that explores a broader range of stimulation sites and collects longer and more stimulation datasets for each site can compute the functional controllability measure introduced here to reliably study its relationship to performance for different stimulation sites. Moreover, our results suggest that computing the at-rest functional controllability from large-scale LFP activity may help identify the optimal stimulation location in DBS without applying stimulation there and only from at-rest neural activity. This can be done by predicting which brain site best modulates the networks involved in a given disorder using at-rest activity and functional controllability, for example, the networks that encode mood in depression[13,22].

Once a stimulation site is selected, assessing the IO models should focus on the power features within a sampled brain network that actually have a response to that site to be modelled. We aimed to focus on these relevant power features by excluding the noisy IO prediction accuracy values associated with the non-predictable power features that were below the noise floor (did not pass the statistical tests). Nevertheless, as a control analysis, we showed that our conclusions also held when considering all power features, showing the robustness of our results.

Controlling for stimulation artefacts. Three lines of evidence confirm that our conclusions about the IO models are not confounded by stimulation artefacts. (1) The stimulation-artefact control analyses (Supplementary Note 1) showed that the stimulation-artefact rejection algorithm only induced a negligible change (<5%) in the studied spectral content of LFP signals (Supplementary Fig. 2g). (2) The same stimulation-artefact rejection algorithm was included in the output-baseline test; thus if the prediction of brain network response was simply caused by the small residuals of the stimulation-artefact rejection algorithm, the IO prediction accuracy in the output-baseline test would not have been different from the actual IO prediction accuracy and the output-baseline test would have failed. Since all reported predictions pass the output-baseline test, this test rules out the effect of stimulation artefacts. (3) Since at-rest LFP signals do not have stimulation artefacts, the fact that at-rest functional controllability can predict the IO prediction accuracy during stimulation provides an independent validation of the IO modelling results.

Future directions. We chose to model the LFP power features at the frequency bands spanning from 1 Hz to 100 Hz because they are informative of various brain states (for example, intended movements[43,51,58,73] or mood[1,22,23]) and disease symptoms (for example, for depressed mood[1,2,13] or in Parkinson's disease[2,44]), and because they change when electrical stimulation with fixed amplitude and frequency is delivered[13,20]. Thus these power features can provide clinically relevant biomarkers[1,2,20,22] (Supplementary Note 2). Beyond modelling power features, it is also important to study the utility of our IO modelling framework for other clinically relevant field-potential features such as phase-amplitude coupling[2,44] or coherence[23]. Field potentials are an important recording modality in clinical applications given their availability and reliability[1,51], which supports clinical feasibility. Beyond field potentials, the IO modelling framework could be extended to single-unit recordings[1,12,73,74] or to multiscale combined single-unit and field-potential recordings[1,43,73,75,93,94]. Such investigations can also help facilitate neuromodulation therapies with these other features or recordings as biomarkers.

We modelled the LFP response to microstimulation in monkeys to develop and demonstrate a general enabling technology. An important future direction is to develop dynamic IO models for brain network responses in the human brain, by performing simultaneous recording and stimulation with ECoG electrodes[1,13,17,18,22,23,61,87,95]. Although the effect of stimulation-induced adaptation was relatively small in our data, developing adaptive IO modelling techniques is important for tracking adaptation and state-dependency effects, further improving the IO prediction accuracy, and combining with plasticity-based approaches[1,53–55,87]. Moreover, with typical data durations for model learning, we find that general nonlinear IO modelling does not improve the linear dynamic IO model. Future long-term chronic stimulation and recordings experiments can be used to further explore the nonlinearity in the IO response when more training data are available. To demonstrate an enabling technology, here we designed our MN to switch across multiple values

in the amplitude range of 0–40 μA and in the frequency range of 0–100 Hz, which are both within the clinical ranges used for DBS, for example, in depression[8] (details in Supplementary Note 3). Similarly, we used an amplitude range of 10–50 μA and a frequency of 100 Hz for constant-stimulation experiments. Future work could investigate other important frequency ranges that are effective for DBS in other applications—for example, 100-185 Hz for Parkinson's disease[96]—and the damping and oscillatory characteristics of the IO response, as well as the adaptation effects in those ranges. Finally, future work could address how these data-driven models may also help guide the development of more mechanistic biophysical models of DBS.

Together, our study shows that large-scale brain network dynamics across multiple brain regions and in response to real-time changes of stimulation parameters can be accurately predicted. These results have important implications for the design of closed-loop neuromodulation systems for treatment of a variety of neurological and neuropsychiatric disorders and for the more precise modulation of brain functions.

Methods

Experimental preparation. Two male rhesus macaques (*Macaca mulatta*) participated in the study (monkey M, 8.4 kg; monkey A, 7 kg). All surgical and experimental procedures were performed in compliance with the National Institute of Health Guide for Care and Use of Laboratory Animals and were approved by the New York University Institutional Animal Care and Use Committee. We surgically implanted a customized large-scale recording chamber over the targeted brain regions over the left (monkey M) and right (monkey A) hemispheres (Gray Matter Research) using magnetic resonance (MR)-guided stereotaxic surgical techniques (Brainsight, Rogue Research). We then mounted a customized semi-chronic microdrive system for large-scale circuit mapping in macaque mesolimbic and basal ganglia systems[97] into the chamber and sealed it with compressed gaskets and room-temperature-vulcanizing sealant (734 flowable sealant, Dow Corning).

The microdrive provided bi-directionally independent control of the position of up to 220 microelectrodes (1.5 mm spacing) along a single axis with a range up to 32 mm (monkey M) and 40 mm (monkey A) with 125 μm pitch[98]. The microdrive provided access to the cortico–subcortical limbic network, including OFC, ACC, prefrontal cortex, motor cortices, parietal cortex, caudate nucleus, putamen, globus pallidus and AMG (Supplementary Tables 1 and 2). For monkey M, we loaded 160 platinum–iridium (Pt/Ir) electrodes (MicroProbes) with impedance 0.1–0.5 MΩ for recording and microstimulation and 60 tungsten electrodes (Alpha Omega) for intracortical recording with impedance 0.8–1.2 MΩ. For monkey A, we loaded 220 Pt/Ir electrodes (MicroProbes) with impedance 0.5 MΩ for recording and microstimulation. Electrode impedances were measured at 1 kHz (Bak Electronics).

Neural recordings. The monkeys were awake, head-restrained and quietly seated in a primate chair placed in an unlit sound-attenuated electromagnetically shielded room (ETS Lindgren). Neural recordings were referenced to a ground screw implanted in the left posterior parietal lobe (monkey M) or left occipital lobe (monkey A). The ground screw was chronically implanted, resting on the dura mater. Neural signals from all channels were simultaneously amplified and digitized at 30 kHz with 16 bits of resolution with the lowest significant bit equal to 0.1 μV (NSpike, Harvard Instrumentation Lab; unit gain headstage, Blackrock Microsystems) and continuously streamed to disk during the experiment (also see Supplementary Table 3).

Dynamic IO model structure. We build a multiple-input–multiple-output model to describe the dynamic effect of stimulation amplitude and frequency (input) on the brain network activity (output). We define the input as a multi-component vector time series $\{\mathbf{u}_k\}$ ($\{\cdot\}$ represents a time series or a set depending on the context) of stimulation amplitude and frequency—that is, $\mathbf{u}_k = \left[u_k^{\text{amp}}, u_k^{\text{freq}}\right]' \in \mathbb{R}^{N_u \times 1}, N_u = 2$—where u_k^{amp} and u_k^{freq} represent the stimulation amplitude and frequency, respectively, and superscript $'$ represents matrix and vector transpose. Here, k represents the discretization time step, which was either 1 s or 0.5 s (T_s in Supplementary Table 3). We define the output as the LFP power features calculated over time from multiple brain regions and put them into a multi-component vector time series $\{\mathbf{y}_k \in \mathbb{R}^{N_y \times 1}\}$. We construct a dynamic multiple-input–multiple-output LSSM as a series of multiple-input-single-output models, each modelling a component in $\{\mathbf{y}_k\}$ (Supplementary Fig. 3). The LSSM corresponding to the ith LFP power feature component is written as:

$$\begin{cases} \mathbf{x}_{k+1}^{(i)} = A^{(i)}\mathbf{x}_k^{(i)} + B^{(i)}\mathbf{u}_k + \mathbf{w}_k^{(i)} \\ \mathbf{y}_k^{(i)} = C^{(i)}\mathbf{x}_k^{(i)} + \mathbf{v}_k^{(i)} \end{cases}, \quad (1)$$

where $\mathbf{y}_k^{(i)} \in \mathbb{R}$ represents the ith component of $\mathbf{y}_k$, $\mathbf{x}_k^{(i)} \in \mathbb{R}^{N_x^{(i)} \times 1}$ is a dynamic latent state[18] (it is dynamic because it is directly related to its own past values via the top state equation) and $\mathbf{w}_k^{(i)} \in \mathbb{R}^{N_x^{(i)} \times 1}, \mathbf{v}_k^{(i)} \in \mathbb{R}$ are zero-mean white Gaussian noise with covariance matrix $Q^{(i)} = \mathbb{E}\left[\begin{pmatrix} \mathbf{w}_k^{(i)} \\ \mathbf{v}_k^{(i)} \end{pmatrix} \begin{pmatrix} \mathbf{w}_k^{(i)\prime} & \mathbf{v}_k^{(i)} \end{pmatrix}\right] \in \mathbb{R}^{(N_x^{(i)}+1)\times(N_x^{(i)}+1)}$. Compared with previous latent state dynamic models, which do not incorporate an input term[22,76,87,95], here the model learns the effect of external stimulation input on neural dynamics through the latent state. For data-driven modelling in general[18,22,43,61,73,75–78,87,95,99,100], the model parameters should be fitted to data. The model parameters here that need to be fitted are $A^{(i)} \in \mathbb{R}^{N_x^{(i)} \times N_x^{(i)}}, B^{(i)} \in \mathbb{R}^{N_x^{(i)} \times N_u}, C^{(i)} \in \mathbb{R}^{1 \times N_x^{(i)}}$ and $Q^{(i)}$, which we collect as $\theta^{(i)} = \{A^{(i)}, B^{(i)}, C^{(i)}, Q^{(i)}\}$, and a hyperparameter $N_x^{(i)}$, which is the dimension of the latent state. For simplicity of expression, we omit the superscript i unless it is not clear from the context that one power feature is being modelled. The number of free parameters in this multiple-input–single-output LSSM in equation (1) is $2N_x^{(i)} + N_u N_x^{(i)} + 1$[52]. In the next few sections, we introduce the design of the stochastic input time series $\{\mathbf{u}_k\}$ to sufficiently excite the brain network in collecting IO data to identify the model parameters, the calculation of the output LFP power feature time series $\{\mathbf{y}_k\}$ and the cross-validation procedure to fit and evaluate the IO model in equation (1) using the collected IO data.

