White Matter Network Architecture Guides Direct Electrical Stimulation Through Optimal State Transitions

Jennifer Stiso[1,2], Ankit N. Khambhati[2], Tommaso Menara[3], Ari E. Kahn[1,2],

Joel M. Stein[4], Sandihitsu R. Das[5], Richard Gorniak[6],

Joseph Tracy[7], Brian Litt[5,8], Kathryn A. Davis[5,8],

Fabio Pasqualetti[3], Timothy Lucas[8,9], & Danielle S. Bassett[2,8,10,11,*]

[1] Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104

[2] Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104

[3] Department of Mechanical Engineering, University of California, Riverside, CA 92521

[4] Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA, 19104

[5] Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA, 19104

[6] Department of Radiology, Thomas Jefferson University Hospital, Philadelphia, PA, 19107

[7] Department of Neurology, Thomas Jefferson University Hospital, Philadelphia, PA, 19107

[8] Penn Center for Neuroengineering and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104

[9] Department of Neurosurgery Hospital of the University of Pennsylvania, Philadelphia, PA 19104

[10] Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104

[11] Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA 19104

[*] To whom correspondence should be addressed: dsb@seas.upenn.edu

Abstract

Electrical brain stimulation is currently being investigated as a potential therapy for neurological disease. However, opportunities to optimize and personalize such therapies are challenged by the fact that the beneficial impact (and potential side effects) of focal stimulation on both neighboring and distant regions is not well understood. Here, we use network control theory to build a formal model of brain network function that makes explicit predictions about how stimulation spreads through the brain's white matter network and influences large-scale dynamics. We test these predictions using combined electrocorticography (ECoG) and diffusion weighted imaging (DWI) data from patients with medically refractory epilepsy undergoing evaluation for resective surgery, and who volunteered to participate in an extensive stimulation regimen. We posit a specific model-based manner in which white matter tracts constrain stimulation, defining its capacity to drive the brain to new states, including states associated with successful memory encoding. In a first validation of our model, we find that the true pattern of white matter tracts can be used to more accurately predict the state transitions induced by direct electrical stimulation than the artificial patterns of a topological or spatial network null model. We then use a targeted optimal control framework to solve for the optimal energy required to drive the brain to a given state. We show that, intuitively, our model predicts larger energy requirements when starting from states that are farther away from a target memory state. We then suggest testable hypotheses about which structural properties will lead to efficient stimulation for improving memory based on energy requirements. We show that the strength and homogeneity of edges between controlled and uncontrolled nodes, as well as the persistent modal controllability of the stimulated region, predict energy requirements. Our work demonstrates that individual white matter architecture plays a vital role in guiding the dynamics of direct electrical stimulation, more generally offering empirical support for the utility of network control theoretic models of brain response to stimulation.

Introduction

Direct electrical stimulation has demonstrated clinical utility in detecting brain abnormalities during surgery [40] as well as in mitigating symptoms of epilepsy, essential tremor, and dystonia [65, 54, 43]. Apart from clinical diagnosis and treatment, direct electrical stimulation has also been used to isolate areas responsible for complex higher-order cognitive functions including language [33, 46], semantic memory [64], and face perception [51]. An open and important question is whether such stimulation can be used to enhance cognitive function, and if so, whether stimulation parameters (e.g., intensity and location) can be optimized and personalized based on individual brain anatomy and physiology. While some studies demonstrate enhancements in spatial learning [39] and memory [19, 38, 18, 37, 66] following direct electrical stimulation, others show decrements [30]. Such conflicting evidence is also present in the literature surrounding other types of stimulation, including transcranial magnetic stimulation. Proposed explanations range from variations in stimulation intensity [56] to individual differences in brain connectivity [16].

A key challenge in circumscribing the utility of stimulation for cognitive enhancement or clinical intervention is the fact that we do not have a fundamental understanding of how an arbitrary stimulation paradigm applied to one brain area alters distributed neural activity in neighboring and distant brain areas [32, 38, 43]. Models of stimulation propagation through brain tissue range in complexity and biophysical realism [47], from those that only model the region being targeted to those that use finite element models to expand predictions throughout different tissue types [76], including both gray matter and white matter [35]. Even in the simpler simulations of the effects of stimulation on a local cell population, there are challenges in accounting for the orientation of cells, and the distance from the axon hillock, which can lead to strikingly different circuit behaviors [47]. In the more expansive studies of the effects of stimulation across the brain, it has been noted empirically that minute differences in electrode location can generate substantial differences in which white matter pathways are directly activated [44, 57], and that an individual's white matter connectivity can predict successful outcomes of stimulation [29]. These differences are particularly important in predicting response to therapy, given recent observations that stimulation to white matter may be particularly efficacious in treating depression [58] and epilepsy [71]. Despite these critical observations, a first-principles intuition regarding how the effects of stimulation might depend on the pattern of white matter connectivity present in a single human brain has remained elusive.

Network control theory provides a potentially powerful approach for modeling direct electrical stimulation in humans [67]. Building on recent advances in physics and engineering, network control theory characterizes a complex system as composed of nodes interconnected by edges [50], and then specifies a
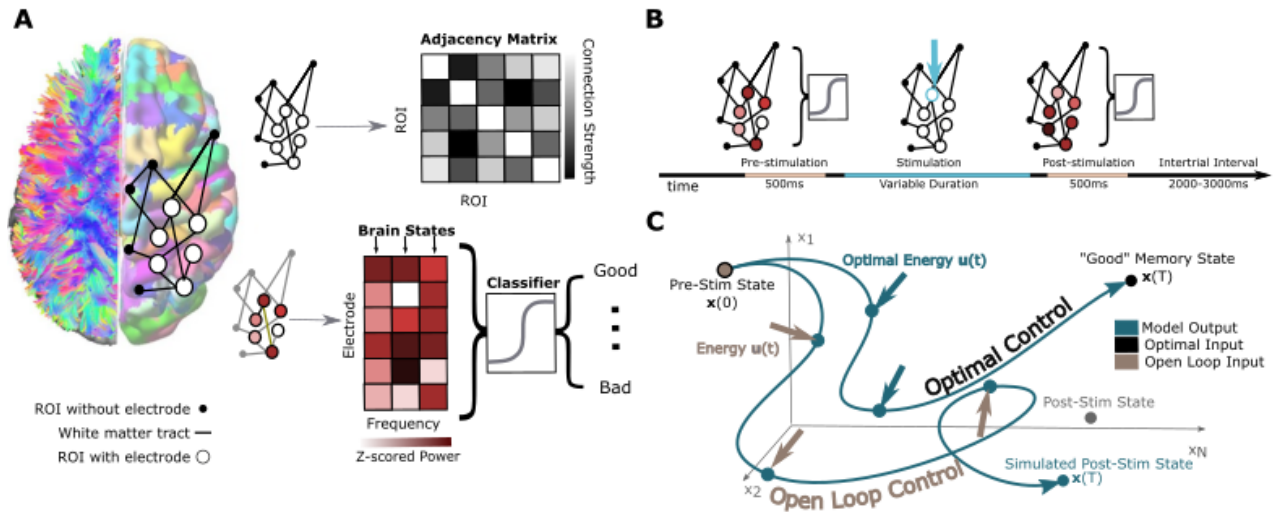
Figure 1: **Schematic of Methods.** *(A)* Depiction of network construction and definition of brain state. *(Left)* We segment subjects' diffusion weighted imaging data into $N = 234$ regions of interest using a Lausanne atlas [8]. We treat each region as a node in a whole-brain network, irrespective of whether or not the region contains an electrode. Edges between nodes represent mean quantitative anisotropy [75] along the streamlines connecting them. *(Right, Top)* Practically, we summarize the network in an $N \times N$ adjacency matrix. *(Right, Bottom)* A brain state is defined as the $N \times 1$ vector comprising activity across the N regions. Any element of the vector corresponding to a region with an electrode is defined as the band-limited power of ECoG activity measured by that electrode. Each brain state is also associated with an estimated probability of being in a good memory state, using a previously validated machine learning classifier approach [18]. *(B)* A schematic of a single stimulation trial. First, ECoG data is collected for 500 ms. Then, stimulation is applied to a given electrode for a variable duration. Finally, ECoG data is again collected after the stimulation. *(C)* A schematic of the open loop and optimal control paradigms. In the open loop design, energy $u(t)$ is applied *in silico* at the stimulation site to the initial, pre-stimulation brain state $x(0)$. The system will travel to some other state $x(T)$ as stipulated by our model of neural dynamics, and we will measure the similarity between that predicted state and the empirically observed post-stimulation state. In the optimal control design, the initial brain state $x(0)$ has some position in space that evolves over time towards a predefined target state $x(T)$. At every time point, we calculate the optimal energy $(u(t))$ required at the stimulating electrode to propel the system to the target state.

4

model of network dynamics to determine how external input affects the nodes' time-varying activity [42]. Drawing on canonical results from linear systems and structural controllability [34], this approach was originally developed in the context of technological, mechanical, and other man-made systems [53], but has notable relevance for the study of natural processes from cell signaling [14] to gene regulation [77]. In applying such a theory to the human brain, one first represents the brain as a network of nodes (brain regions) interconnected by structural edges (white matter tracts) [2], and then one posits a model of system dynamics that specifies how control input affects neural dynamics via propagation along the tracts [27]. Formal approaches built on this model address questions of where control points are positioned in the system [27, 68, 49, 73], as well as how to define spatiotemporal patterns of control input to move the system along a trajectory from an initial state to a desired final state [26, 3]. Intuitively, these approaches may be particularly useful in probing the effects of stimulation [49] and pharmacogenetic activation or inactivation [23] for the purposes of guiding transitions between cognitive states or treating abnormalities of brain network dynamics such as epilepsy [11, 17, 69], psychosis [5], or bipolar disorder [31]. However, this intuition has not yet been validated with direct electrical stimulation data.