Stochastic stimulation input design. The design of the input waveform is critical to enable accurate IO modelling in machine learning[18,52]. An input waveform needs to sufficiently excite the brain network activity so that the IO dataset is informative of the network response[18,52], while consisting of electrical pulses that are charge-balanced in short time intervals as required for stimulation safety[101]. Based on our prior theoretical work[18], we designed an MN-modulated stimulation waveform and delivered it to the stimulation site (Fig. 1a). This waveform consisted of pulse trains whose amplitude and frequency were changed stochastically over time between several fixed levels with equal probability, a pattern we refer to as an MN pattern. An MN pattern has a white spectrum because at each time point the level is generated independently of its past levels[18,52] (next paragraph and Supplementary Fig. 4) and thus can excite the brain network dynamics across all timescales. We constructed the MN-modulated stimulation waveform with standard charge-balanced pulses currently used in direct electrical stimulation[1,4,13,17,18,102].

We used three amplitude levels: 0 μA (no stimulation), 15 μA and 30 μA, and three frequency levels: 0 Hz (no stimulation), 50 Hz and 100 Hz in the construction of the MN-modulated waveform in 9 out of 16 experiments (see Supplementary Note 3 for the choice of these levels). So for a minimum switch period of T_{sw}, at each time point $t = nT_{sw}$ ($n = 1,2,3,\ldots$), we randomly picked the MN parameters from among the following paired values: {(0 μA,0 Hz),(15 μA,50 Hz),(15 μA,100 Hz),(30 μA,50 Hz),(30 μA,100 Hz)} with probabilities of $\{\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$. We chose these pair probabilities such that the amplitude on its own had equal probability to be at each of its possible three levels and similarly for the frequency. We also delivered a two-level MN waveform in the other 7 out of 16 experiments with amplitude levels of 20 μA and 40 μA, and frequency levels of 50 Hz and 100 Hz (Supplementary Note 3). We applied the MN-modulated microstimulation waveform using a bipolar configuration, made by simultaneously sending a biphasic charge-balanced square wave pulse via a pair of Pt/Ir microelectrodes with a fixed pulse width of 100 μs per phase and inter-pulse interval of 53 μs, but opposite polarity (Cerestim R96, Blackrock Microsystems; Fig. 1a). Finally, the switch period T_{sw} was selected from among the following values {0.5 s, 1 s, 4 s, 6 s} to help study the time-scale of the effect of stimulation (Supplementary Note 14).

Stimulation-artefact rejection. Recorded broadband raw signals during stimulation contained stimulation artefacts. We developed an algorithm to reject the stimulation artefacts (Supplementary Fig. 2 and Supplementary Note 1), which typically consist of a short 'instantaneous artefact spike' and a long 'artefact tail' introduced by the analogue circuitries in the amplifier[17,103,104]. We used thresholding to detect the instantaneous artefact-spike timings and then replaced an epoch of 1.17 ms around the instantaneous artefact spike (0.67 ms before and 0.5 ms after) with raw signals recorded immediately before and after the instantaneous artefact spike. This replacement ensured that the amplitude and spectral distribution of the replaced raw data were similar to those of the background raw signals[17]. Then, we used a temporal template subtraction algorithm[103] to reject the artefact tail. Finally, we conducted a control analysis to show that this stimulation-artefact-rejection algorithm introduced only negligible distortion in the signal below 50 Hz (Supplementary Fig. 2g and Supplementary Note 1).

LFP signal preprocessing. After stimulation-artefact rejection, we obtained the LFP signals by (1) downsampling the raw signals to 200 Hz with an anti-aliasing filter with a cut-off frequency of 100 Hz (order 8 Chebyshev type I infinite impulse response (IIR) filter); (2) high-pass filtering above 1 Hz to remove drift (equiripple finite impulse response (FIR) filter with passband cut-off frequency 1 Hz, stopband cut-off frequency 0.5 Hz); (3) removing the line noise at 60 Hz (band-stop equiripple FIR filter with stopband cut-off frequency 59 Hz and 61 Hz,

passband cut-off frequency 58 Hz and 62 Hz); (4) band-stop filtering at 50 Hz to remove any possible residue of stimulation artefacts at the stimulation frequency (equiripple FIR filter with stopband cut-off frequency 49 Hz and 51 Hz, passband cut-off frequency 48 Hz and 52 Hz). Next, we digitally referenced the LFP signal of each channel to its nearest neighbour within 3 mm based on the electrode depths and removed recording channels that had any of these properties: (1) were the stimulation channel and its return channel because they did not record any useful neural signal during stimulation; (2) were not in the brain or were located in white matter as indicated by the co-registered MRI labels; (3) had a large amount of noise during stimulation (during-stimulation s.d. of LFP signals >5× s.d. of pre-stimulation LFP signal).

Note that the number of modelled recording channels after preprocessing in each IO dataset was different (Supplementary Table 3), mainly because the IO datasets were collected from different days and the electrodes in the microdrive were moved across different experiments and days to target and sample different brain regions. Such independently movable electrodes were a key feature of our semi-chronic microdrive design to flexibly assess large-scale multiregional brain networks. Thus on each day, a different set of recording channels were within the grey matter that we modelled and the rest of the electrodes that were in the white matter were removed by our preprocessing step above.

LFP power feature calculation. After stimulation-artefact rejection and obtaining the LFP signals, we extracted the LFP power features. For each LFP channel, we applied a multi-taper spectrogram analysis[105] with a time step $T_s = 1$ s or 0.5 s, an overlapping moving window of $10T_s$ and a frequency resolution of 2 Hz, such that a sufficient number of orthogonal tapers were used to reduce the variance of power spectral density estimation of the LFP signals[105] (Supplementary Note 2). At each time step, we calculated the mean powers in the following 4 frequency bands: 1–8 Hz (delta + theta), 8–12 Hz (alpha), 12–30 Hz (beta) and 30–50 Hz (low gamma). We chose to model these bands because they have been shown to have an important role in many brain functions[1,13,22,43,58,73,95,106–109] and dysfunctions[1,2,44,45,110–112]. We did not focus on modelling higher frequency bands because of the larger stimulation-artefact residues for those bands after artefact rejection (Supplementary Fig. 2g); nevertheless, after controlling for the stimulation artefacts, our control analyses showed that the same IO modelling framework also succeeded for the higher frequency band of 70–100 Hz. We collected the logarithm of the mean powers of each frequency band of each LFP channel in the vector $\mathbf{y}_k \in \mathbb{R}^{N_y \times 1}$. We then removed the mean in the input stimulation time series $\{\mathbf{u}_k\}_{k\in\{1,2,3,\dots,T\}}$ and output LFP power feature time series $\{\mathbf{y}_k\}_{k\in\{1,2,3,\dots,T\}}$ to form the final IO dataset $\{\mathbf{u}_k, \mathbf{y}_k\}_{k\in\{1,2,3,\dots,T\}}$, where $T = 3{,}230 \pm 528$ was the total number of samples used for subsequent IO modelling. Note that removing the mean is a common practice in fitting dynamic IO models to model the changes in an output signal in response to a temporally varying input[52,113]. The mean of the output signal is easily learned by time-averaging the entire signal across all input levels and thus this mean is a single number regardless of input value and does not affect the IO modelling of the output signal variations for different inputs. The dynamic IO model thus describes how brain network activity varies around its mean in response to the variations in stimulation amplitude and frequency over time.

Multi-trial experimental design to dissociate single-trial input-driven dynamics. *Input-driven dynamics and intrinsic dynamics.* The overall brain network dynamics $\mathbf{y}_k$ in equation (1)—that is, the measured LFP power feature time series—can be decomposed into two parts: (1) input-driven dynamics $\mathbf{y}_{k,I}$ and (2) intrinsic dynamics $\mathbf{y}_{k,N}$; that is, $\mathbf{y}_k = \mathbf{y}_{k,I} + \mathbf{y}_{k,N}$ (Fig. 7). The input-driven dynamics $\mathbf{y}_{k,I}$ are driven only by the stimulation input $\mathbf{u}_k$:

$$\begin{cases} \mathbf{x}_{k+1,I} = A\mathbf{x}_{k,I} + B\mathbf{u}_k \\ \mathbf{y}_{k,I} = C\mathbf{x}_{k,I} \end{cases}, \tag{2}$$

where $\mathbf{x}_{k,I}$ is the input-driven part of the latent state. Thus, the same stimulation input would always lead to the same input-driven dynamics. In contrast, the intrinsic dynamics are not dependent on the stimulation input (are input-irrelevant) and thus in the model are only driven by the noise terms $\mathbf{w}_k$ and $\mathbf{v}_k$:

$$\begin{cases} \mathbf{x}_{k+1,N} = A\mathbf{x}_{k,N} + \mathbf{w}_k \\ \mathbf{y}_{k,N} = C\mathbf{x}_{k,N} + \mathbf{v}_k \end{cases}, \tag{3}$$

where $\mathbf{x}_{k,N}$ is the intrinsic part of the latent state. These intrinsic dynamics are not affected by stimulation input and change from trial to trial even with the same stimulation input. Consistently, these intrinsic dynamics are driven by input-irrelevant noise in the model. Note that adding equations (2) and (3) gives equation (1), showing the decomposition.

Dissociating the ground-truth single-trial input-driven dynamics. As the main goal of the IO model is to describe how brain network dynamics respond to stimulation, testing the IO model requires evaluating how well the model can predict the input-driven dynamics $\mathbf{y}_{k,I}$. Thus, we need to dissociate the ground-truth input-driven dynamics from the noisy intrinsic dynamics in the measured overall brain network dynamics; we can then compare the ground-truth input-driven dynamics to model predictions. This dissociation is difficult and requires designing a careful experiment. To achieve this dissociation, we repeated the same stochastic MN input across stimulation trials (Fig. 1b, Supplementary Fig. 5b and Supplementary Notes 6 and 7 contain more details). This experiment design allows us to obtain the ground-truth single-trial input-driven dynamics by averaging the measured overall brain network dynamics across all trials. In particular, as the input is the same in all trials, the single-trial input-driven dynamics $\mathbf{y}_{k,I}$ are also the same for all trials and thus remain intact after averaging (Supplementary Fig. 5d). However, as intended, averaging suppresses the single-trial intrinsic dynamics $\mathbf{y}_{k,N}$ because they change from trial to trial. Thus, averaging dissociates the ground truth of single-trial input-driven dynamics, which can then be used to evaluate the predicted single-trial input-driven dynamics by the IO model within cross-validation as described below.

IO model fitting and evaluation in cross-validation. *Summary.* On the basis of the multi-trial experimental design, we next developed a fourfold cross-validation procedure. Within a cross-validation fold, we took one quarter of the IO data in each trial as the test set (one test set for each trial; Supplementary Fig. 5a,d). We then took the other three quarters of the IO data in each trial as training data and concatenated these single-trial training data across trials to form the training set (Supplementary Fig. 5a,c). We fitted the IO models to the training set without any trial averaging (Supplementary Fig. 5c). We then used the fitted IO model to predict the single-trial input-driven dynamics in a test set from only the history of the stimulation input and without any knowledge of the measured LFP power features in that test set (that is, with forward prediction). This provided a prediction of single-trial input-driven dynamics, which was identical for all test sets (corresponding to different trials) in a given cross-validation fold, because the inputs were identical for all test sets in a given fold (Supplementary Fig. 5d). Finally, we compared the predicted single-trial input-driven dynamics with the ground-truth single-trial input-driven dynamics (ground truth was computed by averaging the measured LFP power features, as explained above). Note that, while for a given cross-validation fold, single-trial input-driven dynamics are the same across trials, these single-trial input-driven dynamics are different for different cross-validation folds and for different experiments. Thus, overall, we can evaluate the IO models for predicting single-trial input-driven dynamics across different stochastic waveforms. Details are given in the next few sections.

Forming test and training sets. We used cross-validation to fit and evaluate the dynamic IO model in equation (1) on the basis of the collected IO dataset $\{\mathbf{u}_k, \mathbf{y}_k\}_{k\in\{1,2,3,\dots,T\}}$ (Supplementary Fig. 5). For each IO dataset, we conducted a fourfold cross-validation. In the jth fold, we took out the same jth quarter of the IO data in each of the N_R stimulation trials as the test sets, resulting in one test set for each trial. We denote the test data in the hth trial by $\mathcal{D}_h^{(\text{test})}$ and the rest of the data in that trial by $\mathcal{D}_h^{(\text{train})}$ as training data. In the jth fold, we concatenated all the single-trial training data to form one single training set as $\mathcal{D}^{(\text{train})} = \bigcup_{h=1}^{N_R} \mathcal{D}_h^{(\text{train})}$ (Supplementary Fig. 5c). We also formed N_R test sets, one per trial (for example $\mathcal{D}_h^{(\text{test})}$ for trial h; Supplementary Fig. 5d). Thus, each test set had a length of $L/4$, where L is the total number of time steps within one trial (Supplementary Note 8) and each training set had a length of $N_R \times 3L/4$. The above procedure ensured that neither the input nor the output in the test sets were seen by the training set. Further, as the input was an MN stochastic process, it was independent in the test sets and the training set (Supplementary Note 9 and Supplementary Fig. 5b).

Forward prediction to predict the single-trial input-driven dynamics in the test sets. Input-driven dynamics describe the effect of stimulation input on LFP power features. Thus, the main evaluation of an IO model is how well it can predict the input-driven dynamics in the test sets. This evaluation can be performed using forward prediction, which proceeded as follows.

Within each cross-validation fold, in each test set corresponding to each trial, we formed a single-trial forward predictor $\hat{\mathbf{y}}_k$ based on equation (1) as:

$$\begin{cases} \mathbf{s}_{k+1} = A\mathbf{s}_k + B\mathbf{u}_k \\ \hat{\mathbf{y}}_k = C\mathbf{s}_k \end{cases}, \tag{4}$$

where $\mathbf{s}_k \in \mathbb{R}^{N_x \times 1}$ is the forward prediction of the latent state. We can see from equation (2) that the above equation (4) provides a prediction of the single-trial input-driven dynamics. By recursively evaluating equation (4), we can write $\mathbf{s}_k$ as a function of the past inputs $\{\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-1}\}$ and the initial state $\mathbf{s}_0$:

$$\mathbf{s}_k = B\mathbf{u}_{k-1} + AB\mathbf{u}_{k-2} + A^2B\mathbf{u}_{k-3} + \dots + A^{k-1}B\mathbf{u}_0 + A^k\mathbf{s}_0, \tag{5}$$

Here the terms $A^{j-1}B\mathbf{u}_{k-j}$ in the sum are evaluated from $j = 1$ to k. This can be seen through expanding the recursion in equation (4) by recursively replacing the equation for $\mathbf{s}_i$ from time $i = 0, \dots, k$. Accordingly, we can see that the predicted single-trial input-driven dynamics $\hat{\mathbf{y}}_k = C\mathbf{s}_k$ in forward prediction is calculated from only the past stimulation inputs $\{\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-1}\}$ and the

initial state $\mathbf{s}_0$, and without using the measured past output LFP power features $\{\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2 \cdots, \mathbf{y}_{k-1}\}$:

$$\hat{\mathbf{y}}_k = CB\mathbf{u}_{k-1} + CAB\mathbf{u}_{k-2} + CA^2B\mathbf{u}_{k-3} + \ldots + CA^{k-1}B\mathbf{u}_0 + CA^k\mathbf{s}_0. \quad (6)$$

Equation (6) again shows that the forward prediction of LFP power feature $\hat{\mathbf{y}}_k$ is assessing the input-driven part of the measured LFP power features—that is, $\hat{\mathbf{y}}_k$ is a function of only past inputs. To consider the most challenging case, we assumed we did not have any prior information on the initial state and thus set it at zero $\mathbf{s}_0 = \mathbf{0}_{N_x \times 1}$ in all forward predictions in all test sets. With this zero initialization and if zero stimulation input is given to the forward prediction, the forward predicted LFP power features remain zero over time as they should by design; this is because forward prediction aims to assess the purely input-driven dynamics of the LFP power features in response to stimulation input (that is predict $\mathbf{y}_{k,I}$ in equation (2)), and so it should give zero when there is no stimulation input. We emphasize that this does not imply that the IO model predicts zero for the LFP power features $\mathbf{y}_k$, as we show with one-step-ahead prediction using the fitted IO models (see Methods, 'One-step-ahead prediction of single-trial overall brain network dynamics during stimulation' and Supplementary Note 6).

Model evaluation. We first fitted the IO model parameters $\theta = \{A, B, C, Q\}$ in the training set with single-trial IO data without any averaging as detailed in 'Model fitting in the training set' section below (Supplementary Fig. 5c). Here, A, B and C are the LSSM matrices in equation (1) and Q is the covariance matrix of the state noise $\mathbf{w}_k$ and the observation noise $\mathbf{v}_k$ in equation (1).

Using the fitted IO models, in each test set that corresponds to one trial, we then predicted the single-trial input-driven dynamics using forward prediction with equation (6)—that is, we computed $\{\hat{\mathbf{y}}_j\}, j = 1, 2, \ldots \frac{L}{4}$, for each test set. From equation (6), each test set had the same predicted single-trial input-driven dynamics because the input stimulation was repeated in each test set by design (Supplementary Fig. 5d).

Finally, we averaged the measured output LFP power features across all test sets to obtain the ground-truth single-trial input-driven dynamics as denoted by $\left\{\bar{\mathbf{y}}_j\right\}, j = 1, 2, \ldots \frac{L}{4}$ as described in 'Dissociating the ground-truth single-trial input-driven dynamics' above. We computed the CC between the ground-truth single-trial input-driven dynamics $\left\{\bar{\mathbf{y}}_j\right\}$ and the predicted single-trial input-driven dynamics $\{\hat{\mathbf{y}}_j\}$:

$$\mathrm{CC}^{\text{per fold}} = \frac{\mathrm{Cov}\left(\{\hat{\mathbf{y}}_j\}, \{\bar{\mathbf{y}}_j\}\right)}{\sqrt{\mathrm{Var}\left(\{\hat{\mathbf{y}}_j\}\right) \times \mathrm{Var}\left(\{\bar{\mathbf{y}}_j\}\right)}}, \quad (7)$$

where $\mathrm{Cov}(\cdot)$ and $\mathrm{Var}(\cdot)$ represent the empirical covariance and variance of a time series, respectively. Note that the CC is the same for all test sets within one cross-validation fold so the CC in one cross-validation fold is taken as the CC for any one of its associated test sets (Supplementary Fig. 5d). However, the CC can be different for different cross-validation folds, as different cross-validation folds have independent input in their test sets (Supplementary Fig. 5b). We then define the IO prediction accuracy as the average CC over all 4 cross-validation folds:

$$\mathrm{CC} = \frac{1}{4}\sum_{m=1}^{4} \mathrm{CC}^{(m)}, \quad (8)$$

where $\mathrm{CC}^{(m)}$ represents the CC in fold m as in equation (7). A higher cross-validated CC represents better prediction and thus a higher IO prediction accuracy. Note that IO prediction accuracy is computed from the dynamic IO model.

We used CC as our main measure because it has been widely used in neuroscience, for example, in quantifying the neural response after stimulation ends[40,42], or the neural decoding accuracy of behaviour[43,58,109,114–116]. Nevertheless, to show the robustness of our results to the performance measure, we also computed the EV of our dynamic IO models in forward prediction. EV quantifies how much variance in the ground-truth single-trial input-driven dynamics can be explained by the predicted single-trial input-driven dynamics from our dynamic IO models. EV is given by:

$$\mathrm{EV}^{\text{per fold}} = \left(1 - \frac{\frac{4}{L}\sum_{k=1}^{L/4}\left(\bar{\mathbf{y}}_k - \hat{\mathbf{y}}_k\right)^2}{\mathrm{Var}\left(\bar{\mathbf{y}}_k\right)}\right) \times 100\%, \quad (9)$$

We thus also computed the average EV over all four cross-validation folds for forward prediction of our dynamic IO models:

$$\mathrm{EV} = \frac{1}{4}\sum_{m=1}^{4} \mathrm{EV}^{(m)}, \quad (10)$$

where $\mathrm{EV}^{(m)}$ represents the EV in fold m as in equation (9).

Assessing generalizability of the IO model across different stochastic waveforms. While within a given cross-validation fold, the input and thus the single-trial input-driven dynamics were the same across test sets, for different cross-validation folds or for different stimulation experiments, we used different stochastic inputs (Supplementary Fig. 5b). In addition, no averaging was ever done across cross-validation folds or experiments. So overall, we are evaluating the single-trial forward prediction across different stochastic waveforms. This allows us to evaluate the generalizability of the IO model.