Here, we posit a simple theory of brain network control, and we test its biological validity and utility in combined electrocorticography (ECoG) and diffusion weighted imaging (DWI) data from patients with medically refractory epilepsy undergoing evaluation for resective surgery. For each subject, we constructed a structural brain network where nodes represented regions of the Lausanne atlas [8] and where edges represented quantitative anisotropy between these regions estimated from diffusion tractography [75] (**Fig. 1A**). Upon this network, we stipulated a noise-free, linear, continuous-time, and time-invariant model of network dynamics [27, 3, 68, 26, 36], from which we built predictions about how regional activity would deviate from its initial state in the presence of exogenous control input to any given node. Using ECoG data acquired from the same individuals during an extensive direct electrical stimulation regimen (**Fig. 1B**), we test these theoretical predictions by representing (i) regional activity as an electrode's power in a given frequency band, (ii) the pre-stimulation state as the power prior to stimulation, and (iii) the post-stimulation state as the power after stimulation (**Fig. 1C**). After quantifying the relative accuracy of our theoretical predictions, we next use the model to make more specific predictions about the control energy required to optimally guide the brain from a pre-stimulation state to a state associated with good memory encoding. We then test these predictions using subject-level power-based biomarkers of good memory encoding extracted with a multivariate classifier from ECoG data collected during a verbal memory task [18]. Finally, we investigate how certain topological [36] and spatial [59] properties of a subject's network alter its response to direct electrical stimulation, and we ask whether that response is also modulated by

control properties of the area being stimulated [27, 49]. Essentially, our study posits and empirically tests a simple theory of brain network control, demonstrating its utility in predicting response to direct electrical stimulation.

Results

Our model assumes the time-invariant network dynamics

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

where the time-dependent state x is an $N \times 1$ vector ($N = 234$) whose i^{th} element gives the band-specific ECoG power in sensor i, $\mathbf{A}$ is the $N \times N$ adjacency matrix estimated from DWI data, $\mathbf{B}$ is an $N \times N$ matrix that selects the control set $\mathcal{K} = u_1, \ldots, u_p$ where p is the number of regions that receive exogenous control input. The input is constant in time and given by $u(t) = \beta \times I \times log(\omega) \times (\Delta t)$, where I is the stimulation amplitude in amperes, ω is the stimulation frequency in hertz, and Δt is the number of simulated samples (here, 950) divided by the stimulation duration in seconds. The free parameter β scales the magnitude of the input (see Materials and Methods). Intuitively, this model formalizes the hypothesis that white matter tracts constrain how stimulation affects brain state.

### Predicting Post-Stimulation States by Open Loop Control

We begin by exercising the model to determine whether our theory accurately predicts changes in brain state induced by direct electrical stimulation. Specifically, we simulate Eq. 1 to predict how stimulation alone (independent of other ongoing intrinsic dynamics) will alter brain state, given the structural adjacency matrix $\mathbf{A}$ and the initial state $x(0)$ comprised of the ECoG power at every node recorded pre-stimulation ($x_i = 1$ if node i is a region without electrodes, and the z-scored power otherwise; see Supplement for further details). For each stimulation event, we calculate the Pearson's correlation coefficient between the empirically observed post-stimulation state (an electrode by frequency matrix) and the predicted post-stimulation state at every time point in the simulated trajectory $x(t)$. To measure the capacity of the model simulation to predict the post-stimulation state, we measure the signed maximum correlation achieved across the model simulation time of arbitrary units. Accordingly, we compute a maximum correlation value between the model prediction and the empirically observed post-stimulation state for each stimulation trial ($\mu = 0.036 \pm 0.019$; **Fig. 2A**) and we observe that the mean of the maximum correlation values is significantly greater than zero (t-test $N = 16$, $t = 5.83$, $p = 3.31 \times 10^{-5}$). We note

6

that this correlation represents the impact of stimulation alone on linear dynamics, and does not take into account any other incoming stimuli from the surrounding environment, any ongoing cognitive or metabolic processes, nonlinear dependencies, or inter-frequency interactions [9, 6, 55]. Complementing this estimate, we were also interested in the time point (measured in arbitrary units) at which the trial reached its largest magnitude correlation (positive or negative) before decaying towards zero. We observed that the time at which the peak magnitude occurred differed across trials, having a mean of 298 ± 114 (**Fig. 2B**).
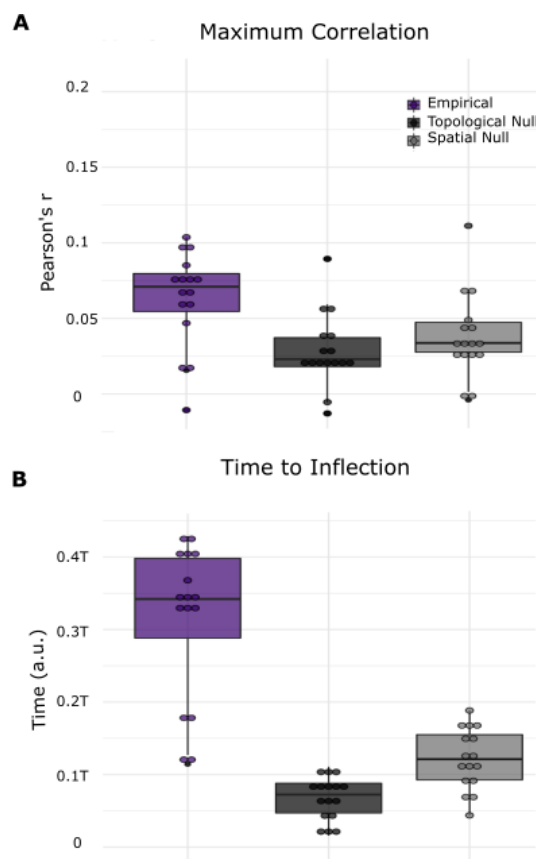
Figure 2: **Post-Stimulation Brain State Depends on White Matter Network Architecture.** *(A)* Box plots depicting the average signed maximum correlation between the empirically observed post-stimulation state and the predicted post-stimulation state at every time point in the simulated trajectory $x(t)$. *(B)* Box plots depicting the average time to reach the peak magnitude (positive or negative) correlation between the empirically observed post-stimulation state and the theoretically predicted post-stimulation state at every time point in the simulated trajectory $x(t)$. Time is measured in arbitrary units (a.u.). Color indicates theoretical predictions from Eq.1 where **A** is (i) the empirical network (purple) estimated from the diffusion imaging data, (ii) the topological null network (dark charcoal), and (iii) the spatial null network (light charcoal).

To determine the influence of network geometry on our model predictions, we compared the empirical observations to those obtained by replacing **A** in the simulation with one of two null model networks, each
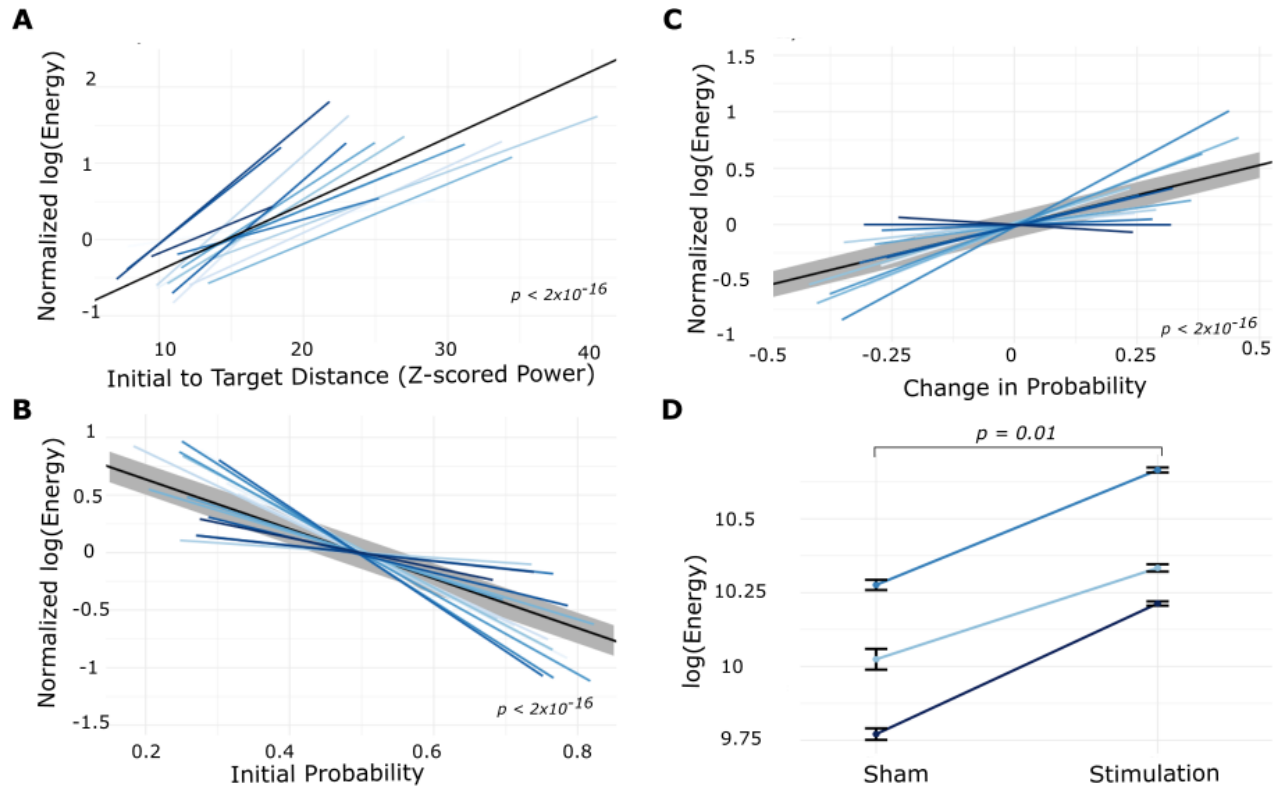
Figure 3: **Longer-Distance Trajectories Require More Stimulation Energy.** *(A)* The normalized energy required to transition between the initial state and the post-stimulation state, as a function of the Frobenius norm between the initial state and the post-stimulation state. The black solid line represents the best linear fit (with grey representing standard error), and is provided simply as a guide to the eye. Normalization is also done as a visual aid. *(B)* The energy required to transition to a good memory state, as a function of the initial probability of being in a good memory state. *(C)* The energy required to transition to a good memory state as a function of the empirical change in memory state resulting from stimulation. *(D)* In three experimental sessions that included both sham and stimulation trials, we calculated the energy required to reach the post-stimulation state or the post-sham state, rather than a target good memory state. Here we show the difference in energy required for sham state transitions in comparison to stimulation state transitions. Error bars indicate standard error of the mean across trials. Across all four panels, different shades of blue indicate different experimental sessions and subjects.