Model fitting in the training set. We first describe how we fitted the model parameters θ within a cross-validation fold given a known hyperparameter N_x and then present how the optimal N_x was selected. The fitting of all model parameters and the selection of the hyperparameter was based only on the training set, without seeing the test sets whose inputs were fully independent of the training set (Supplementary Fig. 5b and Supplementary Note 9). First, given a fixed N_x, we used the prediction error method[52] to fit the model parameters θ. In the prediction error method, we minimize the following cost function for single-trial training data obtained from the N_R trials without any averaging:

$$J(\theta) = \sum_{h=1}^{N_R} \sum_{\{\mathbf{y}_k, \mathbf{u}_k\} \in \mathcal{D}_h^{(\text{train})}} \left(\mathbf{y}_k - \hat{\mathbf{y}}_k(\theta)\right)^2, \quad (11)$$

where $\hat{\mathbf{y}}_k(\theta)$ is the forward predictor in equation (6) and is expressed as a function of the model parameters θ. We then obtain the fitted model parameters $\hat{\theta}$ via standard nonlinear optimization methods[52]:

$$\hat{\theta} = \arg\min_{\theta} J(\theta), \quad (12)$$

where the initial guess of θ is obtained by a fast, projection-based subspace method for fitting LSSMs[52]. As the data in equation (11) are single-trial data and not averaged at all, the model parameters are fitted to single-trial IO data and capture single-trial brain network dynamics. Second, within the training set, we used an inner cross-validation to select the optimal hyperparameter N_x from the grid {1,2,3,4,5,6} (this range was sufficient as described below). For each N_x, we conducted an inner-level cross-validation within the training set to calculate the IO prediction accuracy for that N_x. Then we compared the inner cross-validated IO prediction accuracy within the training set and found the N_x that maximized it as the optimal hyperparameter. In fact, the selected optimal N_x was 2.74 ± 0.08, less than the upper bound of 6, confirming the adequacy of the used upper bound for N_x.

Statistical tests. We conducted two statistical tests to assess the significance of forward prediction as quantified by the cross-validated IO prediction accuracy in equation (8). First, we quantified what the IO prediction accuracy would have been by random chance if we had used a randomly generated input in our modelling that was different and independent from the actual input (that is, independent from the actual time series of amplitude and frequency). To quantify this chance level, we constructed an input-baseline distribution (Supplementary Fig. 6a) for the IO prediction accuracy by repeating the same cross-validation procedure on artificially generated IO datasets. We created the artificially generated IO datasets by (1) replacing the input with random MN inputs that were independent of the actual MN inputs delivered in the IO datasets, and (2) keeping the output the same as the actual output in the IO datasets. We generated 100 independent artificially generated IO datasets and repeated the same cross-validation procedure and computed the IO prediction accuracy for each. The IO prediction accuracies obtained across all these 100 datasets was used to obtain the input-baseline distribution of IO prediction accuracy. We define the input-baseline P value as the probability that IO prediction accuracy from the input-baseline distribution is greater than the IO prediction accuracy from the actual IO dataset. This P value is calculated by fitting a parametric GPD to the tail of the input-baseline distribution (Supplementary Note 10). We calculated the input-baseline P values for each of the power features and corrected the P value using FDR to account for multiple comparisons[117] across the total number of LFP power features being modelled within an IO dataset.

Second, we quantified what the IO prediction accuracy would have been by random chance if instead of the actual brain network activity during stimulation, we had used the pre-stimulation at-rest brain network activity in our IO modelling. To quantify this chance level, we constructed an output-baseline distribution for the IO prediction accuracy (Supplementary Fig. 6b) by repeating the same cross-validation procedure with artificially generated IO datasets that had the same input as the actual IO dataset, but had their outputs replaced as follows. We replaced each second of the recorded during-stimulation raw signal in the IO dataset with a random 1 s period of the pre-stimulation at-rest raw signal. This random 1 s period was chosen by randomly selecting a starting time in the pre-stimulation data. To control for the effect of the stimulation-artefact rejection algorithm on IO modelling, we applied the same stimulation-artefact rejection algorithm to this replaced signal as well. To do this, we first obtained the instantaneous artefact-spike timings and artefact-tail windows from the

actual during-stimulation raw signal; then, using these timings and windows, we repeated the same stimulation-artefact-rejection procedure on the replaced raw signal (details in Supplementary Note 1). Then for the resulting signal, we repeated the same LFP signal processing and power calculation to obtain the output LFP power features. We next repeated the same cross-validation to obtain the IO prediction accuracy and finally repeated the above procedure 100 times to form the output-baseline distribution for the IO prediction accuracy. Finally, we used the same procedure as was used in the input-baseline test to obtain the output-baseline P value. Note that the IO prediction accuracy from the input-baseline and output-baseline tests would be centred around zero, since the input and output were independent in these tests.

We declare that forward prediction is significant if for its IO prediction accuracy, both the FDR-corrected input-baseline P value and output-baseline P value are less than 0.05. A power feature with significant prediction is termed a predictable power feature. Unless otherwise specified, all paired comparisons use the two-sided Wilcoxon signed-rank test and all the non-paired multi-group comparisons use the Kruskal–Wallis test, and significance is declared if the P value is less than 0.05.

One-step-ahead prediction of single-trial overall brain network dynamics during stimulation. Beyond evaluating our IO model in predicting single-trial input-driven dynamics as the main performance measure (using forward prediction), we also evaluated our IO model in predicting the single-trial overall brain network dynamics. As seen from equations (1)–(3), the overall brain network dynamics $\mathbf{y}_k$ consist not only of the input-driven dynamics $\mathbf{y}_{k,I}$ that remain the same from trial to trial, but also of the intrinsic dynamics $\mathbf{y}_{k,N}$ that can change from trial to trial. Thus, to predict the single-trial overall brain network dynamics, in addition to the stimulation input, the measured LFP power features in that single-trial should also be used for prediction. Therefore, instead of forward prediction, we used a different one-step-ahead prediction that predicts the current single-trial LFP power feature $\mathbf{y}_k$ as a function of both the past inputs $\{\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-1}\}$ and the past measured LFP power features $\{\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2 \dots, \mathbf{y}_{k-1}\}$ (paragraphs below). We denote this one-step-ahead prediction by $\tilde{\mathbf{y}}_k$.

In brief, since our IO models were fitted directly using single-trial training data without averaging (Supplementary Fig. 5c), the fitted IO models captured single-trial intrinsic dynamics with the noise terms $\mathbf{w}_k$ and $\mathbf{v}_k$ and could thus directly predict single-trial overall brain network dynamics. The optimal one-step-ahead prediction (to minimize mean-squared error) can be done by the following Kalman recursive form[52]:

$$\begin{cases} \mathbf{z}_k = A\mathbf{z}_{k-1} + B\mathbf{u}_{k-1} + K(\mathbf{y}_{k-1} - C\mathbf{z}_{k-1}) \\ \tilde{\mathbf{y}}_k = C\mathbf{z}_k \end{cases}, \tag{13}$$

where $\mathbf{z}_k$ is the one-step-ahead prediction of the latent state $\mathbf{x}_k$ and K is the total Kalman gain from $\mathbf{z}_{k-1}$ to $\mathbf{z}_k$ computed from the fitted A, C and the fitted noise covariances of $\mathbf{w}_k$ and $\mathbf{v}_k$ (refs. [52,113]). Accordingly, the one-step-ahead-predicted LFP power feature is computed as a function of both the past stimulation inputs and past measured LFP power features in a given trial and can be expanded as (assuming no prior knowledge on initial value and thus using zero initial condition to have the most challenging case):

$$\begin{aligned} \tilde{\mathbf{y}}_k &= CB\mathbf{u}_{k-1} + C(A-KC)B\mathbf{u}_{k-2} + C(A-KC)^2B\mathbf{u}_{k-3} \\ &+ \dots + C(A-KC)^{k-1}B\mathbf{u}_0 + CK\mathbf{y}_{k-1} + C(A-KC)K\mathbf{y}_{k-2} \\ &+ C(A-KC)^2K\mathbf{y}_{k-3} + \dots + C(A-KC)^{k-1}K\mathbf{y}_0 \end{aligned} \tag{14}$$

Here the terms $C(A-KC)^{j-1}B\mathbf{u}_{k-j}$ in the sum are evaluated from $j=1$ to k and similarly for terms associated with $\mathbf{y}_{k-j}$. This can be seen through expanding equation (13) by recursively replacing the equation for $\mathbf{z}_i$ for $i = 0, \dots, k$.

We used the above equation to predict single-trial overall dynamics of LFP power features during stimulation with a similar fourfold cross-validation procedure. The training procedure was exactly the same as in equations (11) and (12). As we are now evaluating single-trial overall brain network dynamics, in contrast to forward prediction, the ground truth in this case is simply the measured LFP power feature values in each test set (and does not require averaging to dissociate). Thus, the ground truth is different in different test sets of a given cross-validation fold. In each test set, we now directly compared the single-trial one-step-ahead prediction with the single-trial LFP power features $\mathbf{y}_k$ measured in that test set and computed their CC (Supplementary Note 6). Note that CCs can now be different for different test sets and thus the CC of the given cross-validation fold is computed as the average of the CC across its test sets. In these analyses no averaging is ever required even in obtaining the ground truth.

The significance of the cross-validated CC in one-step-ahead prediction is evaluated through an output-permutation test in which the single-trial LFP power feature output data were randomly permuted in time for 100 times and the same cross-validation procedure was repeated on this permuted output data. The resulting cross-validated CC of these permuted output data formed the output-permutation distribution. We define the output-permutation P value as the probability that CC from the output-permutation distribution is greater than the CC from the actual IO dataset.

Regression model, smoothing model and non-oscillatory model. The regression model is a non-dynamic feedthrough model written as

$$\mathbf{y}_k = B\mathbf{u}_k + \mathbf{e}_k, \tag{15}$$

where $\mathbf{e}_k \in \mathbb{R}$ is modelled as a zero-mean Gaussian noise with variance $E \in \mathbb{R}$ and $\theta = \{B, E\}$ are model parameters to be fitted. Note that since the output LFP power features were calculated using an overlapping window of $10T_s$, we additionally applied a moving average to the input stimulation parameters using the same overlap to form the input $\mathbf{u}_k$ in equation (15). With this processing, the regression model represents a non-dynamic feedthrough from the stimulation input to the output LFP power features at the same time steps. The forward predictor associated with this model is

$$\mathbf{y}_k = B\mathbf{u}_k. \tag{16}$$

The smoothing model is a special case of LSSM with an identity state transition matrix:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{x}_k + B\mathbf{u}_k + \mathbf{w}_k \\ \mathbf{y}_k = C\mathbf{x}_k + \mathbf{v}_k \end{cases}. \tag{17}$$

The eigenvalues of the state transition matrix A in LSSM in equation (1) characterize the dynamics of the effect of stimulation on the output power features[52]—that is, how stimulation affects the time evolution of power features. This can be seen from equation (6), where the relationship between the current LFP power feature and past stimulation inputs are dictated by A and its powers. An identity state transition matrix with all eigenvalues equal to 1 represents a special form of dynamic response in which the effect of stimulation input at each time is simply an unweighted smoothed average of its past values[52]. The forward predictor in this case is in the same form as equation (6), replacing the state transition matrix with an identity matrix.