designed to independently remove specific geometric features of the structural network. First for each trial, we constructed a topological null: a randomly rewired network that preserved the edge distribution, number of nodes, and number of edges. Second, we constructed a spatial null: a randomly rewired network that additionally preserved the relationship between edge strength and Euclidean distance. Using a one-way ANOVA, we find a significant difference in maximum correlation values ($F(2, 45) = 17.4$, $p = 2.51 \times 10^{-6}$), and the time at which the maximum correlation values occur ($F(2, 45) = 25.6$, $p = 3.77 \times 10^{-8}$). We then performed *post-hoc* analyses and found that the topological null produced significantly weaker maximum correlations between the empirically observed post-stimulation state and the predicted post-stimulation states (paired t-test: $N = 16$, $t = 6.58$, uncorrected $p = 8.76 \times 10^{-6}$), which also peaked significantly earlier in time than the true data ($N = 16$, $t = 7.92$, uncorrected $p = 7.92 \times 10^{-7}$). The spatial null model also produced significantly weaker maximum correlations between the empirically observed post-stimulation state and the predicted post-stimulation states (paired t-test $N = 16$, $t = 5.83$, uncorrected $p = 3.31 \times 10^{-5}$), which also occurred significantly earlier in time than that observed in the true data ($N = 16$, $t = 3.84$, uncorrected $p = 1.60 \times 10^{-3}$). We observed consistent results in individual subjects, across all frequency bands, with different values of β and when using a smaller resolution atlas for whole brain parcellation. The only exception was that spatial null models peaked at the same time as empirical graphs in 2 of the 8 frequency bands (see Supplemental Methods). Overall, these observations support the notion that structural connections facilitate a rich repertoire of system dynamics following cortical stimulation, and directly constrain the dynamic propagation of stimulation energy in the human brain in a manner consistent with a simple linear model of network dynamics.

Inducing State Transitions by Optimal Network Control

We next sought to use the model to better understand the principles constraining brain state transitions in the service of cognitive function, and their response to exogenous perturbations in the form of direct electrical stimulation. Building on the network dynamics stipulated in **Eq. 1**, we used an optimal control framework to calculate the optimal amount of external input **u** to deliver to the control set K containing the stimulating electrode, driving the system from a specific pre-stimulation state towards a target post-stimulation state (**Fig. 1C**). This target post-stimulation state was defined as a period with high probability of successfully encoding a memory, and was operationalized using a previously validated classifier constructed from ECoG data from the same subjects during the performance of a verbal memory task [18] (**Fig. 1A**). Specifically, we use a cost function that minimizes both the energy and the difference of

the current state from the target state:

$$\min_u \int_0^T (x_T - x(t))^T \mathbf{S}(x_T - x(t)) + \rho \; u(t)^T u(t) dt, \tag{2}$$

where x_T is the target state, $\mathbf{S}$ is a diagonal $N \times N$ matrix that selects a subset of states to constrain (here, $\mathbf{S}$ is the identity and all diagonal entries are equal to 1), ρ is the importance of the energy penalty relative to the state penalty, and T is the time allotted for the simulation. Practically, we note that optimizing the cost function in Eq. 2 necessarily identifies simulated optimal control trajectories from the pre-stimulation state to a good memory state reasonably close to the target (final distance from target $\mu = 0.12 \pm 0.06$) with minimal error (range from 3.65×10^{-5} to 5.19×10^{-4}).

We begin by addressing the hypothesis that greater energy should be required to reach the target state when it is farther from the initial state. We operationalize this notion by defining distance in four different ways. First, we define distance as the Frobenius norm of the difference between initial and target states. We fit a linear mixed effects model to the integral of the input squared, or energy (here, $\mathbf{B}u$) in every trial, treating the Frobenius norm distance between initial and final state as a fixed effect, and treating subject as a random effect. We find that the distance between initial and final state is positively related to the energy required for the transition ($\beta = 8.3 \times 10^{-3}$, $p < 2 \times 10^{-16}$) (**Fig. 3A**). Second, we define distance by the memory capacity in the initial state. We fit a linear mixed effects model to the integral of the input squared in every trial, treating the initial state's probability of successfully encoding a memory as a fixed effect, and treating subject as a random effect. We find that the initial state's probability of successfully encoding a memory is negatively related to the energy required for the transition ($\beta = -0.18$, $p < 2 \times 10^{-16}$) (**Fig. 3B**), suggesting that states that begin closer to the target require less energy to reach the target. Third, we define distance as the observed change in memory state resulting from stimulation. We fit a linear mixed effects model to the input squared in every trial, treating the change in memory state as a fixed effect, and treating subject as a random effect. We find that the change in memory state is positively related to the energy required for the transition ($\beta = 8.3 \times 10^{-3}$, $p = 2 \times 10^{-16}$) (**Fig. 3C**).

Taken together, this set of results serves as a basic validation that transitions between nearby brain states will generally require less energy than transitions between distant states. This finding holds whether distance is defined in terms of the difference in Frobenius norm between matrices of regional power, or in terms of the estimated probability to support the cognitive process of memory encoding. In specificity analyses, we also determined whether these relationships were expected in appropriate random network null models. We observed that the relationships were significantly attenuated in theoretical predictions from Eq. 1 where $\mathbf{A}$ is either the topological null network ($p = 6.1 \times 10^{-4}$) or the spatial null network

($p = 0.0017$). Interestingly, we also found that the largest differences between the empirical relationships and those expected in the null networks were observed in the context of biological measures of distance (e.g., initial probability and change in probability), with only modest differences seen in the statistical measure of distance (the Frobenius norm).

As a fourth and final test of the biological relevance of these findings, we considered sham trials, where no stimulation was delivered, as compared to stimulation trials. Intuitively, we expect that the state that the brain reaches after stimulation is farther away from the initial state than the state that the brain reaches naturally at the conclusion of a sham trial. Consistent with this expectation, we observed that 2 out of the 3 experimental sessions that included sham stimulation displayed significantly larger distances (measured by the Frobenius norm) between pre-and post-stimulation states for stimulation conditions than for sham conditions (permutation test, $p < 6.8 \times 10^{-3}$). Given this difference, we tested whether more energy would be required to simulate the transition from the initial pre-stimulation state to the post-stimulation state, than from the initial pre-sham state to the post-sham state. We found consistently greater energy for stimulation trials compared to sham trials (paired t-test, $p = 0.01$; **Fig. 3D**). We further confirmed this finding with a non-parametric permutation test assessing differences in the distribution of energy values across trials for sham conditions and the distribution of energy values across trials for stimulation conditions (permutation test, $p < 2 \times 10^{-16}$ for all subjects). These observations support the notion that transitions between nearby brain states occur without stimulation (sham) and require little predicted energy, whereas transitions between distant brain states occur with stimulation and require greater predicted energy.