The non-oscillatory model is another special case of LSSM in which the state transition matrix A in LSSM in equation (1) is constrained to have positive real eigenvalues (Supplementary Note 13). Since complex conjugate eigenvalue pairs and negative real eigenvalues represent oscillatory dynamics (see also Supplementary Note 13), this non-oscillatory model structure ensures that the fitted LSSM does not model any oscillatory dynamics (Supplementary Note 13). The forward predictor in this case is in the same form as equation (6).

At-rest functional controllability. To calculate at-rest functional controllability (Supplementary Note 17), we first calculated the at-rest LFP power features using at-rest data with the same time step of 1 s (Supplementary Note 16). We denote the four at-rest LFP power features of the stimulation site by $\mathbf{r}_k \in \mathbb{R}^{4\times 1}$, which constitute the stimulation node. We denote the LFP power features at the other sites by $\mathbf{p}_k \in \mathbb{R}^{N_y\times 1}$, which constitute the other network nodes. We concatenated $\mathbf{p}_k$ and $\mathbf{r}_k$ into a single vector $\mathbf{z}_k = \begin{bmatrix} \mathbf{r}_k \\ \mathbf{p}_k \end{bmatrix} \in \mathbb{R}^{(N_y+4)\times 1}$, $k = 1, 2, \dots, T_0$ (T_0=500 is the total number of time steps in the at-rest data (Supplementary Note 16)).

Then, we built a deterministic dynamic network model for $\{\mathbf{z}_k\}$ as

$$\mathbf{z}_k = T\mathbf{z}_{k-1} + Dg_k, \tag{18}$$

where g_k is a nominal scalar variable representing the input strength (whose value does not enter in calculating at-rest controllability[46]) and $D = [1, 1, 1, 1, 0, 0, \dots, 0]'$ is a nominal input matrix representing that the input g_k is delivered at the stimulation node. The first four components in D are 1 because the stimulation input is injected into the stimulation node (four power features at the stimulation site) and the rest of its components are 0 because no input is applied to the other network nodes[46,47]. $T \in \mathbb{R}^{(N_y+4)\times(N_y+4)}$ is the state transition matrix and is required to be stable (all eigenvalues within the unit disc) to calculate controllability[46] (Supplementary Note 17). We thus fitted a stable T from a ridge-regularized multivariable linear regression of $\mathbf{z}_k$ against its immediate past value $\mathbf{z}_{k-1}$ (with g_k=0 because we are using only at-rest data to fit T)—that is, $T = Z_cZ_p'\left(Z_pZ_p' + \lambda I\right)^{-1}$ with $Z_c = [\mathbf{z}_{T_0}, \mathbf{z}_{T_0-1}, \dots, \mathbf{z}_2] \in \mathbb{R}^{(N_y+4)\times T_0}$, $Z_p = [\mathbf{z}_{T_0-1}, \mathbf{z}_{T_0-2}, \dots, \mathbf{z}_1] \in \mathbb{R}^{(N_y+4)\times T_0}$—and the regularization parameter λ was initially chosen as λ_0=0.000001 and gradually increased to $10\lambda_0$, $100\lambda_0$, …, until a stable T was fitted.

Next, on the basis of equation (18), we calculated the infinite-horizon controllability Gramian W_c from T and D as the solution of the following discrete-time Lyapunov equation[46]:

$$TW_cT' - W_c + DD' = 0. \tag{19}$$

It can be shown that the ith diagonal element of W_c is inversely proportional to the energy that is needed to move the ith component in $\mathbf{z}_k$ around the state space (Supplementary Note 17). Therefore, we took the logarithm of the $(i+4)$th diagonal elements of W_c as the functional controllability from the stimulation node

to the ith network node (note that the first 4 elements represent the stimulation node and thus their controllability is not relevant to compute):

$$O^{(i)} = \log(W_c(i+4, i+4)), i = 1, 2, \dots, N_y, \quad (20)$$

where $W_c(i, j)$ represents the i, jth element in the matrix W_c.

To obtain a chance-level baseline distribution for at-rest functional controllability, we randomly shuffled the time index of $\mathbf{z}_k$ 1,000 times and repeated the computations in equations (18)–(20) each time (fitting the model in equation (18) from scratch each time) to get 1,000 random at-rest functional controllability values and then found their empirical distribution. We define the P value of an actual at-rest functional controllability as the probability that at-rest functional controllability from the baseline distribution would be larger than the actual at-rest functional controllability. This P value was calculated by GPD fitting (Supplementary Note 10) and corrected using FDR across all LFP power features being modelled in one IO dataset. We calculated the z-scored at-rest functional controllability by z-scoring the actual at-rest functional controllability.

Prediction of IO prediction accuracy and estimated response strength using at-rest functional controllability. We conducted three analyses to evaluate the relation of IO prediction accuracy and at-rest functional controllability.

First, we test if the value of at-rest functional controllability within a brain region correlates with and can predict the value of the IO prediction accuracy within that region during stimulation. We categorized the predictable power features according to their anatomical regions and included regions that at least had 2 predictable power features (total number of regions $N = 13$). We computed the average at-rest functional controllability of the predictable power features within a brain region as $O^{(j)}_{\text{region}}$ (j representing the jth region) and the corresponding average IO prediction accuracy as $CC^{(j)}_{\text{region}}$. We first performed a linear regression to test if $CC^{(j)}_{\text{region}}$ correlated with $O^{(j)}_{\text{region}}$. Then, we tested if $O^{(j)}_{\text{region}}$ predicted $CC^{(j)}_{\text{region}}$ by constructing a linear predictor using $O^{(j)}_{\text{region}}$:

$$CC^{(j)}_{\text{region}} = a + bO^{(j)}_{\text{region}}. \quad (21)$$

We used leave-one-region-out cross-validation to evaluate the prediction performance using this linear predictor; in the jth fold, we took the average at-rest functional controllability and IO prediction accuracy of all regions except the jth region as training set to fit the parameters $\hat{a}, \hat{b}$ in the regression model in equation (21). We then used this model to perform prediction for the jth region as $\widehat{CC}^{(j)}_{\text{region}} = \hat{a} + \hat{b}O^{(j)}_{\text{region}}$. The final prediction error across all 13 cross-validation folds (which equals the total number of tested regions because we did leave-one-region-out cross-validation) is the mean square error (MSE) computed as $\text{MSE} = \frac{1}{13}\sum_{j=1}^{13}\left(\widehat{CC}^{(j)}_{\text{region}} - CC^{(j)}_{\text{region}}\right)^2$. We evaluated the significance of this predictor by comparing its MSE with the chance level obtained by randomly shuffling the indices of $O^{(j)}_{\text{region}}$ for 1,000 times and calculating the corresponding MSEs by performing the leave-one-region-out cross-validation from scratch. We define the P value as the probability that MSE from this chance-level distribution would be larger than the actual MSE. We refer to this test as the permutation test. This P value was calculated by GPD fitting (Supplementary Note 10).

Second, at the single network node level (that is, single LFP power feature), we test if the value of at-rest functional controllability of a predictable power feature correlates with and can predict the value of its IO prediction accuracy. This analysis is the same as the first analysis except that we replace the average at-rest functional controllability within regions $O^{(j)}_{\text{region}}$ with the at-rest functional controllability of each predictable power feature $O^{(i)}$, and similarly for the IO prediction accuracy.

Third, among all LFP power features (regardless of predictable or not), we test if the value of at-rest functional controllability correlates with the IO prediction accuracy with the same method as the one above. Further, among all LFP power features, we test if the value of at-rest functional controllability can predict whether an LFP power feature will have a predictable response or not. We constructed a threshold-based classifier to do this prediction:

$$\begin{cases} \mathcal{I}^{(i)} = 1, \text{ if } O^{(i)} > \sigma \\ \mathcal{I}^{(i)} = 0, \text{ if } O^{(i)} \leq \sigma \end{cases}, \quad (22)$$

where $\mathcal{I}^{(i)} = 1$ represents a predictable ith LFP power feature and 0 otherwise. We swept the threshold σ to obtain the ROC curve of this classifier across all LFP power features and calculated its AUC. We evaluate the significance of this classifier by comparing its AUC with chance level obtained by random shuffling of the indices of $O^{(i)}$ for 1,000 times and calculating the corresponding AUCs. We define the P value as the probability that AUC from this chance-level distribution would be larger than the actual AUC, and similar to the first analysis above refer to this test as the permutation test. This P value was calculated by GPD fitting (Supplementary Note 10).

Finally, we further studied if at-rest functional controllability can also predict the response strength of the network nodes. To estimate the response strength of a network node (that is, one LFP power feature), we used our fitted IO model to predict its single-trial input-driven dynamics $\hat{\mathbf{y}}_k$ as in equation (6). We then computed the response strength $\mathcal{S}$ as the logarithm of the average energy in the temporal variations of the predicted single-trial input-driven dynamics:

$$\mathcal{S} = \log\left(\frac{4}{L}\sum_{k=1}^{\frac{L}{4}}\left(\hat{\mathbf{y}}_k\right)^2\right), \quad (23)$$

where $\frac{L}{4}$ is the length of the test set. This is indeed the estimate of the logarithm of the average energy in the temporal changes of the LFP power features due to stimulation. We then averaged the response strength $\mathcal{S}$ in equation (23) across the four cross-validation folds to obtain the estimated response strength. We next repeated our region and single-node correlation analyses by replacing the IO prediction accuracy of each power feature with its cross-validated response strength.