The Role of Network Topology on Stimulation-Based Control

While it is natural to posit that the distance between brain states is an important constraint on the ease of a state transition, there are other important principles that are also likely to play a critical role. Paramount among them is the architecture of the network available for the transmission of control signals. We therefore now turn to the question of which features of the network predict the amount of energy required for each transition from the pre-stimulation state to a good memory state. To address this question, we considered the empirical networks as well as the topological and spatial null model networks discussed earlier. We find that the optimal control input energy required for these state transitions differs across network types (one-way ANOVA $F(2, 75) = 4.00$, $p = 0.03$). In *post-hoc* testing, we found that the optimal control energy was significantly different between the empirical network and the topological null network (two-tailed t-test: $t = -2.6$, $p = 0.01$) (**Fig. 4A**), but not between the empirical network and the spatial null network ($p > 0.05$). This observation suggests that the spatial embedding that characterizes both the real
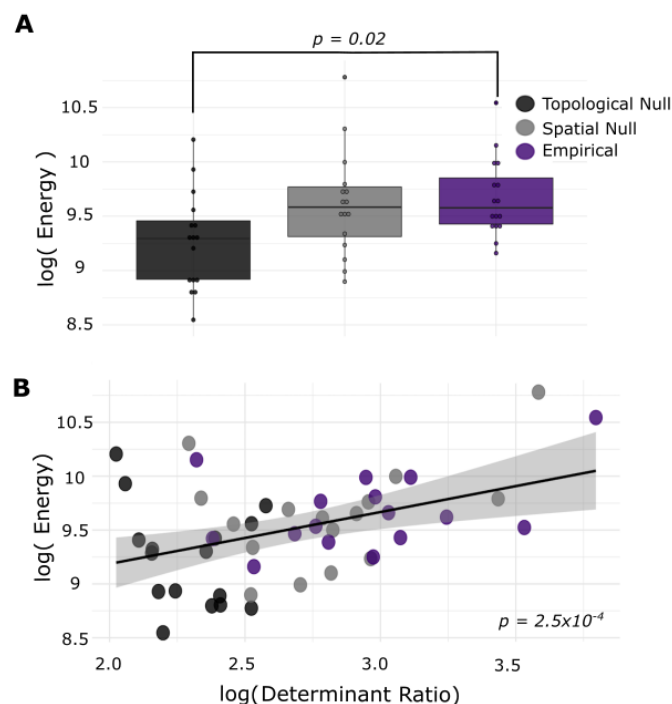
11

Figure 4: **Topological and Spatial Constraints on the Energy Required for Stimulation-Based Control.** *(A)* Average input energy required for each transition from the pre-stimulation state to a good memory state, as theoretically predicted from Eq. 1 where **A** is (i) the empirical network (purple) estimated from the diffusion imaging data, (ii) the topological null network (dark charcoal), and (iii) the spatial null network (light charcoal). *(B)* The relationship between the determinant ratio and the energy required for the transition from the pre-stimulation state to a good memory state. Note: The color scheme is identical to that used in panel *(A)*.

network and the spatial null network may increase the difficulty of control. In supplemental analyses, we test two additional spatially embedded null models that further preserve degree distribution and strength sequence, and we find similar average energies to the empirical and spatial null models discussed here (see Supplement). We hypothesized that the difference in optimal control energy could be mechanistically explained by the determinant ratio, a recently proposed metric quantifying the trade-off between connection strength (facilitating control) and connection homogeneity (hampering control) [36]. Intuitively, a network with a high determinant ratio will have weak, homogenous connections between the control nodes and nodes being controlled. We found that across all networks the determinant ratio explains a significant amount of variance in energy after accounting for network type (linear mixed effects model with network type and determinant ratio as fixed effects: $\chi^2 = 12.2$, $p = 4.7 \times 10^{-4}$) (**Fig. 4B**). These results support the notion that spatial embedding could impose energy barriers by compromising the trade-off between the strength and homogeneity of connections emanating from the stimulating electrode.
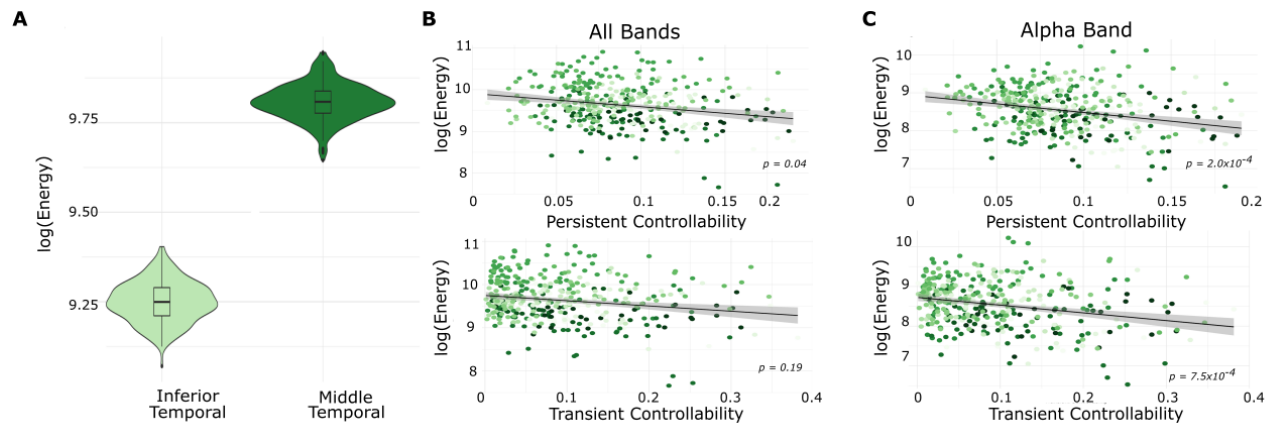
Figure 5: **Role of Local Topology Around the Region Being Stimulated.** *(A)* Transitions from the observed initial state to a good memory state required significantly greater energy when affected by the middle temporal sensors than when affected by the inferior temporal sensors. *(B)* Relationship between persistent (*top*) or transient (*bottom*) controllability of the stimulated region and the energy predicted from optimal transitions from the initial state to a good memory state. We only allow energy to be injected into a single electrode-containing region, and we consider a broadband state matrix. *(C)* As in panel *(B)* but when considering the α band state vector only.

Characteristics of Efficient Regional Controllers

Thus far, we have seen that the distance of the state transition and the architecture of the network available for the transmission of control signals both impact the energy required. However, neither of these factors addresses the potential importance of anatomical characteristics specific to the region being stimulated. Such regional effects are salient in the one subject in our patient sample who had multiple empirical stimulation sites spanning the same number of ROIs. In this patient, we found that transitions from the observed initial state to a good memory state required significantly greater energy when stimulation was delivered to electrodes in the middle temporal region than when stimulation was delivered to the inferior temporal region (permutation test, $p < 2 \times 10^{-16}$) (**Fig. 5A**). We hypothesized that this sensitivity to anatomical location could be mechanistically explained by regional persistent and transient modal controllability, which quantify the degree to which specific eigenmodes of the network's dynamics can be influenced by input applied to that region. Energetic input to nodes with high persistent controllability will result in large perturbations to slowly decaying modes of the system, while energetic input to nodes with high transient controllability will result in large perturbations to quickly decaying modes of the system.

To test our hypothesis, we simulated optimal trajectories from the initial state to a good memory state while only allowing energy to be injected into a single electrode-containing region (irrespective of whether or not empirical stimulation was applied there). We then compared the energy predicted from these

13

simulations to the regional controllability. We found a significant relationship between persistent (but not transient) modal controllability of the region being stimulated and the input energy of the state transition (linear mixed effects model accounting for subject: persistent controllability $\chi^2 = 3.89$, $p = 0.049$, transient controllability $\chi^2 = 1.69$, $p = 0.19$) (**Fig. 5B**). We note that the strength of the region being stimulated was not a significant predictor of energy (linear mixed effects model $\chi^2 = 3.5$, $p = 0.061$). Notably, we found that the broad-band effect was heavily driven by the α band (linear mixed effects model: persistent controllability $\chi^2 = 13.8$, $p = 2.0 \times 10^{-4}$, transient controllability $\chi^2 = 11.4$, $p = 7.5 \times 10^{-4}$; Bonferroni corrected for multiple comparisons across frequency bands) (**Fig. 5C**). Additionally, in the one subject that had two empirical stimulation locations, we observed that the middle temporal stimulation site with larger energy requirements had smaller persistent controllability (0.058) than the inferior temporal site with smaller energy (0.072). These findings suggest that the local white matter architecture of stimulated regions can support the selective control of slowly damping dynamics.

Effective Prediction of Energy Requirements

In the previous section, we presented a series of analyses with the goal of elucidating what aspects of brain state and white matter connectivity affect the energy requirements predicted by our model, in an effort to better understand the network wide effects of direct electrical stimulation. Here, we conclude by synthesizing these results into a single model to predict the energy requirements of a stimulation paradigm, given the persistent controllability of the region to be stimulated, the determinant ratio of the network to be controlled, and the probability of encoding a memory at the time of stimulation (Fig6A). We fit a random forest model to predict energy given these inputs from our data, and we compared the performance of this model to the performance of a distribution of 1000 models in which the association between energy values and predictors was permuted uniformly at random. We found that our model had an out-of-bag mean squared error of 9.28×10^{-3}, substantially lower than the null distribution ($\mu = 9.62 \times 10^{-3} \pm 2.97 \times 10^{-5}$). We also found that our model explained 93.2% of the variance in the predicted energy of the state transition. Random forest models also produce a measure of variable importance, which represents the degree to which including these variables tends to reduce the prediction error. We found that the determinant ratio was the most important (increased node purity = 627), followed by the persistent controllability (320), followed by the initial probability of encoding a memory (23.0). Broadly, these results suggest that the energy requirements for a specific state transition can be accurately predicted given simple features of the connectome and the current brain state.
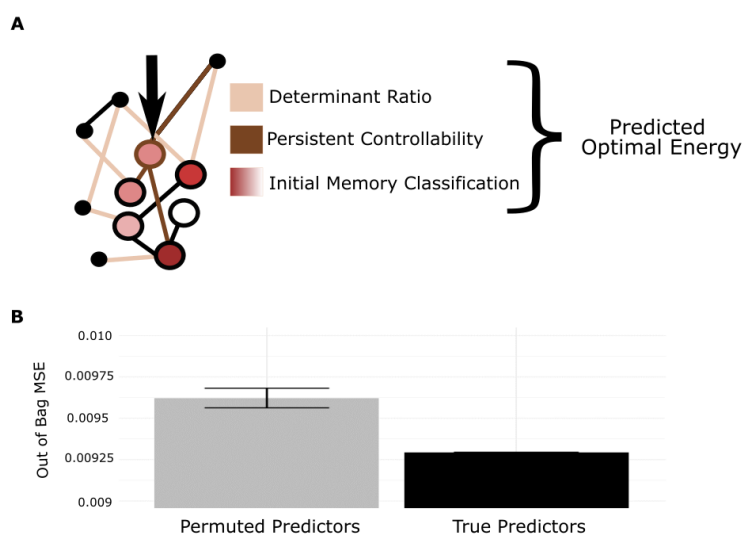
14

Figure 6: **Network Topology and Brain State Predict Energy Requirements.** *(A)* Schematic of the three topology and state features included in the random forest model that we built to predict energy requirements. Network level effects (tan) are captured by the determinant ratio, regional effects (brown) are captured by persistent controllability, and state-dependent effects (red) are captured by the initial memory state. *(B)* Comparison of the out-of-bag mean squared error for a model where each subject's determinant ratio, persistent controllability, and initial memory state are used to predict their required energy. We compared the performance of this model to the performance of a distribution of 1000 models in which the association between energy values and predictors was permuted uniformly at random.

Discussion

While direct electrical stimulation has great therapeutic potential, its optimization and personalization remains challenging, in part due to a lack of understanding of how focal stimulation impacts both neighboring and distant regions. Here use network control theory to test the hypothesis that the effect of direct electrical stimulation on brain dynamics is constrained by an individual's white matter connectivity. By stipulating a simplified noise-free, linear, continuous-time, and time-invariant model of neural dynamics, we demonstrate that time-varying changes in the pattern of ECoG power across brain regions is better predicted by an individual's true white matter connectivity than either topological or spatial network null models. We build on this observation by positing a model for brain state transitions in which the energy required for the state transition is minimized, as is the length of the trajectory through the available state space. We demonstrate that transitions between more distant states are predicted to require greater energy than transitions between nearby states; these results are particularly salient when distance is defined based on differences in the probability with which a cross-regional pattern of ECoG power supports memory encoding. In addition to the distance between initial and target states, we also find that regional and global characteristics of the network topology predict the energy required for the state transition: networks with

smaller determinant ratios (stronger, less homogeneous connections), and stimulation regions with higher persistent controllability, tend to demand less energy. Finally, we demonstrate that these two topological features in combination with the initial brain state explain 93% of the variance in required energy across subjects. Overall, our study supports the notion that control theoretic models of brain network dynamics provide biologically grounded, individualized hypotheses of response to direct electrical stimulation by accounting for how white matter connections constrain state transitions.

A Role for Control Theory in Modern Neuroscience

Developing theories, models, and methods for the control of neural systems is not a new goal in neuroscience. Whether in support of basic science (e.g., seminal experiments from Hodgkin and Huxley) or in support of clinical therapies (e.g., technological development in brain-machine interfaces or deep brain stimulation), efforts to control neural activity have produced a plethora of experimental tools with varying levels of complexity [63]. Building on these empirical advances, the development of a theory for control in neuroscience is a logical next step. Network control theory is one particularly promising option. In assimilating brain state and connectivity in a mathematical model [63], network control theory offers a first-principles approach to modeling neural dynamics, predicting its response to perturbations, and optimizing those perturbations to produce a desired outcome. In cellular neuroscience, network control theory has offered predictions of the functional role of individual neurons in *C. elegans*, and those predictions have been validated by perturbative experiments [74]. While the theory has also offered predictions in humans [27, 49, 11, 69, 31], these predictions have not been validated in accompanying perturbative experiments. Here we address this gap by examining the utility of network control theory in predicting empirically recorded brain states, and by validating the fundamental assumption that state transitions are constrained by an individual's white matter connectivity. The work provides theoretical support for emerging empirical observations that structural connectivity can predict the behavioral effects of stimulation [29], thus constituting an important first step in establishing the promise and utility of control theoretic models of brain stimulation.

The Principle of Optimal Control in Brain State Transitions

By positing a model for optimal brain state transitions, we relate expected energy expenditures to a change in the probability with which a pattern of ECoG power is associated with good memory encoding, directly relating the theory to a desired behavioral feature. This portion of the investigation was made possible by an important modeling advance addressing the challenge of simulating a trajectory whose control is

dominated by a single node: the stimulating electrode. This type of control is an intuitive way to model stimulation, where you only want to capture changes resulting from a single input source. However, prior work has demonstrated that while the brain is theoretically controllable from a single point, the amount of energy required can be so large as to make the control strategy impractical [27]. Here we extend prior models of optimal control [3, 26] by relaxing the input matrix $\mathbf{B}$ such that it allows large input to stimulated regions, but also allows small, randomly generated amounts of input at other nodes in the network. Practically, this approach greatly lowers the error of the calculation and also produces narrowly distributed trajectories for the same inputs (see Supplemental Methods).

Topological Influencers of Control

Beyond the distance of the state transition, we found that both local and global features of the network topology were important predictors of control energy. In line with previous work investigating controllability radii [48], energy requirements were lower for randomly rewired networks. Both empirical and topological graphs share the common feature of modularity [10], which is destroyed in random topological null models [59]. Prior theoretical work has demonstrated that modularity is one way in which to decrease the energy of control by decreasing the determinant ratio, a quantification of the relationship between the strength and heterogeneity of direct connections from the controlling node to others [36]. Here we confirmed that the determinant ratio accurately predicted the required energy, while leaving a small amount of variance unexplained. We expected that this unexplained variance could be somewhat accounted for by features of the local network topology surrounding the stimulated node [68]. Consistent with our expectation, we found that persistent controllability was the only significant predictor of energy across all frequency bands, indicating a specific role of slow modes in these state transitions. The effect was particularly salient in the α-band, whose role in memory encoding is well-known [21, 7].

Methodological Considerations

Primary Data

As with any model of complex biological systems, our results must be interpreted in the context of the underlying data. First, we note that DWI data provides an incomplete picture of white matter organization, and even state-of-the-art tractography algorithms can identify spurious connections [70]. As higher resolution imaging, reconstruction, and tractography methods emerge, it will be important to replicate the results we report here. Second, while ECoG data provides high temporal resolution, it is collected from

17

patients with epilepsy and results might not generalize to a healthy population [52]. However, it is worth noting that recent work has shown that tissue damage resulting from recurrent seizures can be minimal [60], and most electrodes are not placed in epileptic tissue [52]. Nevertheless, this population can display atypical physiological signatures of memory [22], as well as atypical white matter connectivity [25]. It will be important in future to extend this work to non-invasive techniques accessible to healthy individuals.

Modelling Assumptions

Our results must also be interpreted in light of model assumptions. First, we note that our model assumes linear network dynamics. While the brain is not a linear system, such simplified approximations can predict features of fMRI data [28], predict the control response of nonlinear systems of coupled oscillators [20], and more generally provide enhanced interpretability over nonlinear models. Nevertheless, considering control in nonlinear models of neural dynamics will constitute an important next step. Second, we consider a relaxed input matrix to ensure that state transitions are primarily influenced by the set of stimulating electrodes and to a lesser extent non-stimulating electrodes. This choice is not a true representation of single point control, but instead reflects the fact that the system is constantly modulated by endogenous sources [26, 3]. Lastly, our model uses a time-invariant connectivity matrix. While DWI data is relatively stable over short time-scales, repeated stimulation can result in dynamic changes in plasticity [45] that are not captured here.

Defining Brain States

In our model, a brain state represents the z-scored power across electrodes in eight logarithmically spaced frequency bands from 1 to 200 Hz. This choice was guided by (i) the goal of maintaining consistency with the brain states on which the memory classifier was trained, and (ii) the fact that power spectra are well-documented behavioral analogs for memory [18, 21, 7]. Yet, since many power calculations require convolution with a sine wave, power is insensitive to non-sinusoidal and phase-dependent features of the signal [62, 13, 72]. It would be interesting in future to explore transitions in other state spaces, such as instantaneous voltage [62]. Lastly, it is important to note that our algorithm controls each frequency band independently, although incorporating inter-frequency coupling [55, 4, 9] could be an interesting direction for future work.

Conclusions and Future Directions

Our study begins to explore the role of white matter connectivity in guiding direct electrical stimulation, with the goal of driving brain dynamics towards states with a high probability of memory encoding. We demonstrate that our model of targeted direct electrical stimulation tracks well with biological intuitions, and is influenced by both regional and global topological properties of underlying white matter connectivity. Overall, we have shown that our control theoretic model is a promising method that has potential to inform hypotheses about the outcome of direct electrical stimulation.