NLARX model. The NLARX model structure[118] models the output LFP power feature at each time k as

$$\mathbf{y}_k = \boldsymbol{\phi}'\mathbf{R}_{N_a,N_b} + \sum_{m=1}^{M}\alpha_m\beta_m^{-\frac{(N_a+N_bN_u)}{2}}\psi\left(\frac{\mathbf{R}_{N_a,N_b} - \boldsymbol{\gamma}_m}{\beta_m}\right), \quad (24)$$

where $\mathbf{R}_{N_a,N_b} = \left[\mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \dots, \mathbf{y}_{k-N_a}, \mathbf{u}'_{k-1}, \mathbf{u}'_{k-2}, \dots, \mathbf{u}'_{k-N_b}\right]' \in \mathbb{R}^{(N_a+N_bN_u)\times 1}$ is the regressor vector containing past output LFP power features and input stimulation amplitude and frequency, $\psi : \mathbb{R}^{(N_a+N_bN_u)\times 1} \to \mathbb{R}$ is a radial wavelet function taken as $\psi(\mathbf{x}) = (N_a + N_bN_u - \mathbf{x}'\mathbf{x})e^{-\frac{\mathbf{x}'\mathbf{x}}{2}}, \mathbf{x} \in \mathbb{R}^{(N_a+N_bN_u)\times 1}$ and M is the number of wavelets. The model parameters are the linear regression coefficients $\boldsymbol{\phi} \in \mathbb{R}^{(N_a+N_bN_u)\times 1}$, the wavelet weights $\alpha_m \in \mathbb{R}$, $m = 1, 2, \dots, M$, the nonlinear wavelet dilation parameters $\beta_m \in \mathbb{R}$, $m = 1, 2, \dots, M$ and the nonlinear wavelet translation parameters $\boldsymbol{\gamma}_m \in \mathbb{R}^{(N_a+N_bN_u)\times 1}$, $m = 1, 2, \dots, M$. We collect all the above parameters as $\theta = \{\boldsymbol{\phi}, \alpha_m, \beta_m, \boldsymbol{\gamma}_m, m = 1, 2, \dots, M\}$. The NLARX model in equation (24) has three hyperparameters N_a, N_b and M, which represent the delay of the output included in the regressor, the delay of the input included in the regressor and the number of wavelets, respectively. The number of free parameters in this NLARX model is $(N_a + N_bN_u) + M(2 + N_a + N_bN_u)$ (ref. [118]). The forward predictor associated with this model is

$$\hat{\mathbf{y}}_k = \boldsymbol{\phi}'\hat{\mathbf{R}}_{N_a,N_b} + \sum_{m=1}^{M}\alpha_m\beta_m^{-\frac{(N_a+N_bN_u)}{2}}\psi\left(\frac{\hat{\mathbf{R}}_{N_a,N_b} - \boldsymbol{\gamma}_m}{\beta_m}\right), \quad (25)$$

$$\hat{\mathbf{R}}_{N_a,N_b} = \left[\hat{\mathbf{y}}_{k-1}, \hat{\mathbf{y}}_{k-2}, \dots, \hat{\mathbf{y}}_{k-N_a}, \mathbf{u}'_{k-1}, \mathbf{u}'_{k-2}, \dots, \mathbf{u}'_{k-N_b}\right]', \quad (26)$$

with zero initial conditions—that is, $\left[\hat{\mathbf{y}}_0, \hat{\mathbf{y}}_{-1}, \dots, \hat{\mathbf{y}}_{-N_a+1}, \mathbf{u}'_0, \mathbf{u}'_{-1}, \dots, \mathbf{u}'_{-N_b+1}\right]' = \mathbf{0}_{(N_a+N_bN_u)\times 1}$. This predictor again only uses the values of the past stimulation inputs ($\mathbf{u}_k$) for forward prediction of the single-trial input-driven dynamics.

We repeated the same cross-validation procedure using the NLARX model structure, except that (1) in the training set, we used a specialized fast algorithm[118] to fit the model parameters θ; (2) the hyperparameters N_a and N_b were selected on the basis of an inner cross-validation within the training set with grid $N_a \in \{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$, $N_b \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$, where the value 0 for N_a represents not including the past values of $\mathbf{y}_k$ in the model of the present $\mathbf{y}_k$. The upper bounds of these grids were chosen such that the data length is larger than the number of parameters even when the upper bound number of parameters is used in any IO dataset; (3) in each inner cross-validation within the training set, the number of wavelets M was chosen on the basis of minimizing the Akaike information criterion[119].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. The raw and analysed datasets generated during the study are too large to be publicly shared, but are available for research purposes from the corresponding author on reasonable request.

Code availability

The custom computer code in this study is available at https://github.com/ShanechiLab/DynamicStimulation.

Received: 25 June 2019; Accepted: 24 November 2020;
Published online: 01 February 2021

References

1. Shanechi, M. M. Brain–machine interfaces from motor to mood. *Nat. Neurosci.* **22**, 1554–1564 (2019).

2. Hoang, K. B., Cassar, I. R., Grill, W. M. & Turner, D. A. Biomarkers and stimulation algorithms for adaptive brain stimulation. *Front. Neurosci.* **11**, 564 (2017).
3. Lo, M. C. & Widge, A. S. Closed-loop neuromodulation systems: next-generation treatments for psychiatric illness. *Int. Rev. Psychiatry* **29**, 191–204 (2017).
4. Ashkan, K., Rogers, P., Bergman, H. & Ughratdar, I. Insights into the mechanisms of deep brain stimulation. *Nat. Rev. Neurol.* **13**, 548–554 (2017).
5. Deuschl, G. & Agid, Y. Subthalamic neurostimulation for Parkinson's disease with early fluctuations: balancing the risks and benefits. *Lancet Neurol.* **12**, 1025–1034 (2013).
6. Fisher, R. et al. Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy. *Epilepsia* **51**, 899–908 (2010).
7. Boccard, S. G., Pereira, E. A. & Aziz, T. Z. Deep brain stimulation for chronic pain. *J. Clin. Neurosci.* **22**, 1537–1543 (2015).
8. Dandekar, M., Fenoy, A., Carvalho, A., Soares, J. & Quevedo, J. Deep brain stimulation for treatment-resistant depression: an integrative review of preclinical and clinical findings and translational implications. *Mol. Psychiatry* **23**, 1094 (2018).
9. Koning, P. P., de, Figee, M., Munckhof, P., van den, Schuurman, P. R. & Denys, D. Current status of deep brain stimulation for obsessive–compulsive disorder: a clinical review of different targets. *Curr. Psychiatry Rep.* **13**, 274–282 (2011).
10. Williams, Z. M. & Eskandar, E. N. Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat. Neurosci.* **9**, 562 (2006).
11. Chang, E. F., Kurteff, G. & Wilson, S. M. Selective interference with syntactic encoding during sentence production by direct electrocortical stimulation of the inferior frontal gyrus. *J. Cogn. Neurosci.* **30**, 411–420 (2018).
12. Whitmire, C. J., Millard, D. C. & Stanley, G. B. Thalamic state control of cortical paired-pulse dynamics. *J. Neurophysiol.* **117**, 163–177 (2016).
13. Rao, V. R. et al. Direct electrical stimulation of lateral orbitofrontal cortex acutely improves mood in individuals with symptoms of depression. *Curr. Biol.* **28**, 3893–3902 (2018).
14. Hartevelt, T. Jvan et al. Neural plasticity in human brain connectivity: the effects of long term deep brain stimulation of the subthalamic nucleus in Parkinson's disease. *PLoS ONE* **9**, e86496 (2014).
15. Saenger, V. M. et al. Uncovering the underlying mechanisms and whole-brain dynamics of deep brain stimulation for Parkinson's disease. *Sci. Rep.* **7**, 9882 (2017).
16. Basu, I. et al. Consistent linear and non-linear responses to invasive electrical brain stimulation across individuals and primate species with implanted electrodes. *Brain Stimul.* **12**, 877–892 (2019).
17. Crowther, L. J. et al. A quantitative method for evaluating cortical responses to electrical stimulation. *J. Neurosci. Methods* **311**, 67–75 (2019).
18. Yang, Y., Connolly, A. T. & Shanechi, M. M. A control-theoretic system identification framework and a real-time closed-loop clinical simulation testbed for electrical brain stimulation. *J. Neural Eng.* **15**, 066007 (2018).
19. Osorio, I. et al. An introduction to contingent (closed-loop) brain electrical stimulation for seizure blockage, to ultra-short-term clinical trials, and to multidimensional statistical analysis of therapeutic efficacy. *J. Clin. Neurophysiol.* **18**, 533–544 (2001).
20. Little, S. et al. Bilateral adaptive deep brain stimulation is effective in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **87**, 717–721 (2016).
21. Shirvalkar, P., Veuthey, T. L., Dawes, H. E. & Chang, E. F. Closed-loop deep brain stimulation for refractory chronic pain. *Front. Comput. Neurosci.* **12**, 18 (2018).
22. Sani, O. G. et al. Mood variations decoded from multi-site intracranial human brain activity. *Nat. Biotechnol.* **36**, 954–961 (2018).
23. Kirkby, L. A. et al. An amygdala–hippocampus subnetwork that encodes variation in human mood. *Cell* **175**, 1688–1700 (2018).
24. Etkin, A. & Wager, T. D. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* **164**, 1476–1488 (2007).
25. Kupfer, D. J., Frank, E. & Phillips, M. L. Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *Lancet* **379**, 1045–1055 (2012).
26. Williams, L. M. Defining biotypes for depression and anxiety based on large-scale circuit dysfunction: a theoretical review of the evidence and future directions for clinical translation. *Depress. Anxiety* **34**, 9–24 (2017).
27. Montgomery, E. B. & Baker, K. B. Mechanisms of deep brain stimulation and future technical developments. *Neurol. Res.* **22**, 259–266 (2000).
28. Rubin, J. E. & Terman, D. High frequency stimulation of the subthalamic nucleus eliminates pathological thalamic rhythmicity in a computational model. *J. Comput. Neurosci.* **16**, 211–235 (2004).
29. McIntyre, C. C. & Hahn, P. J. Network perspectives on the mechanisms of deep brain stimulation. *Neurobiol. Dis.* **38**, 329–337 (2010).
30. Hahn, P. J. & McIntyre, C. C. Modeling shifts in the rate and pattern of subthalamopallidal network activity during deep brain stimulation. *J. Comput. Neurosci.* **28**, 425–441 (2010).
31. Santaniello, S. et al. Therapeutic mechanisms of high-frequency stimulation in Parkinson's disease and neural restoration via loop-based reinforcement. *Proc. Natl Acad. Sci. USA* **112**, E586–E595 (2015).
32. Stefanescu, R. A., Shivakeshavan, R. & Talathi, S. S. Computational models of epilepsy. *Seizure* **21**, 748–759 (2012).
33. Sritharan, D. & Sarma, S. V. Fragility in dynamic networks: application to neural networks in the epileptic cortex. *Neural Comput.* **26**, 2294–2327 (2014).
34. Feng, X. J., Shea-Brown, E., Greenwald, B., Kosut, R. & Rabitz, H. Optimal deep brain stimulation of the subthalamic nucleus–a computational study. *J. Comput. Neurosci.* **23**, 265–282 (2007).
35. Brocker, D. T. et al. Optimized temporal pattern of brain stimulation designed by computational evolution. *Sci. Transl. Med.* **9**, eaah3532 (2017).
36. Liu, J., Khalil, H. K. & Oweiss, K. G. Model-based analysis and control of a network of basal ganglia spiking neurons in the normal and parkinsonian states. *J. Neural Eng.* **8**, 045002 (2011).
37. Santaniello, S., Fiengo, G., Glielmo, L. & Grill, W. M. Closed-loop control of deep brain stimulation: a simulation study. *IEEE Trans. Neural Syst. Rehabil. Eng.* **19**, 15–24 (2011).
38. Millard, D. C., Wang, Q., Gollnick, C. A. & Stanley, G. B. System identification of the nonlinear dynamics in the thalamocortical circuit in response to patterned thalamic microstimulation in vivo. *J. Neural Eng.* **10**, 066011 (2013).
39. Bolus, M., Willats, A., Whitmire, C., Rozell, C. & Stanley, G. Design strategies for dynamic closed-loop optogenetic neurocontrol in vivo. *J. Neural Eng.* **15**, 026011 (2018).
40. Basu, I. et al. A neural mass model to predict electrical stimulation evoked responses in human and non-human primate brain. *J. Neural Eng.* **15**, 066012 (2018).
41. Khambhati, A. N. et al. Functional control of electrophysiological network architecture using direct neurostimulation in humans. *Netw. Neuroscience* **3**, 848–877 (2019).
42. Stiso, J. et al. White matter network architecture guides direct electrical stimulation through optimal state transitions. *Cell Rep.* **28**, 2554–2566 (2019).
43. Hsieh, H.-L., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Multiscale modeling and decoding algorithms for spike-field activity. *J. Neural Eng.* **16**, 016018 (2018).
44. de Hemptinne, C. et al. Therapeutic deep brain stimulation reduces cortical phase-amplitude coupling in Parkinson's disease. *Nat. Neurosci.* **18**, 779–786 (2015).
45. Kondabolu, K. et al. Striatal cholinergic interneurons generate beta and gamma oscillations in the corticostriatal circuit and produce motor deficits. *Proc. Natl Acad. Sci. USA* **113**, E3159–E3168 (2016).
46. Pasqualetti, F., Zampieri, S. & Bullo, F. Controllability metrics, limitations and algorithms for complex networks. *IEEE Trans. Control Netw. Syst.* **1**, 40–52 (2014).
47. Gu, S. et al. Controllability of structural brain networks. *Nat. Commun.* **6**, 8414 (2015).
48. Muldoon, S. F. et al. Stimulation-based control of dynamic brain networks. *PLoS Comput. Biol.* **12**, e1005076 (2016).
49. Tang, E. et al. Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nat. Commun.* **8**, 1252 (2017).
50. Medaglia, J. D. et al. Network controllability in the inferior frontal gyrus relates to controlled language variability and susceptibility to TMS. *J. Neurosci.* **38**, 6399–6410 (2018).
51. Pesaran, B. et al. Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation. *Nat. Neurosci.* **21**, 903–919 (2018).
52. Ljung, L. *System Identification* (Prentice Hall, 1999).
53. Tass, P. A. A model of desynchronizing deep brain stimulation with a demand-controlled coordinated reset of neural subpopulations. *Biol. Cybern.* **89**, 81–88 (2003).
54. Tass, P. A. & Hauptmann, C. Therapeutic modulation of synaptic connectivity with desynchronizing brain stimulation. *Int. J. Psychophysiol.* **64**, 53–61 (2007).
55. Tass, P. A. et al. Coordinated reset has sustained aftereffects in Parkinsonian monkeys. *Ann. Neurol.* **72**, 816–820 (2012).
56. Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl Acad. Sci. USA* **101**, 3747–3752 (2004).
57. Van Wijk, B. C., Stam, C. J. & Daffertshofer, A. Comparing brain networks of different size and connectivity density using graph theory. *PLoS ONE* **5**, e13701 (2010).