Materials and Methods

Data Acquisition and Preprocessing

Electrocorticography data were collected from eleven subjects (age 32 +/- 100 years, 63.6% male and 36.4% female) at Thomas Jefferson University Hospital and the Hospital of the University of Pennsylvania as part of a multi-center project designed to assess the effects of electrical stimulation on memory-related brain function. The research protocol was approved by the institutional review board (IRB) at each hospital and informed consent in writing was obtained from each participant. Electrophysiological data were collected from electrodes implanted subdurally on the cortical surface as well as deep within the brain parenchyma. In each case, the clinical team determined the placement of the electrodes to best localize epileptogenic regions. Subdural contacts were arranged in both strip and grid configurations with an inter-contact spacing of 10 mm. Depth electrodes had 8-12 contacts per electrode, with 3.5 mm spacing. In our model, a brain state represented the z-scored power across electrodes in eight logarithmically spaced frequency bands from 1 to 200 Hz.

Diffusion volumes were skull-stripped using FSL's BET, v5.0.10. Volumes were subsequently corrected for eddy currents and motion using FSL's EDDY tool, v.5.0.10 [1]). Anatomical scans were processed with FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) v6.0.0. Surface reconstructions were used to generate subject-specific parcellations based on the Lausanne atlas from the Connectome Mapper Toolbox [15]. Each parcel was then individually warped into the subject's diffusion space. Using DSI-Studio (http://dsi-studio.labsolver.org), orientation density functions (ODFs) within each voxel were reconstructed from the corrected scans using GQI [75]. We then used the reconstructed ODFs to perform a whole-brain deterministic tractography using the derived QA values in DSI-Studio [75]. We generated 1,000,000 streamlines per subject, with a maximum turning angle of 35 degrees and a maximum length of 500mm [12]. We held the number of streamlines between participants constant [24].

Memory State Classification

Prior to collecting the stimulation data used in this study, non-stimulation ECoG data was collected from each subject as they were performing a verbal memory task. This data was used to train a memory classifier, built on the spectral power averaged across the time dimension for each word-encoding epoch. Each subject's personalized classifier was then used to return a probability of being in a good memory state for each pre- and post-stimulation recording (for further details, see [18]). We used this information to define the target state for our simulations as the average of the top 5% of states with the largest probabilities associated with them. The threshold of 5% was chosen as the smallest threshold that reliably included a sufficient average number of trials, with the goal of only selecting for memory-relevant features and not noise. The probabilities associated with these final target states ranged from 0.61 to 0.74.

Post-Stimulation State Correlations

We simulated stimulation to a given region in the Lausanne atlas from the observed pre-stimulation state ($x(i)$ is the z-scored power if i is a region with an electrode, $x(i) = 1$ otherwise). We then calculated the two-dimensional Pearson's correlation coefficient between the empirically observed post-stimulation state and the predicted post-stimulation state at time points $t = 5$ to $t = T$ in the simulated trajectory $x(t)$. The time points $t < 5$ were excluded to prevent the initial state from being considered as the peak. We calculated two statistics of interest: the maximum correlation reached and the time at which the largest magnitude (positive or negative) correlation occurred.

Null Models

We compared the empirically observed values – of the maximum correlation reached and the time at which the largest magnitude correlation occurred – to those expected under two null models: (i) a topological null model that preserved only the number of edges and their total strength, and (ii) a spatially embedded null model that also preserved the relationship between edge strength and edge distance. Instantiations of the topological null model were generated using the Brain Connectivity Toolbox [61]. The rewiring algorithm begins by randomly choosing two pairs of edges ($i \rightarrow j$ and $k \rightarrow l$) and continues by swapping their origin and termination points ($i \rightarrow k$ and $j \rightarrow l$). Here, we performed 2×10^4 bidirectional edge swaps per network. Instantiations of the spatially embedded model were generated using code from [59]. The rewiring algorithm begins by calculating the Euclidean distance between the average coordinates of all regions in the Lausanne atlas, and continues by removing the effect of distance on the mean and variance

of the edge weights, randomly rewiring, and then adding the effect of distance back to the newly rewired graph. For both topological and spatial null model analyses, a new random graph was generated for every trial (minimum number of trials was 192). Null models were created from the stabilized rather than raw versions of the structural matrices, and – in the optimal control analyses – were also scaled by a parameter γ to reduce the error of the calculation (see Supplemental Materials).

Optimal Network Control

To quantify the ease of controlling the system, we calculated a single measure of energy for every trajectory. We used a measure of total input energy that incorporated the weights of $\mathbf{B}$ in addition to the input $\mathbf{u}$ because the entries of $\mathbf{B}$ were graded:

$$E_{\kappa, \mathbf{x}_0 \mathbf{x}_T} = \int_0^T ||\mathbf{B}_\kappa \mathbf{u}_{\mathbf{x}_0 \mathbf{x}_T}||_2^2 \mathrm{d}t. \tag{3}$$

More specifically, rather than being characterized by binary state values, regions without electrodes were given a value of approximately 5×10^{-5} at their corresponding diagonal entry in $\mathbf{B}$. This additional input ensured that the calculation of optimal energy was computationally tractable (which is not the case for input applied to a very small control set), but also necessitated the incorporation of $\mathbf{B}$ into the calculation of energy for a more representative estimate.

Trajectories were simulated for each frequency band, and these trajectories were combined into a single state matrix for all analyses, unless otherwise specified (e.g., as in Fig. 5C and in some figures in the Supplementary Materials). More specifically, comparisons of brain state were calculated as the two-dimensional Pearson's correlation coefficient between simulated region-by-frequency matrices and empirical region-by-frequency matrices (Fig. 2). Only regions with electrodes were included in correlations, as they were the only regions with initial state measurements. Energy in all optimal control analyses was calculated in each band independently, and then summarized in a region-by-frequency matrix at each time point (Fig. 3–6). A single measure of energy for a trial was calculated by integrating the Frobenius norm of the energy matrix over time.

Network Statistics

To probe the role of graph architecture in the energy required for optimal control trajectories, we calculated the determinant ratio, which is defined as the ratio of the strength to the homogeneity of the connections between the first degree driver (anything with a non-zero entry in $\mathbf{B}$) and the non-driver (anything with a zero entry in $\mathbf{B}$) [36]. This metric was derived assuming that a system has a greater number of driver

nodes than non-driver nodes, and that the initial and final states are distributed around zero; importantly, both assumptions are accurate for our simulations. Quantitatively, the trade-off between strength and homogeneity is embodied in the ratio between the determinant of the Gram matrix of all driver to non-driver connections, and the determinant of that same matrix with each non-driver node removed iteratively. The gram matrix here is the inner product of the vectors giving connections from driver nodes to and non-driver nodes. More specifically, if C is the Gram matrix of all driver to non-driver connections, and C_k is the matrix of all connections from driver nodes to all but the kth non-driver node, the determinant ratio is defined by $N^{-1} \sum_{k=1}^{N} \frac{det(C_k)}{det(C)}$. Since the calculation of the determinant of large matrices can be computationally challenging, we use the equivalent estimate of the trace of the inverse of the Gram matrix, $\text{Trace}(C^{-1})$, to calculate the average determinant ratio (see Kim et. al. for a full derivation) [36].

To understand the expected differences in stimulation-induced dynamics based on which region is actually being stimulated, we calculated two network control statistics: the *persistent modal controllability* and the *transient modal controllability*. Intuitively, the persistent (transient) controllability is high in nodes where the addition of energy will result in large perturbations to the slow (fast) modes of the system [27]. Typically, modal controllability is computed from the eigenvector matrix $V = [v_{ij}]$ of the adjacency matrix $\mathbf{A}$. The jth mode of the system is poorly controllable from node i if the entry for v_{ij} is small. Modal controllability is then calculated as $\phi_i = \sum_{j=1}^{N}(1 - \lambda_j^2(A))v_{ij}^2$. We adapt this discrete-time estimate to continuous-time by defining modal controllability to be $\phi_i = \sum_{j=1}^{N}(1 - (e^{\lambda_j(A)\delta t})^2)v_{ij}^2$. Here, δt is the time step of the trajectory and $e^{\lambda_j(A)\delta t}$ is the conversion from continuous to discrete eigenvalues of the system. Persistent (transient) modal controllability are computed in the same way, but using only the 10% largest (smallest) eigenvalues of the system. We explore how this cutoff affects estimates of persistent and transient controllability in the Supplement.

Random Forest Model

Random forest models are constructed by averaging predictions over a large number of decision trees (here: 500), where each branch in the tree splits one of the predictors into two groups, the means of which are used as a predicted value for observations in each branch [41]. Splits are selected to reduce prediction error. Random forest models rely on bootstrapping data for each split, and a random selection of the variable to split on to avoid overfitting the data. Out-of-bag mean squared error is calculated as the prediction error of the samples that were not included in bootstrapped selection for each tree, and therefore are samples that the model has not been trained on [41]. For further details, see Supplemental Methods.

Data Sharing

Code for simulations and select metrics is available at https://github.com/jastiso. Data will be made available upon request.

Acknowledgments

Author Contributions

DSB, ANK, and JS designed analyses; JS analyzed data; FP, TM, and JS wrote code; AEK constructed DWI matrices; ANK preprocessed data; FP and TM developed control framework; JS and DSB wrote the manuscript. Other authors were involved in data collection and manuscript editing.

Author Declaration

The authors declare no conflicts of interest.

References

[1] J. L. Andersson and S. N. Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125:1063–1078, 2016.

[2] D. S. Bassett and O. Sporns. Network neuroscience. *Nature Neuroscience*, 20(3):353–364, 2017.

[3] R. F. Betzel, S. Gu, J. D. Medaglia, F. Pasqualetti, and D. S. Bassett. Optimally controlling the human connectome: The role of network topology. *Scientific Reports*, 6:1–14, 2016.

[4] M. Bonnefond, S. Kastner, and O. Jensen. Communication between Brain Areas Based on Nested Oscillations. *Eneuro*, 4(2):ENEURO.0153–16.2017, 2017.

[5] U. Braun, A. Schaefer, R. F. Betzel, H. Tost, A. Meyer-Lindenberg, and D. S. Bassett. From maps to multi-dimensional network mechanisms of mental disorders. *Neuron*, 97(1):14–31, 2018.

[6] G. Buzsáki, C. A. Anastassiou, and C. Koch. The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–420, 2012.

[7] G. Buzsáki and E. I. Moser. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16(2):130–138, 2013.

[8] L. Cammoun, X. Gigandet, D. Meskaldji, J. P. Thiran, O. Sporns, K. Q. Do, P. Maeder, R. Meuli, and P. Hagmann. Mapping the human connectome at multiple scales with diffusion spectrum MRI. *Journal of Neuroscience Methods*, 203(2):386–397, 2012.

[9] R. T. Canolty and R. T. Knight. The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11):506–515, 2010.

[10] Y. Chen, S. Wang, C. C. Hilgetag, and C. Zhou. Trade-off between Multiple Constraints Enables Simultaneous Formation of Modules and Hubs in Neural Systems. *PLoS Comput Biol*, 9(3), 2013.

[11] S. Ching, E. N. Brown, and M. A. Kramer. Distributed control in a mean-field cortical network model: implications for seizure suppression. *Phys Rev E*, 86(2 Pt 1):021920, 2012.

[12] M. Cieslak and S. T. Grafton. Local termination pattern analysis: A tool for comparing white matter morphology. *Brain Imaging and Behavior*, 8(2):292–299, 2014.

[13] S. R. Cole, R. van der Meij, E. J. Peterson, C. de Hemptinne, P. A. Starr, and B. Voytek. Nonsinusoidal Beta Oscillations Reflect Cortical Pathophysiology in Parkinson's Disease. *The Journal of Neuroscience*, 37(18):4830–4840, 2017.

[14] S. P. Cornelius, W. L. Kath, and A. E. Motter. Realistic control of network dynamics. *Nat Commun*, 4:1942, 2013.

24

[15] A. Daducci, S. Gerhard, A. Griffa, A. Lemkaddem, L. Cammoun, X. Gigandet, R. Meuli, P. Hagmann, and J. P. Thiran. The Connectome Mapper: An Open-Source Processing Pipeline to Map Connectomes with MRI. *PLoS ONE*, 7(12), 2012.

[16] J. Downar, J. Geraci, T. V. Salomons, K. Dunlop, S. Wheeler, M. P. McAndrews, N. Bakker, D. M. Blumberger, Z. J. Daskalakis, S. H. Kennedy, A. J. Flint, and P. Giacobbe. Anhedonia and reward-circuit connectivity distinguish nonresponders from responders to dorsomedial prefrontal repetitive transcranial magnetic stimulation in major depression. *Biol Psychiatry*, 76(3):176–185, 2014.

[17] D. Ehrens, D. Sritharan, and S. V. Sarma. Closed-loop control of a fragile network: application to seizure-like dynamics of an epilepsy model. *Front Neurosci*, 9:58, 2015.

[18] Y. Ezzyat, J. E. Kragel, J. F. Burke, D. F. Levy, A. Lyalenko, P. Wanda, L. O'Sullivan, K. B. Hurley, S. Busygin, I. Pedisich, M. R. Sperling, G. A. Worrell, M. T. Kucewicz, K. A. Davis, T. H. Lucas, C. S. Inman, B. C. Lega, B. C. Jobst, S. A. Sheth, K. Zaghloul, M. J. Jutras, J. M. Stein, S. R. Das, R. Gorniak, D. S. Rizzuto, and M. J. Kahana. Direct Brain Stimulation Modulates Encoding States and Memory Performance in Humans. *Current Biology*, 27(9):1251–1258, 2017.

[19] Y. Ezzyat, P. A. Wanda, D. F. Levy, A. Kadel, A. Aka, I. Pedisich, M. R. Sperling, A. D. Sharan, B. C. Lega, A. Burks, R. E. Gross, C. S. Inman, B. C. Jobst, M. A. Gorenstein, K. A. Davis, G. A., J. M. S. Worrell, Michal T. Kucewicz, R. Gorniak, S. R. Das, Daniel, S. Rizzuto, and M. J. Kahana. Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. *Nature Communications*, 2018.

[20] S. Feldt Muldoon, I. Soltesz, and R. Cossart. Spatially clustered neuronal assemblies comprise the microstructure of synchrony in chronically epileptic networks. *Proceedings of the National Academy of Sciences*, 110(9):3567–3572, 2013.

[21] J. Fell, E. Ludowig, B. P. Staresina, T. Wagner, T. Kranz, C. E. Elger, and N. Axmacher. Medial Temporal Theta/Alpha Power Enhancement Precedes Successful Memory Encoding: Evidence Based on Intracranial EEG. *Journal of Neuroscience*, 31(14):5392–5397, 2011.

[22] H. Glowinski. Cognitive deficits in temporal lobe epilepsy: an investigation of memory functioning. *Journal of Nervous and Mental Disease*, 157(2):129–137, aug 1973.

[23] D. S. Grayson, E. Bliss-Moreau, C. J. Machado, J. Bennett, K. Shen, K. A. Grant, D. A. Fair, and D. G. Amaral. The Rhesus Monkey Connectome Predicts Disrupted Functional Networks Resulting from Pharmacogenetic Inactivation of the Amygdala. *Neuron*, 91(2):453–466, 2016.

[24] A. Griffa, P. S. Baumann, J. P. Thiran, and P. Hagmann. Structural connectomics in brain diseases. *NeuroImage*, 80:515–526, 2013.

[25] D. W. Gross, L. Concha, and C. Beaulieu. Extratemporal White Matter Abnormalities in Mesial Temporal Lobe Epilepsy Demonstrated with Diffusion Tensor Imaging. *Epilepsia*, 47(8):1360–1363, aug 2006.

[26] S. Gu, R. F. Betzel, M. G. Mattar, M. Cieslak, P. R. Delio, S. T. Grafton, F. Pasqualetti, and D. S. Bassett. Optimal trajectories of brain state transitions. *NeuroImage*, 148(January):305–317, 2017.

[27] S. Gu, F. Pasqualetti, M. Cieslak, Q. K. Telesford, A. B. Yu, A. E. Kahn, J. D. Medaglia, J. M. Vettel, M. B. Miller, S. T. Grafton, and D. S. Bassett. Controllability of structural brain networks. *Nature Communications*, 6:1–10, 2015.

[28] C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*, 104(24):10240–10245, 2007.

[29] A. Horn, M. Reich, J. Vorwerk, N. Li, G. Wenzel, Q. Fang, T. Schmitz-Hübsch, R. Nickl, A. Kupsch, J. Volkmann, A. A. Kühn, and M. D. Fox. Connectivity Predicts deep brain stimulation outcome in Parkinson disease. *Annals of Neurology*, 82(1):67–78, jul 2017.

[30] J. Jacobs, J. Miller, S. A. Lee, T. Coffey, A. J. Watrous, M. R. Sperling, A. Sharan, G. Worrell, B. Berry, B. Lega, B. C. Jobst, K. Davis, R. E. Gross, S. A. Sheth, Y. Ezzyat, S. R. Das, J. Stein, R. Gorniak, M. J. Kahana, and D. S. Rizzuto. Direct Electrical Stimulation of the Human Entorhinal Region and Hippocampus Impairs Memory. *Neuron*, 92(5):983–990, 2016.

[31] J. Jeganathan, A. Perry, D. S. Bassett, G. Roberts, P. B. Mitchell, and M. Breakspear. Fronto-limbic dysconnectivity leads to impaired brain network controllability in young people with bipolar disorder and those at high genetic risk. *NeuroImage: Clinical*, In Revision, 2018.

[32] M. D. Johnson, H. H. Lim, T. I. Netoff, A. T. Connolly, N. Johnson, A. Roy, A. Holt, K. O. Lim, J. R. Carey, J. L. Vitek, and B. He. Neuromodulation for brain disorders: challenges and opportunities. *IEEE Trans Biomed Eng*, 60(3):610–624, 2013.

[33] S. E. Jones, S. Y. Mahmoud, and M. D. Phillips. A practical clinical method to quantify language lateralization in fMRI using whole-brain analysis. *Neuroimage*, 54(4):2937–2949, 2011.

[34] T. Kailath. *Linear Systems*. Prentice-Hall, 1980.

[35] D. Kim, S. C. Jun, and H. I. Kim. Computational study of subdural and epidural cortical stimulation of the motor cortex. *Conf Proc IEEE Eng Med Biol Soc*, 2011:7226–7229, 2011.

[36] J. Z. Kim, J. M. Soffer, A. E. Kahn, J. M. Vettel, F. Pasqualetti, and D. S. Bassett. Role of graph architecture in controlling dynamical networks with applications to neural systems. *Nature Physics*, 14:91–98, 2018.

[37] M. T. Kucewicz, B. M. Berry, L. R. Miller, F. Khadjevand, Y. Ezzyat, J. M. Stein, V. Kremen, B. H. Brinkmann, P. Wanda, M. R. Sperling, R. Gorniak, K. A. Davis, B. C. Jobst, R. E. Gross, B. Lega, J. Van Gompel, S. M. Stead, D. S. Rizzuto, M. J. Kahana, and G. A. Worrell. Evidence for verbal memory enhancement with electrical brain stimulation in the lateral temporal cortex. *Brain*, Epub Ahead of Print, 2018.

[38] A. W. Laxton, D. F. Tang-Wai, M. P. McAndrews, D. Zumsteg, R. Wennberg, R. Keren, J. Wherrett, G. Naglie, C. Hamani, G. S. Smith, and A. M. Lozano. A phase i trial of deep brain stimulation of memory circuits in Alzheimer's disease. *Annals of Neurology*, 68(4):521–534, 2010.

[39] D. J. Lee, A. Izadi, M. Melnik, S. Seidl, A. Echeverri, K. Shahlaie, and G. G. Gurkoff. Stimulation of the medial septum improves performance in spatial learning following pilocarpine-induced status epilepticus. *Epilepsy Res*, 130:53–63, 2017.

[40] F. Li, E. M. Deshaies, G. Allott, G. Canute, and R. Gorji. Direct cortical stimulation but not transcranial electrical stimulation motor evoked potentials detect brain ischemia during brain tumor resection. *Am J Electroneurodiagnostic Technol*, 51(3):191–197, 2011.

[41] A. Liaw and M. Wiener. Classification nad Regression by RandomForest. *R News*, 2/3:–438, 2002.

[42] Y. Y. Liu, J. J. Slotine, and A. L. Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.

[43] A. M. Lozano and N. Lipsman. Probing and Regulating Dysfunctional Circuits Using Deep Brain Stimulation. *Neuron*, 77(3):406–424, 2013.

[44] J. L. Lujan, A. Chaturvedi, K. S. Choi, P. E. Holtzheimer, R. E. Gross, H. S. Mayberg, and C. C. McIntyre. Tractography-activation models applied to subcallosal cingulate deep brain stimulation. *Brain Stimul*, 6(5):737–739, 2013.

[45] R. C. Malenka and M. F. Bear. Review LTP and LTD: An Embarrassment of Riches useful to conceptualize LTP and LTD as a general class of cellular/synaptic phenomena. Just as different neurons express different complements of ion channels to. *Neuron*, 44:5–21, 2004.

[46] J. Mani, B. Diehl, Z. Piao, S. S. Schuele, E. Lapresto, P. Liu, D. R. Nair, D. S. Dinner, and H. O. Luders. Evidence for a basal temporal visual language center: cortical stimulation producing pure alexia. *Neurology*, 71(20):1621–1627, 2008.

[47] C. C. McIntyre, M. Savasta, B. L. Walter, and J. L. Vitek. How Does Deep Brain Stimulation Work? Present Understanding and Future Questions. *Journal of Clinical Neurophysiology*, 21(1):40–50, 2004.

[48] T. Menara, V. Katewa, D. S. Bassett, and F. Pasqualetti. The Structured Controllability Radius of Symmetric (Brain) Networks. 2018.

[49] S. F. Muldoon, F. Pasqualetti, S. Gu, M. Cieslak, S. T. Grafton, J. M. Vettel, and D. S. Bassett. Stimulation-Based Control of Dynamic Brain Networks. *PLoS Computational Biology*, 12(9), 2016.

[50] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[51] J. Parvizi, C. Jacques, B. L. Foster, N. Witthoft, V. Rangarajan, K. S. Weiner, and K. Grill-Spector. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J Neurosci*, 32(43)):14915–14920, 2012.

[52] J. Parvizi and S. Kastner. Human intracranial EEG: Promises and Limitations. *Nature Neuroscience*, 2017.

[53] F. Pasqualetti, S. Zampieri, and F. Bullo. Controllability metrics, limitations and algorithms for complex networks. *IEEE Transactions on Control of Network Systems*, 1(1):40–52, 2014.

[54] J. S. Perlmutter and J. W. Mink. Deep Brain Stimulation. *Annu Rev Neurosci*, 2006.

[55] E. J. Peterson and B. Voytek. Alpha oscillations control cortical gain by modulating excitatory-inhibitory background activity. *bioRxiv*, pages 1–25, 2017.

[56] A. Reichenbach, K. Whittingstall, and A. Thielscher. Effects of transcranial magnetic stimulation on visual evoked potentials in a visual suppression task. *Neuroimage*, 54(2):1375–1384, 2011.

[57] P. Riva-Posse, K. S. Choi, P. E. Holtzheimer, C. C. McIntyre, R. E. Gross, A. Chaturvedi, A. L. Crowell, S. J. Garlow, J. K. Rajendra, and H. S. Mayberg. Defining critical white matter pathways mediating successful subcallosal cingulate deep brain stimulation for treatment-resistant depression. *Biol Psychiatry*, 76(12):963–969, 2014.

[58] P. Riva-Posse, P. E. Holtzheimer, S. J. Garlow, and H. S. Mayberg. Practical considerations in the development and refinement of subcallosal cingulate white matter deep brain stimulation for treatment-resistant depression. *World Neurosurg*, 80(3-4):S27.e25–34, 2013.

[59] J. A. Roberts, A. Perry, A. R. Lord, G. Roberts, P. B. Mitchell, R. E. Smith, F. Calamante, and M. Breakspear. The contribution of geometry to the human connectome. *NeuroImage*, 124:379–393, 2016.

[60] L. Rossini, R. Garbelli, V. Gnatkovsky, G. Didato, F. Villani, R. Spreafico, F. Deleo, G. Lo Russo, G. Tringali, F. Gozzo, L. Tassi, and M. de Curtis. Seizure activity per se does not induce tissue damage markers in human neocortical focal epilepsy. *Annals of Neurology*, 82(3):331–341, 2017.

[61] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

[62] G. Schalk, J. Marple, R. T. Knight, and W. G. Coon. Instantaneous voltage as an alternative to power- and phase-based interpretation of oscillatory brain activity. *NeuroImage*, 157(May):545–554, 2017.

[63] S. J. Schiff. *Neural Control Engineering: The Emerging Intersection between Control Theory and Neuroscience*. MIT Press, 2011.

[64] A. Shimotake, R. Matsumoto, T. Ueno, T. Kunieda, S. Saito, P. Hoffman, T. Kikuchi, H. Fukuyama, S. Miyamoto, R. Takahashi, A. Ikeda, and R. M. A. Lambon. Direct exploration of the role of the ventral anterior temporal lobe in semantic memory: Cortical stimulation and local field potential evidence from subdural grid electrodes. *Cereb Cortex*, 10:3802–3817, 2015.

[65] V. A. Sironi. Origin and Evolution of Deep Brain Stimulation. *Frontiers in Integrative Neuroscience*, 5:42, 2011.

[66] N. Suthana, Z. Haneef, J. Stern, R. Mukamel, E. Behnke, B. Knowlton, and I. Fried. Memory Enhancement and Deep-Brain Stimulation of the Entorhinal Area. *N Engl J Med*, 366:502–10, 2012.

[67] E. Tang and D. S. Bassett. Control of dynamics in brain networks. *arXiv*, 1701:01531, 2017.

[68] E. Tang, C. Giusti, G. L. Baum, S. Gu, E. Pollock, A. E. Kahn, D. R. Roalf, T. M. Moore, K. Ruparel, R. C. Gur, R. E. Gur, T. D. Satterthwaite, and D. S. Bassett. Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nature Communications*, 8(1), 2017.

[69] P. N. Taylor, J. Thomas, N. Sinha, J. Dauwels, M. Kaiser, T. Thesen, and J. Ruths. Optimal control based seizure abatement using patient derived connectivity. *Frontiers in Neuroscience*, 9(MAY), 2015.

[70] C. Thomas, F. Q. Ye, M. O. Irfanoglu, P. Modi, K. S. Saleem, D. A. Leopold, and C. Pierpaoli. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proceedings of the National Academy of Sciences*, 111(46):16574–16579, 2014.

[71] S. Toprani and D. M. Durand. Fiber tract stimulation can reduce epileptiform activity in an in-vitro bilateral hippocampal slice preparation. *Exp Neurol*, 240:28–43, 2013.

[72] M. Vinck, R. Oostenveld, M. Van Wingerden, F. Battaglia, and C. M. Pennartz. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, 55(4):1548–1565, 2011.

[73] E. Wu-Yan, R. F. Betzel, E. Tang, S. Gu, F. Pasqualetti, and D. S. Bassett. Benchmarking measures of network controllability on canonical graph models. 2017.

[74] G. Yan, P. E. Vértes, E. K. Towlson, Y. L. Chew, D. S. Walker, W. R. Schafer, and A. L. Barabási. Network control principles predict neuron function in the Caenorhabditis elegans connectome. *Nature*, 550(7677):519–523, 2017.

[75] F. C. Yeh, T. D. Verstynen, Y. Wang, J. C. Fernández-Miranda, and W. Y. I. Tseng. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS ONE*, 8(11), 2013.

[76] N. Yousif and X. Liu. Investigating the depth electrode-brain interface in deep brain stimulation using finite element models with graded complexity in structure and solution. *J Neurosci Methods*, 184(1):142–151, 2009.

[77] J. G. T. Zanudo, G. Yang, and R. Albert. Structure-based control of complex networks with nonlinear dynamics. *Proc Natl Acad Sci U S A*, 114(28):7234–7239, 2017.