58. Sani, O. G., Abbaspourazad, H., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nat. Neurosci.* **24**, 140–149 (2021).
59. Herzfeld, D. J., Kojima, Y., Soetedjo, R. & Shadmehr, R. Encoding of action by the Purkinje cells of the cerebellum. *Nature* **526**, 439 (2015).
60. Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51 (2012).
61. Anumanchipalli, G. K., Chartier, J. & Chang, E. F. Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498 (2019).
62. Vaz, A. P., Inati, S. K., Brunel, N. & Zaghloul, K. A. Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. *Science* **363**, 975–978 (2019).
63. Markowitz, D. A., Curtis, C. E. & Pesaran, B. Multiple component networks support working memory in prefrontal cortex. *Proc. Natl Acad. Sci. USA* **112**, 11084–11089 (2015).
64. Denfield, G. H., Ecker, A. S., Shinn, T. J., Bethge, M. & Tolias, A. S. Attentional fluctuations induce shared variability in macaque primary visual cortex. *Nat. Commun.* **9**, 2654 (2018).
65. Han, X., Xian, S. X. & Moore, T. Dynamic sensitivity of area V4 neurons during saccade preparation. *Proc. Natl Acad. Sci. USA* **106**, 13046–13051 (2009).
66. Jamali, M. et al. Dorsolateral prefrontal neurons mediate subjective decisions and their variation in humans. *Nat. Neurosci.* **22**, 1010–1020 (2019).
67. Zavala, B. A., Jang, A. I. & Zaghloul, K. A. Human subthalamic nucleus activity during non-motor decision making. *eLife* **6**, e31007 (2017).
68. Herzfeld, D. J., Kojima, Y., Soetedjo, R. & Shadmehr, R. Encoding of error and learning to correct that error by the Purkinje cells of the cerebellum. *Nat. Neurosci.* **21**, 736–743 (2018).
69. Zheng, H. J., Wang, Q. & Stanley, G. B. Adaptive shaping of cortical response selectivity in the vibrissa pathway. *J. Neurophysiol.* **113**, 3850–3865 (2015).
70. Froudarakis, E. et al. Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci.* **17**, 851–857 (2014).
71. Susilaradeya, D. et al. Extrinsic and intrinsic dynamics in movement intermittency. *eLife* **8**, e40145 (2019).
72. Hall, T. M., Carvalho, Fde & Jackson, A. A common structure underlies low-frequency cortical dynamics in movement, sleep, and sedation. *Neuron* **83**, 1185–1199 (2014).
73. Abbaspourazad, H., Choudhury, M., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Multiscale low-dimensional motor cortical state dynamics predict naturalistic reach-and-grasp behavior. *Nat. Commun.* (in the press).
74. Shenoy, K. V. & Carmena, J. M. Combining decoder design and neural adaptation in brain-machine interfaces. *Neuron* **84**, 665–680 (2014).
75. Abbaspourazad, H., Hsieh, H.-L. & Shanechi, M. M. A multiscale dynamical modeling and identification framework for spike-field activity. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**, 1128–1138 (2019).
76. Kao, J. C. et al. Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat. Commun.* **6**, 7759 (2015).
77. Irwin, Z. et al. Neural control of finger movement via intracortical brain–machine interface. *J. Neural Eng.* **14**, 066004 (2017).
78. Vaskov, A. K. et al. Cortical decoding of individual finger group motions using ReFIT Kalman filter. *Front. Neurosci.* **12**, 751 (2018).
79. Keller, C. J. et al. Mapping human brain networks with cortico-cortical evoked potentials. *Phil. Trans. R. Soc. B* **369**, 20130528 (2014).
80. Holtzheimer, P. E. et al. Subcallosal cingulate deep brain stimulation for treatment-resistant depression: a multisite, randomised, sham-controlled trial. *Lancet Psychiat.* **4**, 839–849 (2017).
81. Dougherty, D. D. et al. A randomized sham-controlled trial of deep brain stimulation of the ventral capsule/ventral striatum for chronic treatment-resistant depression. *Biol. Psychiatry* **78**, 240–248 (2015).
82. Ezzyat, Y. et al. Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. *Nat. Commun.* **9**, 365 (2018).
83. Deadwyler, S. A. et al. A cognitive prosthesis for memory facilitation by closed-loop functional ensemble stimulation of hippocampal neurons in primate brain. *Exp. Neurol.* **287**, 452–460 (2017).
84. Zanos, S., Richardson, A. G., Shupe, L., Miles, F. P. & Fetz, E. E. The Neurochip-2: an autonomous head-fixed computer for recording and stimulating in freely behaving monkeys. *IEEE Trans. Neural Syst. Rehabil. Eng.* **19**, 427–435 (2011).
85. Zanos, S., Rembado, I., Chen, D. & Fetz, E. E. Phase-locked stimulation during cortical beta oscillations produces bidirectional synaptic plasticity in awake monkeys. *Curr. Biol.* **28**, 2515–2526 (2018).
86. Etkin, A. et al. Using fMRI connectivity to define a treatment-resistant form of post-traumatic stress disorder. *Sci. Transl. Med.* **11**, eaal3236 (2019).
87. Ahmadipour, P., Yang, Y., Chang, E. F. & Shanechi, M. M. Adaptive tracking of human ECoG network dynamics. *J. Neural Eng.* https://doi.org/10.1088/1741-2552/abae42 (2020).
88. Mazzoni, A. et al. Computing the local field potential (LFP) from integrate-and-fire network models. *PLoS Comput. Biol.* **11**, e1004584 (2015).
89. Tehovnik, E., Tolias, A., Sultan, F., Slocum, W. & Logothetis, N. Direct and indirect activation of cortical neurons by electrical microstimulation. *J. Neurophysiol.* **96**, 512–521 (2006).
90. Haber, S. N. in *Decision Neuroscience: An Integrative Perspective* (eds Dreher, J.-C. & Tremblay, L.) 3–19 (Elsevier, 2017).
91. Choi, J., Goncharov, V., Kleinbart, J., Orsborn, A. & Pesaran, B. Monkey-MIMMS: Towards automated cellular resolution large-scale two-photon microscopy in the awake macaque monkey. In *40th Conf. Proc. IEEE Eng. Med. Biol. Soc.* 3013–3016 (IEEE, 2018).
92. Kleinbart, J. E. et al. A modular implant system for multimodal recording and manipulation of the primate brain. In *40th Conf. Proc. IEEE Eng. Med. Biol. Soc.* 3362–3365 (IEEE, 2018).
93. Bighamian, R., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Sparse model-based estimation of functional dependence in high-dimensional field and spike multiscale networks. *J. Neural Eng.* **16**, 056022 (2019).
94. Wang, C. & Shanechi, M. M. Estimating multiscale direct causality graphs in neural spike-field networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**, 857–866 (2019).
95. Yang, Y., Sani, O., Chang, E. F. & Shanechi, M. M. Dynamic network modeling and dimensionality reduction for human ECoG activity. *J. Neural Eng.* **16**, 056014 (2019).
96. Garcia, L., d Alessandro, G., Bioulac, B. & Hammond, C. High-frequency stimulation in Parkinson's disease: more or less? *Trends Neurosci.* **28**, 209–216 (2005).
97. Qiao, S., Brown, K. A., Orsborn, A. L., Ferrentino, B. & Pesaran, B. Development of semi-chronic microdrive system for large-scale circuit mapping in macaque mesolimbic and basal ganglia systems. In *38th Conf. Proc. IEEE Eng. Med. Biol. Soc.* 5825–5828 (IEEE, 2016).
98. Dotson, N. M., Hoffman, S. J., Goodell, B. & Gray, C. M. A large-scale semi-chronic microdrive recording system for non-human primates. *Neuron* **96**, 769–782 (2017).
99. Yang, Y. et al. Developing a personalized closed-loop controller of medically-induced coma in a rodent model. *J. Neural Eng.* **16**, 036022 (2019).
100. Yang, Y. & Shanechi, M. M. An adaptive and generalizable closed-loop system for control of medically induced coma and other states of anesthesia. *J. Neural Eng.* **13**, 066019 (2016).
101. Lilly, J. C., Hughes, J. R., Alvord, E. C. Jr & Galkin, T. W. Brief, noninjurious electric waveform for stimulation of the brain. *Science* **121**, 468–469 (1955).
102. Herrington, T. M., Cheng, J. J. & Eskandar, E. N. Mechanisms of deep brain stimulation. *J. Neurophysiol.* **115**, 19–38 (2015).
103. Hashimoto, T., Elder, C. M. & Vitek, J. L. A template subtraction method for stimulus artifact removal in high-frequency deep brain stimulation. *J. Neurosci. Methods* **113**, 181–186 (2002).
104. Erez, Y., Tischler, H., Moran, A. & Bar-Gad, I. Generalized framework for stimulus artifact removal. *J. Neurosci. Methods* **191**, 45–59 (2010).
105. Babadi, B. & Brown, E. N. A review of multitaper spectral analysis. *IEEE Trans. Biomed. Eng.* **61**, 1555–1564 (2014).
106. Schwartz, A. B., Cui, X. T., Weber, D. J. & Moran, D. W. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron* **52**, 205–220 (2006).
107. Thakor, N. V. Translating the brain-machine interface. *Sci. Transl. Med.* **5**, 210ps17 (2013).
108. So, K., Dangi, S., Orsborn, A. L., Gastpar, M. C. & Carmena, J. M. Subject-specific modulation of local field potential spectral power during brain–machine interface control in primates. *J. Neural Eng.* **11**, 026002 (2014).
109. Stavisky, S. D., Kao, J. C., Nuyujukian, P., Ryu, S. I. & Shenoy, K. V. A high performing brain–machine interface driven by low-frequency local field potentials alone and together with spikes. *J. Neural Eng.* **12**, 036009 (2015).
110. Sarnthein, J. & Jeanmonod, D. High thalamocortical theta coherence in patients with Parkinson's disease. *J. Neurosci.* **27**, 124–131 (2007).
111. Neumann, W.-J. et al. Subthalamic synchronized oscillatory activity correlates with motor impairment in patients with Parkinson's disease. *Mov. Disord.* **31**, 1748–1751 (2016).
112. Wijk, B. Cvan et al. Subthalamic nucleus phase–amplitude coupling correlates with motor impairment in Parkinson's disease. *Clin. Neurophysiol.* **127**, 2010–2019 (2016).
113. Van Overschee, P. & De Moor, B. *Subspace Identification for Linear Systems: Theory, Implementation and Applications* (Springer Science & Business Media, 2012).
114. Schalk, G. et al. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *J. Neural Eng.* **4**, 264 (2007).
115. Pistohl, T., Ball, T., Schulze-Bonhage, A., Aertsen, A. & Mehring, C. Prediction of arm movement trajectories from ECoG-recordings in humans. *J. Neurosci. Methods* **167**, 105–114 (2008).

116. Bansal, A. K., Truccolo, W., Vargas-Irwin, C. E. & Donoghue, J. P. Decoding 3D reach and grasp from hybrid signals in motor and premotor cortices: spikes, multiunit activity, and local field potentials. *J. Neurophysiol.* **107**, 1337–1355 (2011).
117. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
118. Zhang, Q. Using wavelet network in nonparametric estimation. *IEEE Trans. Neural Netw.* **8**, 227–236 (1997).
119. Akaike, H. in *Selected Papers of Hirotugu Akaike* (eds Parzen, E. et al.) 199–213 (Springer, 1998).

Acknowledgements

We acknowledge support of the Army Research Office under contract W911NF-16-1-0368 (to M.M.S.) as part of the collaboration between the US Department of Defense, the UK Ministry of Defence and the UK Engineering and Physical Research Council under the Multidisciplinary University Research Initiative. We also acknowledge support of US National Institutes of Health BRAIN grant R01-NS104923 (to B.P. and M.M.S.). Finally, the we acknowledge the Defense Advanced Research Projects Agency under Cooperative Agreement Number W911NF-14-2-0043 (to M.M.S. and B.P.), issued by the Army Research Office contracting office in support of the DARPA SUBNETS programme. The views, opinions and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government. We thank B. Goodell, C. Gray, J. E. Kleinbart and A. Orsborn for assistance with chamber and microdrive system design; S. Frey and B. Hynes for custom modifications to the Brainsight system; R. Shewcraft, J. Choi, M. Rubiano, Y. Jang and O. Martin for help with animal preparation and care; and K. Brown for help with MRI analysis.

Author contributions

M.M.S. and Y.Y. conceived the study and developed the IO modelling framework. Y.Y. and M.M.S. designed the multi-trial stochastic stimulation and cross-validation. Y.Y., O.G.S., S.Q., B.P. and M.M.S. designed the stimulation experiments. S.Q. and B.P. implemented the stimulation experiments. S.Q., J.I.S, B.F. and B.P. performed the experiments and data collection. Y.Y. and M.M.S. implemented and performed the modelling and analyses. O.G.S., Y.Y. and M.M.S. designed and implemented the closed-loop simulations. M.M.S. supervised all the modelling and analysis work. B.P. supervised all the experimental work. Y.Y. and M.M.S. wrote the manuscript with input from S.Q., O.G.S. and B.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41551-020-00666-w.

Correspondence and requests for materials should be addressed to M.M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Maryam M. Shanechi

Last updated by author(s): Aug 21, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The computer code used to stream the raw neural data to disk is implemented in MATLAB R2015a and C. The custom computer code used to generate and deliver the designed stimulation waveform is implemented in MATLAB R2015a. The custom computer code in this study is available at https://github.com/ShanechiLab/DynamicStimulation.
Data analysis	Offline analyses are implemented as custom MATLAB R2018b code. The custom computer code in this study is available at https://github.com/ShanechiLab/DynamicStimulation.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data supporting the results in this study are available within the paper and its Supplementary Information. The raw data are too large to be shared publicly but they are available for research purposes from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We applied our modeling framework to sixteen independent MN stimulation datasets collected in two non-human primates (NHPs) and ranging across four different stimulation sites. To demonstrate our modeling framework as an enabling technology, we used two NHPs to show its generalization across subjects and we stimulated from four different brain sites to show its generalization across stimulation sites. We applied our neural adaptation analyses to additional sixteen independent constant stimulation datasets collected in the same two NHPs and ranging across the same four stimulation sites as the MN stimulation datasets. Demonstration in two NHPs is standard for NHP electrophysiology studies. For main comparisons made in the main figures, the effect size (Cohen's d) was at least 0.2955, and by assuming a normal distribution of data, with the above sample size, the estimated power was at least 0.9946 given a type-I error of 0.05.
Data exclusions	No data were excluded from the analyses.
Replication	The results of the modeling framework were replicated for all sixteen independent MN stimulation datasets collected in two NHPs and ranging across four different stimulation sites. The results of the neural adaptation were replicated for all sixteen independent MN stimulation datasets and all sixteen independent constant stimulation datasets. All attempts at replication were successful.
Randomization	Not relevant to the study. There was no group allocation.
Blinding	Not relevant to the study. There was no blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Two male rhesus macaques (Macaca mulatta, Monkey A and Monkey M) participated in the study. Monkey A and Monkey M were 16 and 7 years old, respectively, at the time of experiments.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All surgical and experimental procedures were performed in compliance with the National Institute of Health Guide for Care and Use of Laboratory Animals and were approved by the New York University Institutional Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Not relevant to the study. Structural and diffusion weighted imaging during anesthesia. MRI data was used to visualize the brain regions on the 3D reconstructed monkey brains from MRI data.
Design specifications	Not relevant to the study. No task/trial involved.
Behavioral performance measures	Not relevant to the study. No behaviors involved. Animals were anesthetized with a constant IV infusion of 0.5mg/hr/kg of atracurium and 4mg/hr/kg of sufentanil, or isoflourane alone and placed in the scanner in the sphinx position.

Acquisition

Imaging type(s)	Structural and diffusion (Siemens Allegra using 3 elements out of a 4-channel phased array from Nova Medical Inc.)
Field strength	3 Tesla
Sequence & imaging parameters	T1-weighted magnetization-prepared rapid acquisition gradient-echo (MPRAGE) sequence (0.6 x 0.6 or 0.5 x 0.5 mm2 in-plane resolution, slice thickness: 0.6 or 0.5 mm)
Area of acquisition	A whole brain scan

Diffusion MRI ☒ Used ☐ Not used

Parameters	We performed multishell high-angular resolution diffusion imaging (HARDI) tractography. We acquired data using 64 gradient directions in 1.2 mm2 in-plane resolution (TR = 7000 ms; TE = 110 ms; b-values: 0, 750, 1500, 2250 s/mm2; FOV: 80 x 64 pixels; slices: 48; slice thickness: 1.2 mm; DWI to b0 ratio 65:1).

Preprocessing

Preprocessing software	FSL5.0
Normalization	Data were transformed using linear affine transformation in FSL to permit registration across scans within subject and to localize recording sites based on individual brain anatomy.
Normalization template	A template was not used to transform the data.
Noise and artifact removal	To correct for geometric distortions from field inhomogeneities caused by the non-zero off-resonance fields, data were collected with reversed phase-encode blips, forming pairs of images with distortions going in opposite directions. From these pairs the susceptibility-induced off-resonance field was estimated using FSL's TOPUP tool and the two images were combined into a single corrected image. Eddy currents generated by the 64 gradient directions were subsequently corrected using FSL's eddy tool.
Volume censoring	N/A

Statistical modeling & inference

Model type and settings	Not relevant to the study
Effect(s) tested	Not relevant to the study

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	Not relevant to the study
Correction	Not relevant to the study

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis