Inferring Neural Communication Dynamics from Field Potentials Using Graph Diffusion Autoregression

Felix Schwock[1,3*], Julien Bloch[2,3], Karam Khateeb[2,3], Jasmine Zhou[2,3], Les Atlas[1], Azadeh Yazdan-Shahmorad[1,2,3*]

Estimating dynamic network communication is attracting increased attention, spurred by rapid advancements in multi-site neural recording technologies and efforts to better understand cognitive processes. Yet, traditional methods, which infer communication from statistical dependencies among distributed neural recordings, face core limitations: they do not incorporate possible mechanisms of neural communication, neglect spatial information from the recording setup, and yield predominantly static estimates that cannot capture rapid changes in the brain. To address these issues, we introduce the graph diffusion autoregressive model. Designed for distributed field potential recordings, our model combines vector autoregression with a network communication process to produce a high-resolution communication signal. We successfully validated the model on simulated neural activity and recordings from subdural and intracortical micro-electrode arrays placed in macaque sensorimotor cortex demonstrating its ability to describe rapid communication dynamics induced by optogenetic stimulation, changes in resting state communication, and neural correlates of behavior during a reach task.

Introduction

The coordinated interactions across different brain networks and subnetworks underlies cognitive processes[1–6], and disruptions of these interactions are linked to a range of neurological disorders[7–10]. Despite this demonstrated importance, we still do not fully understand how brain networks perform computations through the coordinated signaling of connected neurons and neural populations during natural behavior, following a disease or injury, or as the result of rehabilitative intervention. The development of new electrophysiological recording technologies such as large-scale micro-electrode arrays provides unique opportunities for measuring brain network activity simultaneously over multiple areas with high spatial and temporal resolution[11–16].

A common signal extracted from subdural and intracortical micro-electrode arrays is the local field potential (LFP), which describes voltage fluctuations in the extracellular space of neuronal tissue. For these signals, the most common approach for estimating neural communication is through functional connectivity (FC) analysis[17,18]. In general, FC measures define neural communication as the undirected (symmetric) or directed (asymmetric) statistical dependence between different measurements that can be inferred directly from data using either model-free approaches or very general model classes such as vector autoregressive (VAR) models[19–22]. While these techniques are a popular choice for electrophysiology analysis, they predominantly yield static estimates of neural communication. Additionally, they rarely incorporate information about the structural network connectivity of the underlying brain region, particularly when analyzing

[1]Department of Electrical and Computer Engineering, University of Washington, Seattle, WA, USA. [2]Department of Bioengineering, University of Washington, Seattle, WA, USA. Washington National [3]Primate Research Center, Seattle, WA, USA. *e-mail: fschwock@uw.edu; azadehy@uw.edu

recordings from high-resolution intracranial electrophysiology arrays. Lastly, most FC metrics stem from general-purpose statistical methods that have found widespread use across many scientific disciplines, but critically lack mechanistic assumptions relevant to modeling neural communication.

In contrast to heavily data-driven FC analysis, neural interactions can also be modeled using tools from dynamical systems theory that incorporate knowledge about the mechanisms through which different neural populations interact[23,24]. For field potentials, a popular technique are neural field models that use a combination of differential equations to model temporal dynamics and integrals to incorporate spatial interactions[23]. Such models can generate neural dynamics that match empirical observations, such as the wave-like propagation of oscillatory activity observed in the sensorimotor cortex[25,26]. Furthermore, these ideas have been extended to model the flow of information across brain networks for example via structurally guided diffusion[27]. However, these models are typically not used in a data-driven framework, where functional interactions are directly inferred from measured neural activity.

Here we propose a new technique for estimating dynamic neural communication that 1) naturally incorporates the spatial layout of the recording array and the local connectivity structure of the cortex[28] as a structural prior, 2) integrates a mechanism of neural communication into a data-driven FC model, and 3) produces a highly dynamic information flow signal that can be used to study transient network events. Specifically, we combine the classical autoregressive framework for the treatment of temporal dynamics with the graph Laplacian of a predefined structural connectivity graph to incorporate network interactions[29]. Because the graph Laplacian is commonly used to model diffusion processes on networks[30], we refer to our approach as the graph diffusion autoregressive (GDAR) model. To the best of our knowledge, the GDAR model is the first approach to integrate the above three aspects – structural priors, a mechanism of neural communication, and a highly dynamic information flow signal – into a single data-driven model.

To demonstrate the utility of our framework, we tested the GDAR model on five, highly diverse datasets. First, using synthetic data from various networks of Wilson-Cowan oscillators we demonstrate that the high-resolution communication signal estimated by our model aligns with the simulated interactions more accurately than standard VAR models. Next, using three micro-electrocorticography (μECoG) and one Utah array dataset we demonstrate that the GDAR model can be used to uncover transient communication dynamics evoked *during* cortical optogenetic stimulation, uncover neural correlates of a monkey's reach behavior that are dependent on the spatial frequency, and analyze changes in resting state neural communication *after* electrical stimulation. We show that the GDAR model outperforms standard VAR models and other FC measures and provides insights that cannot be obtained by other models. Finally, we show that the GDAR model better generalizes to unseen data than VAR models.

Results

Graph diffusion autoregressive (GDAR) model. An overview of the GDAR model is shown in Fig. 1 and a more detailed mathematical description can be found in Methods. First, the electrode layout of the recording array is used to construct a sparse and locally connected graph, with each electrode representing a node and with edges connecting nearby nodes Fig. 1a left). This graph serves as a structural prior that incorporates information about the local connectivity of

the cortex into the model[28]. By modeling the spatiotemporal evolution of neural activity observed at the nodes of the graph as a parameterized diffusion process, the GDAR model transforms the neural activity into a directed communication or information flow signal defined on the graph edges (Fig. 1a right). This communication signal, which we will refer to as *GDAR flow*, describes the moment to moment signaling between the nodes. Unlike classical functional connectivity analysis, which aggregates information over multiple time points thereby estimating an average information flow, the GDAR model transforms the neural activity at each time point into a flow signal without losing temporal resolution. Therefore, it can be used to study transient communication events in the brain.

As with VAR models, the GDAR model can be formalized as a predictive model with order p, which describes the number of lags used for predicting future neural activity. A model order of $p = 1$ describes a classic graph diffusion process, where temporal changes in neural activity are driven by the discrete approximation of the surface Laplacian, i.e., the second spatial derivative (see Eq. (3) in Methods). Increasing the model order increases the capacity of the model and adds "memory" to the diffusion process, thus offering the increased ability to model complex spatiotemporal neural dynamics. An overview of the p^{th} order model is shown in Fig. 1b and c. The neural activity at time t at each node $s_i[t]$ is modeled using a linear combination of its own p past samples plus the time varying GDAR flow from all neighboring nodes:

$$s_i[t] = \sum_{k=1}^{p} m_k^{\{i\}} s_i[t-k] + \sum_{j \in \mathcal{N}(i)} f^{\{i,j\}}[t]. \qquad (1)$$

The GDAR flow $f^{\{i,j\}}$ between node i and j is given by a linear combination of the p past activity gradients between the two nodes (Fig. 1c; see Methods for more details)

$$f^{\{i,j\}}[t] = \sum_{k=1}^{p} w_k^{\{i,j\}} \left(s_j[t-k] - s_i[t-k]\right) \qquad (2)$$

and can be positive or negative, depending on whether information flows into or out of node i. The node and edge parameters of the GDAR model $m_k^{\{i\}}$ and $w_k^{\{i,j\}}$ can be estimated from neural recordings using linear regression (see Methods) and are assumed to be static within a predefined time window.

Computing the second spatial derivative is equivalent to computing the current source densities (CSDs), which is a popular technique for analyzing field potential recordings obtained from technologies such as ECoG or electroencephalography (EEG)[31,32]. Therefore, the GDAR model can also be considered a combination of CSD analysis and VAR model. For field potential recordings the model parameters $w_1^{\{i,j\}}, \ldots, w_p^{\{i,j\}}$ can be interpreted as conductivities such that voltage gradients multiplied by conductivity yields current flow. Summing the current flows at each node is analogous to computing the current sources and sinks in CSD analysis.

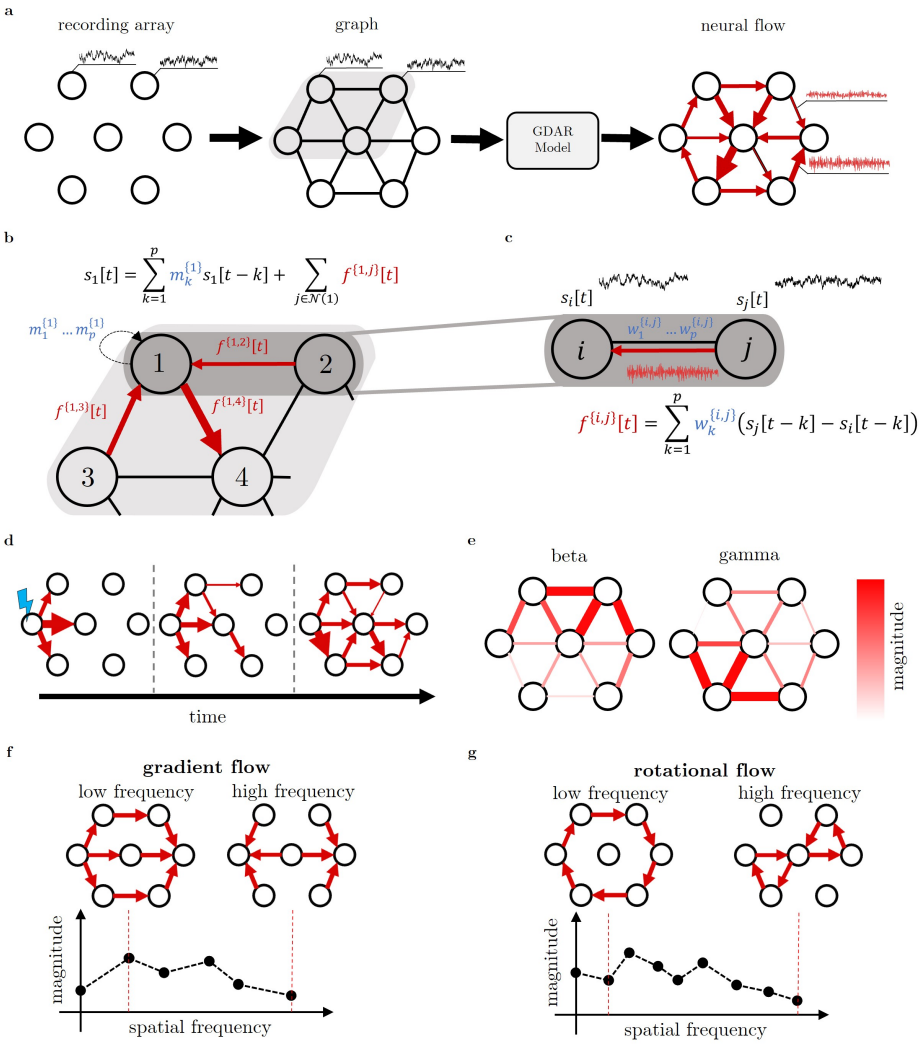


Fig. 1: Overview of the graph diffusion autoregressive (GDAR) model. (a) The recording array is used to form a sparse, locally connected graph, where each electrode represents a node, and edges connect neighboring nodes. The GDAR model then transforms the neural activity observed at the nodes into a directed flow signal defined on the graph edges, representing the real-time signaling between nodes. (b) The model incorporates an order p autoregressive system, where at time t each node's neural activity is modeled using a combination of its past p samples and the flow from all adjacent nodes. (c) The directed GDAR flow at time t between node i and j, denoted as $f^{\{i,j\}}[t]$, is calculated as the weighted sum of the previous p activity gradients between these nodes. In analogy to current source density analysis, the edge parameters $w_1^{\{i,j\}}, \ldots, w_p^{\{i,j\}}$ can be interpreted as conductivities for local field potential measurements, such that conductivity times potential gradient yields a current flow. The model parameters are assumed to be static within a particular time window and can be estimated using linear regression (see Methods). (d) The GDAR flow can be used to study transient communication events, for example due to cortical stimulation. (e) For resting state recordings, power spectral density estimates of the GDAR flow signal can be used to study frequency band specific communication patterns. (f) and (g) Akin to classical Fourier analysis for time series, the GDAR flow signal can also be decomposed into gradient (directional) and rotational flow modes of different spatial frequency to study the smoothness and spatial composition of the flow signal across the network.

The high temporal resolution of the GDAR flow signal $f^{\{i,j\}}[t]$ is ideal to study transient signaling events. For example, the propagation of neural activity due to cortical stimulation can be tracked by concatenating consecutive time steps of $f^{\{i,j\}}[t]$ and analyzing its spatiotemporal evolution

(Fig. 1d). Alternatively, the model can be applied to resting state recording in which case it may be reasonable to compute the power spectrum of $f^{\{i,j\}}[t]$. This results in a similar frequency decomposition as is typical for VAR based FC measures (Fig. 1e) with the distinction that the GDAR flow power spectrum is determined by magnitude and phase differences between neighboring channels that are modulated by the Fourier transform of the model parameters (see Methods) whereas VAR based FC measures only utilize the model parameters for the spectral representation.

Furthermore, modeling neural communication on top of a graph allows us to use recently developed theory from signal processing over simplicial complexes[33–35] to decompose the GDAR flow signal into gradient (directional) and rotational modes of different spatial frequencies (Fig. 1f and g and Methods). The resulting gradient and rotational flow spectra can be used to quantify the degree of smoothness or coordination of the neural signaling (e.g., flow spectra with stronger low-frequency components are considered to represent more coordinated signaling).

The GDAR model outperforms VAR models in inferring communication dynamics in a network of coupled Wilson-Cowan oscillators: To assess the accuracy of the GDAR flow, we fit the model to simulated data generated by 10 randomly connected 16-node networks of coupled Wilson-Cowan oscillators (Fig. 2 and Extended Fig. 1). The networks are used to generate a ground truth neural flow signal, as well as simulated neural activity, which is used to fit GDAR and VAR models of varying model orders (see Methods for more details). In contrast to the GDAR model, VAR models assume no structural connectivity and may find communication links between any pair of nodes in the network even if these nodes are not directly connected. Therefore, we also compare the GDAR model to a VAR model with access to the ground truth structural connectivity network, denoted as enhanced VAR (eVAR) model (see Methods). GDAR and eVAR model only differ in the aspect that the neural flow for the latter one is not driven by spatial activity gradients, but rather by the neural activity itself. All models are used to transform the simulated neural activity into a neural flow signal, which is compared to the ground truth neural flow using various metrics (Fig. 2a). Furthermore, we estimate the neural flow using the CSD approach and compare it to the ground truth flow.

First, we compare the Pearson correlation coefficients (PCCs) between ground truth and estimated flow from GDAR, VAR, and eVAR models, as well as the CSD approach and found that the GDAR model significantly outperforms all other models for model orders $p \geq 12$ (Wilcoxon rank-sum test, $p < 0.05$), thus providing the most accurate estimate of the true neural flow dynamics. (Fig. 2b). We found the same result holds on two additional network structures – a 7-node locally connected graph and a 16-node grid graph that has a connectivity structure similar to the one we assume for our electrophysiology datasets below (Extended Fig. 1).

Despite the superior performance of the GDAR over the competing models, the amount of correlation between estimated and ground truth flow is relatively low for all models. In principle, this can arise from amplitude and phase mismatches between the signals. We investigated this by transforming the ground truth and estimated flow into the frequency domain followed by computing magnitude and phase differences, as well as correlation coefficients between the

magnitude spectra. We found that the GDAR model exhibits consistently lower magnitude errors than the VAR and eVAR models for all frequencies and lower phase errors for frequencies below 50 Hz (Fig. 2c). Furthermore, for all models and model orders, correlations between the estimated and ground truth magnitude spectra were significantly higher compared to those between their corresponding time series (Fig. 2d). This suggests that all models have the capacity to accurately capture magnitude features of the flow signal.

A notable observation is that the spectral magnitude of the ground truth flow is well approximated by lower order VAR, eVAR, and GDAR models. Specifically, the median correlation between estimated and ground truth spectral magnitude for VAR and eVAR models decreases with increasing model order. In contrast, the correlation for the GDAR model reaches another local maximum at higher orders, where it significantly outperforms the other two models (Fig. 2d). Despite this second maximum, the highest median correlations for the GDAR model still occur at low model orders, which contrasts with the results shown in Fig. 2b and suggest that low-order models are sufficient for approximating parts of the communication dynamics. At the same time, because low-order models rely on fewer past time steps for predicting future activity, they may have limited capacity to capture more complex spatiotemporal dynamics that are not fully reflected in spectral magnitude alone. To quantify the ability of our model to capture complex spatiotemporal dynamics, we used a recently developed tool from dynamical systems theory, called dynamical similarity analysis (DSA)[36], which uses dynamical mode decomposition and shape analysis to provide a dissimilarity score between two (high-dimensional) time series (see Methods). Indeed, we found that with increasing model order the accuracy in capturing the spatiotemporal dynamics improves for all models (decreased DSA dissimilarity score) before plateauing at an order around $p = 26$ (Fig. 2e). Hence, to accurately capture the complex dynamic properties of the flow signal, higher order models are needed. For these higher orders, the GDAR model significantly outperforms the VAR and eVAR models in terms of correlations between estimated and ground truth flow in both time and frequency domain (Fig. 2b, and d).

Application of GDAR model to electrophysiological recordings: To show the versatility of the GDAR model to analyze communication dynamics evoked by cortical stimulation, during behavior, and at rest, we have applied the model to electrophysiological recordings from four separate experiments that either use a μECoG array (3 datasets) or a Utah array (1 dataset) to record LFPs from the sensorimotor cortex of macaques. For all datasets, the layout of the recording array was first used to construct a locally connected and sparse graph, where each node corresponds to a recording channels (see Fig. 3 -Fig. 5). Next, GDAR models of different orders were fit to the recorded LFPs. The resulting model coefficients were then used to transforms the LFPs into GDAR flow signals, which were post-processed depending on the experimental setup.

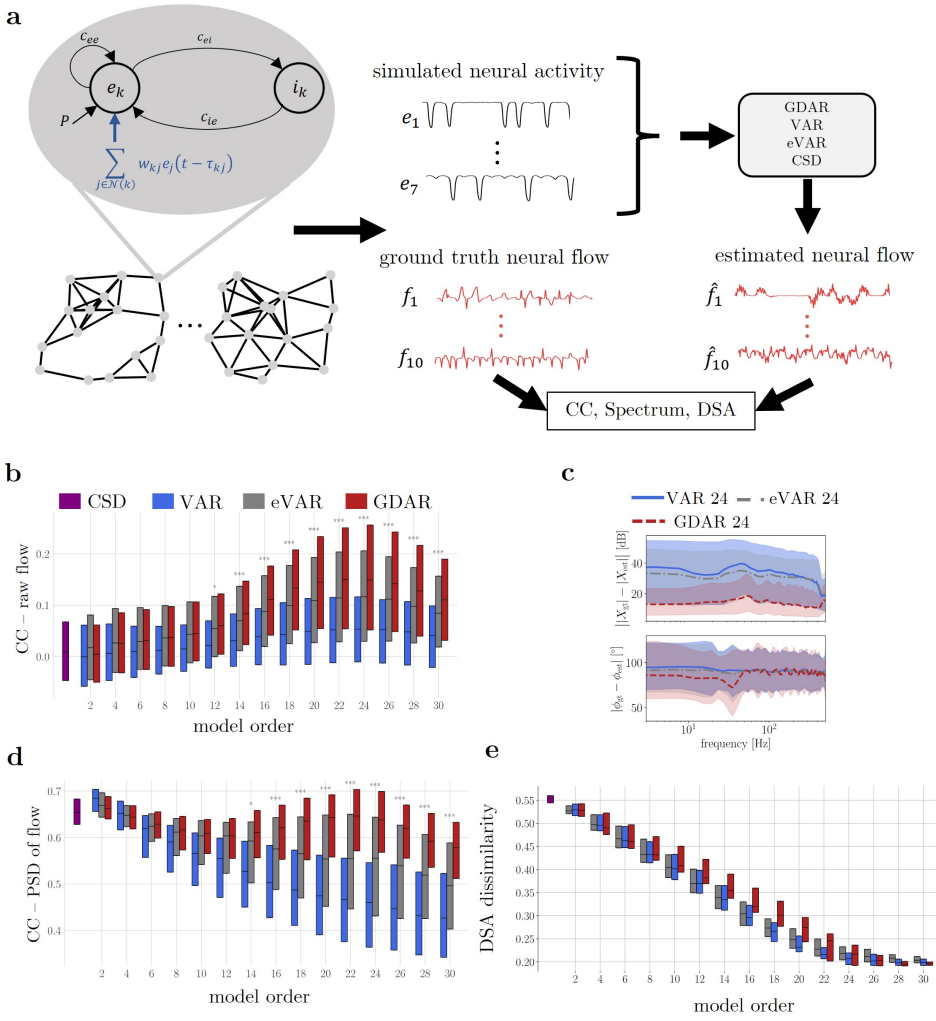


Fig. 2: Evaluation of the GDAR model's accuracy in capturing neural communication dynamics using networks of Wilson Cowan oscillators. (a) 10 randomly connected 16-node toy networks were used to simulate neural activity at each network node as well as compute a ground truth flow across each edge (see Methods). The estimated neural activity is transformed into the estimated neural flow signal using GDAR, VAR, and eVAR (VAR model with knowledge of the structural connectivity) models of different model orders, as well as the CSD approach. Ground truth and estimated neural flow are then compared using various metrics. (b) Distribution (medians, upper and lower quartiles) of the Pearson correlation coefficient (CC) using data from 100 independent simulation trials (10 per network) pooled over all graph edges between ground truth and estimated neural flow for varying model orders. The GDAR model significantly outperforms all other models for orders $p \geq 12$*, thus providing the most accurate overall estimate of the ground truth flow (c) Magnitude and phase difference between the spectrum of ground truth and estimated flow (median, upper and lower quartile). The GDAR model shows consistently lower magnitude errors for all frequencies and phase errors below 50 Hz. (d) Same as in (b) but now comparing the power spectral density (PSD) of the estimated and ground truth flow. The GDAR model again significantly outperforms the other models for higher model orders. (e) Dissimilarity scores between estimated and ground truth flow signals obtained via dynamical similarity analysis (DSA) to assess the accuracy of the estimated flow dynamics. Low dissimilarity scores for high model orders* ($p \geq 26$) *suggest an accurate estimation of the flow dynamics by all three models. All statistical tests use Wilcoxon rank-sum tests at a significance level* $p \leq 0.001$*. Significance markers compare GDAR with eVAR model. Exact p-values can be found in Supplementary Table 1.*

The GDAR model uncovers communication dynamics evoked by cortical optogenetic stimulation: First, we show that the GDAR model can uncover fast, stimulation induced communication dynamics that match the experimental protocol. To do so we use three sessions from an optogenetic stimulation experiment performed in macaques, where two lasers repeatedly stimulated different locations of the primary motor (M1) and somatosensory (S1) cortex expressing the opsin C1V1, and fit a 5^{th} order GDAR model to the LFPs recorded by a 96-channel μECoG array during stimulation (see Fig. 3a-c and Methods)[11,37,38]. GDAR flow signals averaged over all stimulation trials in the milliseconds before and after stimulation for Session 1 are shown in Fig. 3d and Supplementary Video 1. When the network is at rest, flow levels across the network are small. Activation by the first laser located in M1 causes the GDAR flow to immediately increase near the stimulation location before spreading further into the network and reaching S1. After the second laser was activated, the flow increases near the second stimulation location and spreads into most parts of the network within the next few milliseconds.

It is apparent from Fig. 3d that the GDAR flow exhibits complex spatiotemporal dynamics within milliseconds after stimulation. To test how these dynamics depend on the stimulation pattern, we project the high-dimensional flow signal from three sessions, which only differ in their stimulation location (see Fig. 3b), onto their first two principal components (PCs) and compare the flow dynamics in this lower dimensional subspace (see Methods). We found that these low dimensional communication dynamics are very consistent within each session and strongly differ between sessions (Fig. 3e). Furthermore, the communication dynamics show some remarkable similarities with the stimulation patterns. For Sessions 1 and 2, where the flow trajectories largely align in the PC space, the stimulation patterns are similar in that the second stimulation occurs to the top right of the first stimulation. On the other hand, for Session 3, which results in flow trajectories orthogonal to Session 1 and 2, the second stimulation occurs to the top left of the first stimulation. Furthermore, the magnitude of the PC reduced GDAR flow dynamics is noticeably smaller for Session 2 compared to Sessions 1 and 3. This might be a result of the spatial separation between Laser 1 and 2, which is smallest for Session 2. Our findings extend previous work showing that LFP power in monkeys and humans distinctly depend on stimulation parameters such as amplitude and frequency[39,40].

We also tested whether the VAR or CSD flow, or node signal like raw LFPs or its second spatial derivative, which resembles traditional current source density, can uncover dynamics that depend on the stimulation pattern but found that this is not the case (Fig. 3f and g, and Extended Fig. 2). Perhaps it is not surprising that the raw LFPs or simple, model-free transformations thereof (CSD flow, second spatial derivative of raw LFP) fail to describe stimulation dependent dynamics using PC analysis as these signals may be dominated by noise and non-stimulation specific variation. Autoregressive models on the other hand may effectively filter out some of these non-stimulation noise sources. Our results suggest that the GDAR model is more effective at uncovering such transient stimulation-dependent communication dynamics compared to standard VAR models. We also note that the dependence on the stimulation location can be observed when

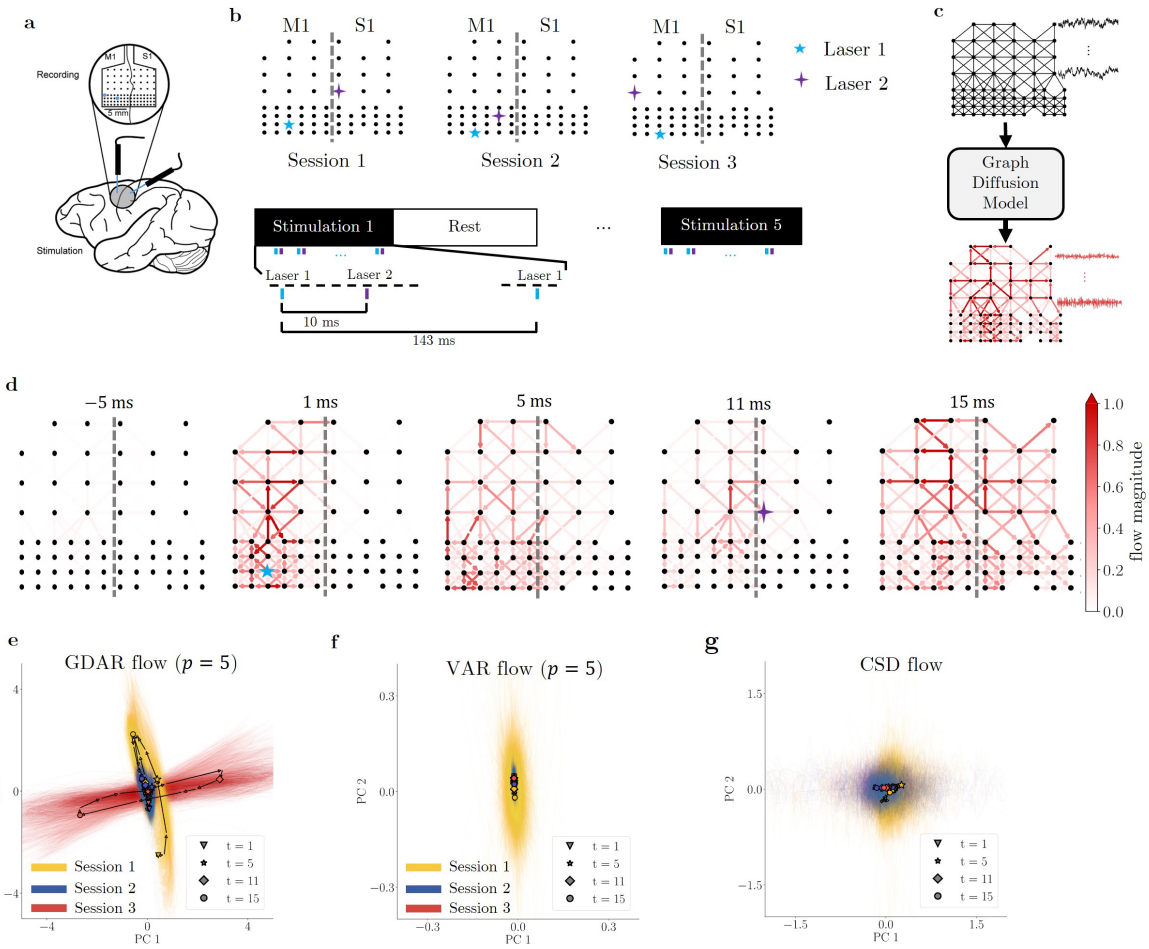


Fig. 3: GDAR model applied to optogenetic stimulation experiment to study transient communication events. (a) LFPs from the primary motor (M1) and somatosensory (S1) cortex of a signal non-human primate were recorded using a 96-channel micro-ECoG array, while repeated paired stimulation was performed using two lasers (modified from Bloch et al.[34]). (b) The relative positions of the electrodes after rejecting bad channels and the locations of the two lasers are shown at the top for the three sessions that were analyzed in this work. The location of the sulcus between M1 and S1 is approximated by the thick gray line. The electrode array was not moved between the sessions. At the bottom, the stimulation protocol is shown. Each laser stimulates alternatingly for 5 ms, with a 10 ms delay between stimulation by Laser 1 and 2. This paired stimulation is repeated every 143 ms. Each stimulation block lasts approximately 7 min and is intermitted shorter long resting blogs during which no stimulation is performed. (c) The recording array is used to construct a sparsely connected graph and the recorded LFPs are then transformed into a flow signal using a 5^{th} order GDAR model. (d) The GDAR flow for Session 1 averaged over all stimulation blocks and trials is shown for different time steps before (first plot) and after (remaining four plots) onset of stimulation from the first laser. The graphs suggest complex spatiotemporal signaling patterns evoked by cortical stimulation. (e) Flow snapshots from the first 25 ms after onset of the first laser stimulation for all trials, blocks, and sessions were stacked into a single matrix and the flow snapshots were projected onto its first two principal components (PCs). The PC reduced GDAR flow trajectories for different sessions are indicated by different colors. Average trajectories are shown as black solid lines with markers indicating different times point after the onset of stimulation by the first laser. Thin colored lines show trajectories by individual paired pulse trials. The plot highlights that GDAR flow trajectories are very consistent within and distinct between sessions, demonstrating that transient communication dynamics depend on the stimulation parameters. (f), (g) PC reduced flow trajectories similar to (e) but using a 5^{th} order VAR model and the CSD approach. In contrast to the GDAR flow, VAR and CSD flow do not exhibit significant time and session dependent dynamics, thus, highlighting the utility of the GDAR model in stimulation induced transient communication dynamics.

plotting a low-dimensional representation of the model parameters itself, where the GDAR model shows a stronger separation between sessions than the VAR model (Extended Fig. 2). Finally, the model can also be adapted to model longer signal propagation paths between specific nodes in the network as it would be reasonable to assume for connections across the sulcus between M1 and S1 (see Extended Fig. 2, Supplementary Video 2, and Supplementary Note).

The GDAR model can track changes in resting state neural communication that are consistent across experiments and recording modalities: Like classical FC analysis, the GDAR model can be utilized to study frequency specific changes in neural communication from resting state recordings. We demonstrate this using two distinct electrical stimulation experiments that employ either intracortical recordings via a 96-channel microelectrode array (Utah array) or subdural recordings via two 32-channel ECoG arrays (Fig. 4 and Fig. 5). For both experiments, repeated electrical stimulation of the macaque sensorimotor cortex is performed for 10-minute blocks at a time either at a single site or alternatingly at two sites (paired-stim), and resting state neural activity is recorded before, after, and between the stimulation blocks. Changes in gamma (30-70 Hz) flow power due to stimulation are estimated via 10^{th} order GDAR models and tested for statistical significance (significance level $p \leq 0.01$; two-sample Kolmogorov-Smirnov test) (see Methods).

For the Utah array data we analyzed recordings from three separate sessions that employ either single site or paired stimulation. For single site stimulation, we observe an increase in GDAR flow power proximal to the stimulation site (Fig. 4b, left). For paired stimulations, a localized increase was only observed near the first stimulation site (Fig. 4b, middle and right), with no notable changes near the second site. We quantified this increase by computing the average flow magnitude over all edges connected to the stimulation site and adjusted it for changes in LFP power (Fig. 4c-e; also see Methods and Extended Fig. 3). Across all sessions, stimulation led to a sustained increase in resting-state communication near the first stimulation site above baseline levels. Notably, for paired stimulation sessions, this augmented communication persisted for at least 17 minutes following the final stimulation block.

For the ECoG data we analyzed changes in resting state neural communication due to single-site electrical stimulation performed during the acute phase after focal ischemic lesioning of the macaque sensorimotor cortex (Fig. 5a and b). In the ipsilesional hemisphere, gamma GDAR flow power increases locally near the stimulation site and decreases across other parts of the network (Fig. 5c). In the contralesional hemisphere, this effect is weakly mirrored, as we observe an increase in GDAR flow power for some edges in the area corresponding to stimulation in the ipsilesional hemisphere but not across the entire network. Similar patterns can be observed for changes in CSD flow power. On the other hand, VAR flow power suggest increased communication within both hemispheres, showing that the GDAR model can describe changes in resting state communication that are not captured by the VAR model. The changes in GDAR flow power from the ECoG dataset are also consistent with our finding from the Utah array dataset, underscoring the GDAR model's ability to robustly capture neural interactions across different experimental and recording modalities.

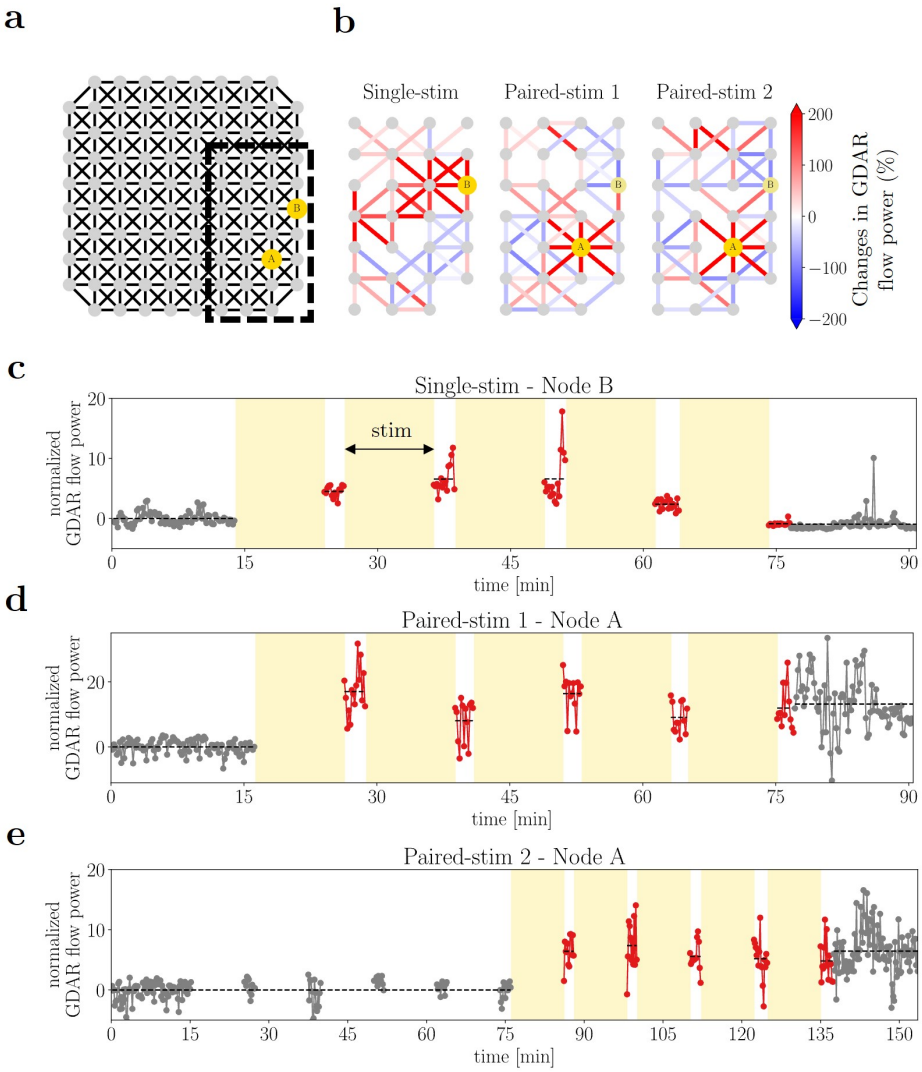


Fig. 4: Application of GDAR model to LFP data recorded from a macaque monkey using a Utah array during a paired electrical stimulation experiment. (a) The locations of the electrodes were used to construct a locally connected sparse graph as input to the GDAR model. Electrode A and B are used for single site and paired stimulation. For paired stimulation, electrode A stimulates before electrode B. (b) Changes in gamma (30-70 Hz) GDAR flow power due to stimulation are shown for three separate sessions. For the single-stim session, only electrode B stimulates. For the paired-stim sessions, electrode A and B stimulate repeatedly and alternatingly for a total of 50 minutes divided into five 10min blocks (see Methods for more details on the stimulation protocol). An increase in gamma GDAR flow power near the first stimulation location can be observed for all sessions. (c)-(e) Temporal evolution of the normalized gamma GDAR flow power averaged over all edges adjacent to the first stimulation site for all three sessions. The 2-minute blocks immediately after stimulation are highlighted in red. The gamma GDAR flow power has been corrected for linear changes in LFP power at the stimulation site (see Methods). For the paired stimulation sessions, the GDAR flow power remains elevated even after the stimulation period ends.

Finally, for the ECoG dataset we have compared our results with changes of three traditional FC measures: gamma coherence, gamma partial directed coherence (PDC), and gamma directed transfer function (DTF) (Fig. 5d). Note that in contrast to the GDAR model and CSD approach, coherence and VAR based measures assumes fully connected graphs, which typically results in much denser communication networks. In the ipsilesional hemisphere we found that PDC and DTF most closely agree with the results from the GDAR model, with the main difference that

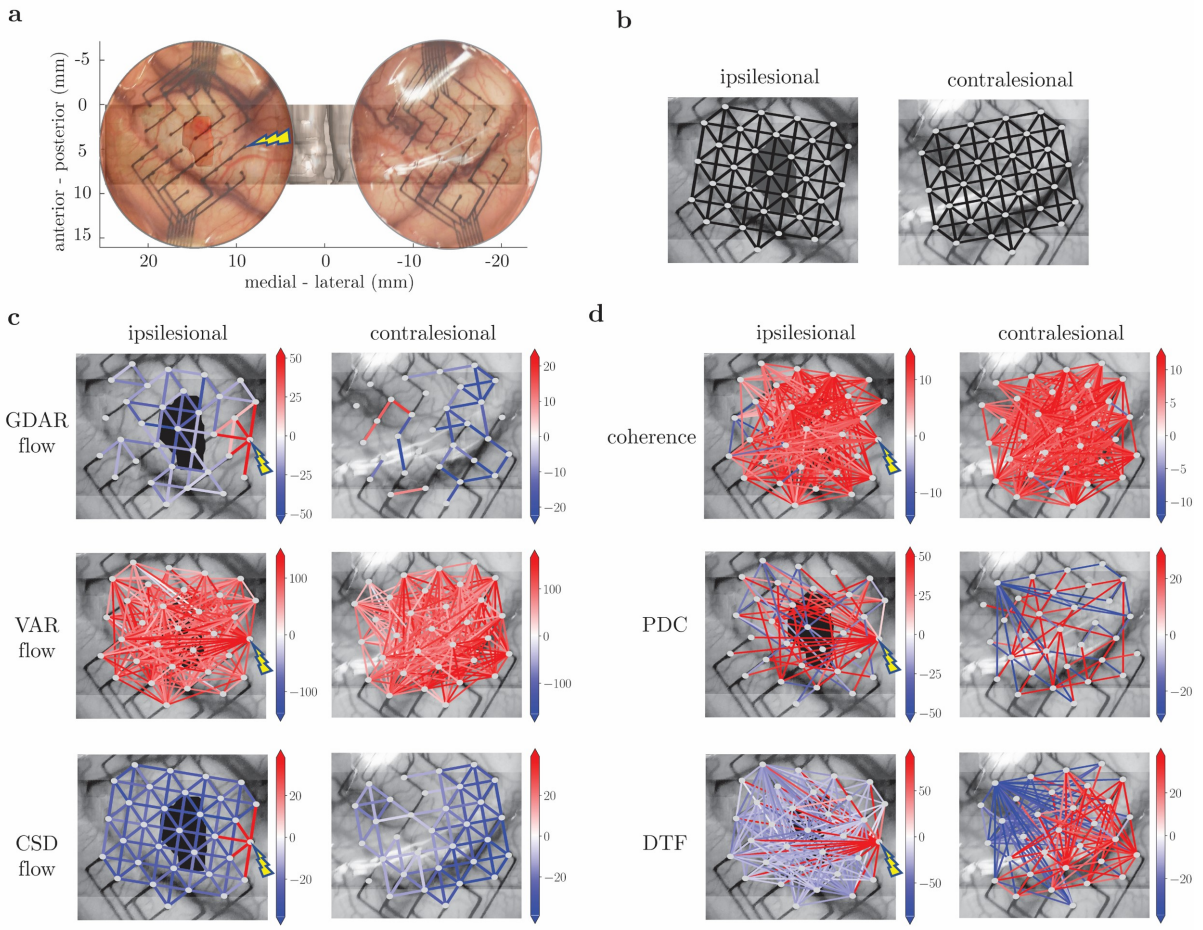


Fig. 5: Analyzing changes in resting state neural communication following electrical stimulation in the acute phase after ischemic lesioning using various flow and FC measures. (a) Two 32-channel ECoG arrays were placed over the left and right hemisphere of a macaque monkey and an ischemic stroke lesion was induced in the left hemisphere using the photothrombotic technique (see Methods). The estimated lesion size is indicated by the red patch in the center of the left hemisphere. One hour after lesioning, electrical stimulation was performed approximately 8 mm away from the lesion location. Neural activity was recorded before and after stimulation to assess the effects of stimulation on network connectivity in the acute phase after stroke. (b) The locations of the electrodes were used to construct two locally connected sparse graphs. (c) Similar to previous applications, 10^{th} order GDAR and VAR models, as well as the CSD approach were used to transform the recorded neural activity before and after stimulation into neural flow signals. Next, power spectra of the flow signals were estimated and changes in gamma $(30-70\ Hz)$ flow power due to stimulation were computed. The changes in gamma flow power (in percent) are shown as undirected edges. Only changes with a significance threshold of $p \leq 0.01$ (two-sample Kolmogorov-Smirnov test) are shown. The lesion location is indicated by the black patch and the stimulation location is shown by the yellow marker. The plots show a local increase in GDAR gamma flow power in the ipsilesional hemisphere near the stimulation location. (d) Same as (c) but instead using coherence, partial directed coherence (PDC), and directed transfer function (DTF) to assess changes in neural communication. Similar to the GDAR flow, PDC and DTF increase for connections with the stimulation location. However, in contrast to the GDAR flow, these changes are less localized and instead affect communication across the entire network.

changes in communication by these two measures affect larger parts of the network. In the contralesional hemisphere, PDC and DTF changes are somewhat opposing the GDAR flow

changes. We note that the patterns observed in Fig. 5c and d are highly frequency dependent (see Extended Fig. 4).

The GDAR flow shows frequency specific correlations with reach velocity and exhibits directional tuning during a center-out reach task: Previous studies have shown that reach movements have strong neural correlates in M1 that can be detected from single neuron recordings as well as intracortical and surface field potentials[41–46]. Using μECoG recordings from M1 of a macaque monkey performing a center-out reach task (Fig. 6a), we show that the GDAR model can be used to study such neural correlates of behavior on the level of network dynamics. To do so, we leverage the graph spectral decomposition of the GDAR flow signal described above and in Methods to decompose the high-gamma GDAR flow into its gradient and rotational flow spectrum (Extended Fig. 5a-e). An example of the time-varying flow spectra for a single reach trial along with the corresponding reach velocity is shown in Fig. 6b. Furthermore, using data from all directions and trials, Fig. 6c shows how the spectral power time series at each spatial frequency correlates with reach velocity. We found that an increase in reach velocity generally correlates with increases in gradient and rotational flow power. Remarkably, the increase in gradient flow power is most pronounced only for the 15 lowest spatial frequencies. In contrast to higher frequencies, such low-frequency flow patterns are more coherent across the graph (see Extended Fig. 5b for examples), suggesting that coordinated activity across a larger cortical area facilitates reach movements. We also observed a similar phenomenon when studying the trial-to-trial variability of the GDAR flow for a single reach direction (see Extended Fig. 6).

To quantify the extent to which the gradient flow spectrum is dominated by low frequencies, we defined the alignment index, which is computed as the ratio of the average power within the 15 lowest spatial frequencies to the average power within the 15 highest spatial frequencies (see Methods). The alignment index shows strong directional tuning with a preference for the 90° (up) and 135° (up-left) directions (Fig. 6d, and Extended Fig. 5g) and a similar cosine characteristic as reported in the literature for other recording modalities[41,42,45]. We also observe a similarly strong directional tuning characteristic for the average power of the rotational spectrum (Fig. 6e and Extended Fig. 5h). In contrast, the directional tuning for the high-gamma envelope of the raw LFP signal averaged across all channels is significantly weaker (Extended Fig. 5i). This suggests that latent patterns of network activity extracted by the GDAR model rather than overall changes in signal power are better correlated with different behaviors.

Finally, we investigated if spectral network features derived from the GDAR flow are correlated to preparatory activity prior to movement onset. We found that the gradient flow alignment index computed during the last 100 ms prior to the go cue shows a strong negative correlation with the reaction time, which is defined as the time between go cue and initiation of movement (Fig. 6f). This effect cannot be explained by any potential premature movements (Extended Fig. 5j). These findings suggest that the degree of neural coordination as captured by the GDAR flow alignment index does not only predict how fast movements are performed, but also how well the monkey is prepared at the start of the go cue.

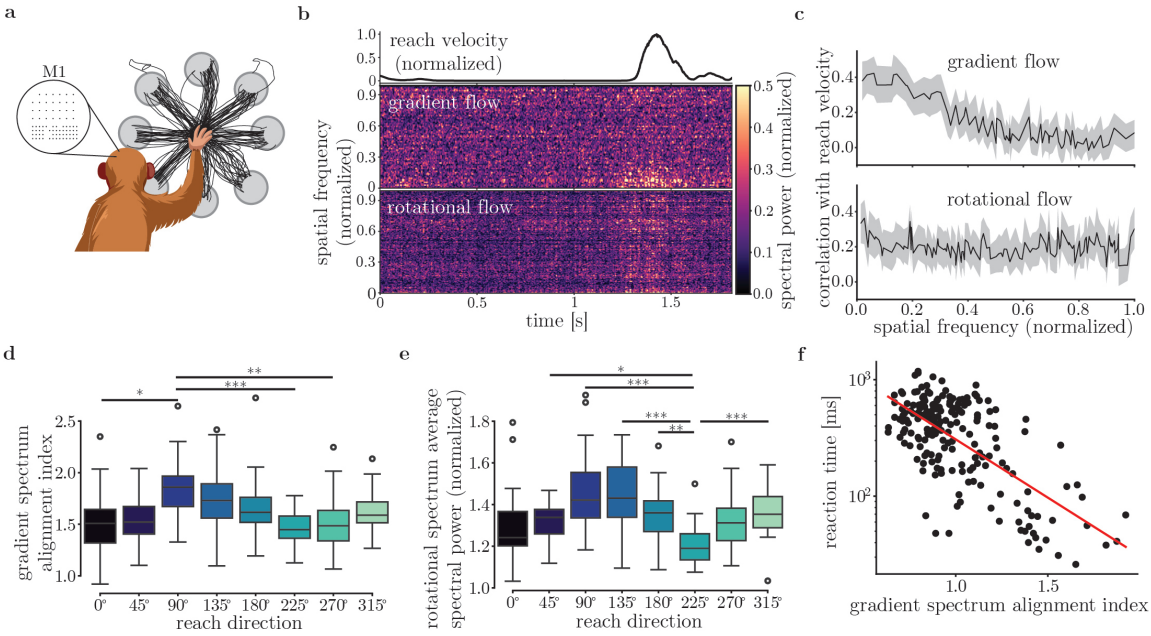


Fig. 6: Applying the GDAR model to ECoG recordings during a center-out reach tasks. (a) A rhesus macaque monkey performed an eight directional reach tasks with 25 trials per direction while LFPs were recorded using a 96-channel micro-ECoG array placed over the primary motor cortex. The GDAR flow was computed for each reach trial, bandpass filtered between $70-200$ *Hz, and decomposed into its gradient and rotational spectrum for each time bin (see Methods). (b) Gradient and rotational spectrogram for a single reach trial. The black line shows the reach velocity. During the reach we observe an increase in rotational flow power across all frequencies and gradient flow power for low spatial frequencies. (c) Correlation (median and interquartile range) between reach velocity and flow spectral power for all spatial frequencies pooled across all trials and directions. Low-frequency gradient flow components show the highest correlation with reach velocity suggestion more activity is coordinated across the brain network during reaching. (d) Alignment index, defined as the ratio of the 15 lowest to the 15 highest gradient flow spectral coefficients for all eight reach directions (shown are quartiles,* 1.5 *times the interquartile range, and outliers). The alignment index forms a cosine-like tuning curve with a preference for the 90° (up) and 135° (up-left) directions. (e) Same as in (d) but for the average rotational flow power. (f) Reaction time, defined as the time between the go cue and movement onset, as a function of the gradient spectrum alignment index during the last 100 ms before the go cue. The strong negative correlation suggests that more coordinated network activity a faster reaction time (correlation coefficient:* -0.674*; p-value:* $7.99 \cdot 10^{-28}$*).*

The GDAR model generalizes to unseen data better than VAR models: To test the GDAR model's ability to generalize to unseen data, we evaluate the model's one-step ahead prediction performance on data that were not included for estimating its parameters. Using the simulated field potentials from the 7-node network of Wilson-Cowan oscillators shown in Extended Fig. 1e (100 independent trials with 5s of simulated neural activity per trial), the GDAR model was trained on the initial N samples of each trial and then tested on the remaining samples (see Methods for more details). An advantage of the GDAR model, owing to its fewer parameters compared to both VAR and eVAR models, is its reduced need for extensive training samples to accurately estimate model parameters. This results in the flatter RMSE curves for both training and testing data (Fig. 7a left and middle) as well as the lowest generalization gap (difference between test and train RMSE) among all models (Fig. 7a, right). Notably, the generalization gap for the GDAR model is nearly an order of magnitude smaller than that for the eVAR model.

We found similar results across the electrophysiology datasets (Fig. 7b-d). Despite the VAR model possessing significantly more parameters – ranging from 8 to 20 times more, depending on the size of the electrode array – the GDAR model exhibits comparable predictive performance on the training set (Fig. 7b). Unlike the VAR model, the GDAR model also generalizes almost perfectly to unseen data, as evidenced by a median generalization gap very close to zero for all datasets (Fig. 7c). Finally, we tested how well the model generalizes to data separated by longer time periods from the training set (Fig. 7d). The GDAR model again maintains a lower generalization gap across all time gaps and datasets compared to the VAR model, with the generalization gap remaining relatively stable as gap lengths increased. An intriguing observation emerged from the optogenetic stimulation dataset, where both the GDAR and VAR models exhibited increasing generalization gaps for larger gap lengths. We believe that this trend stems from plasticity mechanisms within each stimulation block, as repeated paired stimulation induces sustained alterations in brain network activity, thereby challenging the models' ability to generalize over extended time periods.

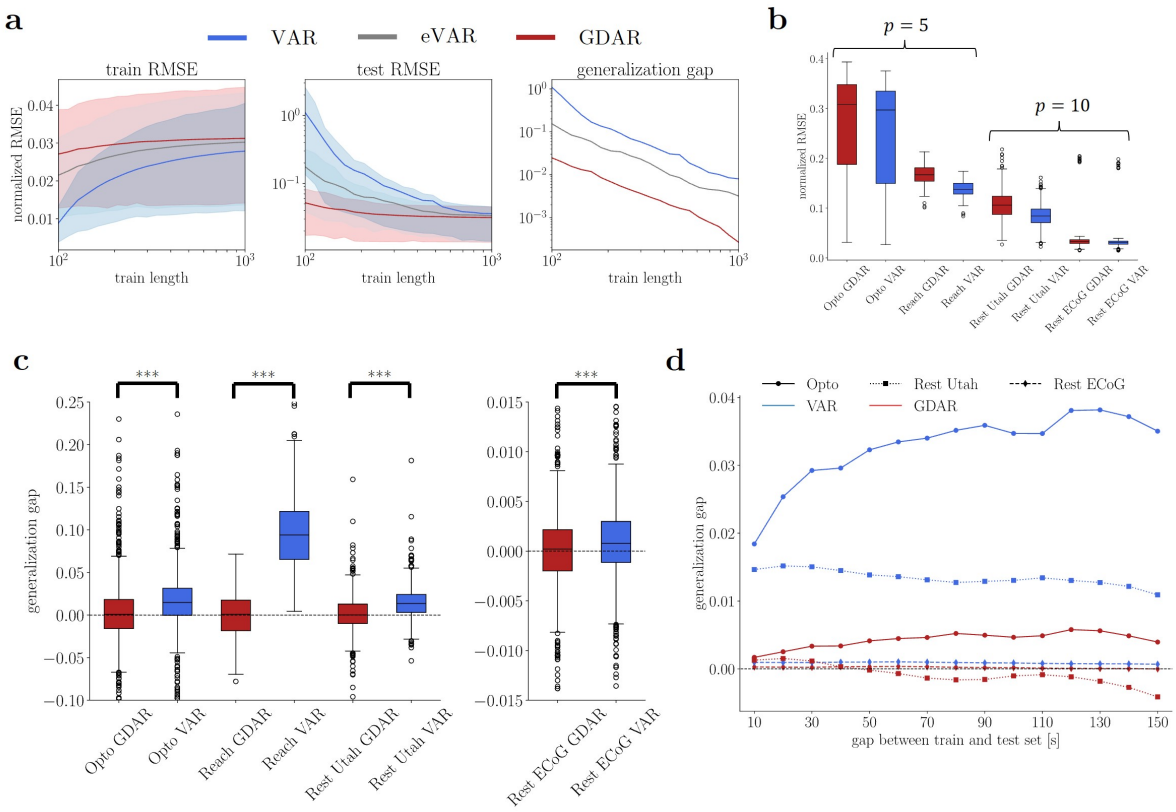


Fig. 7: Generalization Performance of GDAR model on simulated data and electrophysiological recordings. (a) 10^{th} order GDAR, VAR and eVAR models were fit to the first N samples (train length) of neural activity simulated by the network of Wilson-Cowan oscillators shown in Fig. 2a independently for each of the 100 simulation trials. The model coefficients are then used to perform a one-step ahead prediction on the training data as well as the remaining samples in the trial (test data). The left and middle panel show the mean, 10^{th}, and 90^{th} percentile of the root mean square prediction error (RMSEs). The generalization gap (right panel) is defined as the difference between mean test and train RMSE. The GDAR model generalizes significantly better to unseen data than the VAR and eVAR model. (b) Normalized train RMSE of GDAR and VAR models for all four electrophysiology datasets considered in this study. The RMSE generally decreases with increasing model order p and is comparable between GDAR and VAR model within each dataset. (c) Generalization gap (see Methods) of both models for the electrophysiology datasets. The GDAR model almost perfectly generalizes to unseen data. On the other hand, the VAR model always shows some degree of overfitting (Wilcoxon rank-sum test, $p \leq 0.001$). (d) The generalization gap for the optogenetic stimulation and two resting state datasets as a function of the time gap between train and test set. Except for the optogenetic stimulation experiment, the generalization gap remains constant or decreases as the time gap between train and test set increases. For all time gaps, the GDAR model outperforms the VAR model.

Discussion

By drawing insights from both computational neuroscience and statistical modeling, we have introduced the GDAR model as a novel framework for estimating network level neural communication dynamics from field potential recordings. Our approach is defined by three key components – each previously explored in isolation by different communities, but not yet integrated into a unified framework. First, we combine the modeling capabilities of classical VAR models with a network diffusion process that serves as a plausible mechanistic constraint for neural communication. Second, the spatial layout of the recording array is incorporated as a structural prior, significantly reducing the model complexity while mimicking cortical connectivity on a

macroscopic scale. Third, our model produces a communication signal with the same temporal resolution as the original recordings, making it well suited for analyzing both transient and long-term patterns of neural communication. Using simulations and four electrophysiology datasets from macaque sensory motor cortex, we have demonstrated that the GDAR model outperforms competing approaches (standard VAR models, CSD flow) in estimating fast communication dynamics, provides complementary insights into resting state communication that are consistent across different experiments and recording technologies, and can uncover neural flow dynamics that correlate with behavior.

Why does the GDAR model perform better than standard VAR models? VAR models and other approaches for estimating functional brain connectivity have successfully been used to study neural interactions for many years. Yet these techniques remain relatively generic and lack a mechanism through which neural populations interact. In contrast, the GDAR model assumes that information propagates via a diffusive process, which has previously been proposed as a mechanism for neural communication[27,47]. It has also been shown that diffusion processes can explain functional connectivity estimates[48] and model the propagation of activity evoked by intracranial stimulation more accurately than alternative models of neural communication[49]. Furthermore, the Laplacian that drives the diffusion process in the GDAR model has been used in neural field models to simulate realistic large-scale brain dynamics[23,50]. In particular, our finding that the GDAR model outperforms the enhanced VAR model – which differs from GDAR only in lacking the diffusion constraint – highlights the importance of including mechanistic assumptions into data-driven modeling.

Another drawback of standard VAR models is that they generally ignore spatial relations between the recording electrodes, which means that interactions between nearby sensors are treated equally to interactions between distant ones. The idea of integrating spatial information in the form of structural priors into standard VAR models and other FC measures has recently been proposed in magnetic resonance imaging (MRI), electroencephalography (EEG), and magnetoencephalography (MEG) studies, where it has been shown to improve the estimation of FC networks[51–54]. However, this direction remains under-explored and, to the best of our knowledge, has not been applied to localized recording arrays that focus on network dynamics within one or two cortices. Furthermore, the studies that incorporate spatial information lack mechanistic assumptions about the neural communication process and have almost exclusively focused on static FC metrics. In contrast, the GDAR framework naturally produces a dynamic network flow signal by integrating structural priors and mechanistic constraints into a single model thereby likely contributing to its superior performance over VAR models.

Our framework also uses a different mechanism for obtaining temporally resolved communication dynamics. Unlike existing approaches, which derive such dynamics through sliding windows[55–57] or adaptive parameter estimation[54,58], the GDAR model achieves this by combining static model parameter with the recorded neural activity. This approach offers several advantages: it reduces the number of parameters that need to be estimated, and it enables the detection of transient communication events that may be smoothed out by sliding window approaches or are difficult to track using linear adaptive parameter estimation techniques.

The GDAR model also has additional practical advantages for processing field potential signals. For the electrode arrays used in our analysis, the GDAR model has approximately ten times fewer parameters than the full VAR model. This larger number of parameters for the VAR model can cause them to overfit to idiosyncrasies in the data that do not correspond to meaningful neural interactions, which is particularly evident in the poor generalization performance for the reach dataset where model fitting relies on a limited number of observations (see Fig. 7c). Furthermore, field potential recordings can suffer from spurious correlations due to volume conduction, signal artifacts that are shared across channels, and the common reference signal problem. Such spurious correlations are known to cause erroneous connectivity estimates in classical measures of neural communication such as coherence, phase locking value, or metrics based on standard VAR models[17]. These spurious correlations can be addressed by preprocessing field potentials using CSD (i.e., the second spatial derivative) or activity gradients (i.e., the first spatial derivative) instead of using the raw neural activity[17,59,60]. Since the GDAR model employs the second spatial derivative, the effects of spurious correlations are strongly mitigated and should not negatively impact the performance of the model.

While the assumption of a locally connected nearest neighbor graph as a structural prior is inspired by the cortical connectivity found in both mice and macaque monkeys, which is dominated by short range connections[28], it neglects the potential existence of any direct long-range propagation paths. Since it can be difficult to determine the best underlying network structure as structural information is often not available, we suggest that in the future the structural connectivity graph could be designed in a more data-driven way, for example, using sparsity and distance regularizers. Furthermore, we currently make no distinction between nodes corresponding to electrodes in the interior versus the boundary of the array. Especially the boundary nodes may exhibit a large exchange of information with regions outside the array, which is not captured by the model, but could be incorporated by adding additional input terms. Another promising avenue would be to explore how other proposed mechanisms of neural signaling, such as biased random walks or shortest path routing[27,47], could be incorporated as constraints into data-driven models of network communication. Furthermore, the GDAR model could be extended to model non-linear communication dynamics by introducing activation functions into the divergence step of the model.

Although we developed the GDAR for field potential recordings and have applied it to a range of cortical electrophysiology datasets to evaluate its performance and demonstrate its versatility, we believe the model can be extended to other neural recording modalities and applications. For instance, it may be adapted for spiking data, either by modifying the autoregressive component to accommodate for discrete point processes – such as through generalized linear models[61] – or by first converting spikes into firing rates. The model should also be readily applicable to human ECoG and stereoelectroencephalography (sEEG) recordings, which share similar signal properties with the recordings analyzed here. Finally, the GDAR model can be applied to brain-wide recoding modalities such as EEG, MEG or functional MRI (fMRI), which – combined with estimated structural connectivity networks – can enable the analysis of large-scale brain dynamics. In a preliminary investigation, we found that our model reliably estimates neural communication dynamics from resting-state fMRI data and is sensitive to age-related changes in neural flow[62], highlighting its potential for broader applications in systems neuroscience and clinical research.

Acknowledgements

We thank Daniel Silversmith and Joseph O'Doherty for their help with data collection and Philip Sabes for his laboratory in which some of the data was collected. We also thank Toni Haun, Sandi Thelen, and Christopher English for their help with animal surgeries and experimentation for the stroke experiment. This work was supported by the American Heart Association (FS, AY), the National Institute of Health R01NS119593 (JB, AY) and R01MH125429 (FS, AY), the Washington Research Foundation (AY), the Big Data for Genomics and Neuroscience Training Grant NIH 5T32LM012419 (JB), the Center for Neurotechnology NSFERC 1028725 (JB), the Washington National Primate Research Center NIH P51OD010425, U42 OD011123 (AY), the Eunice Kennedy Shriver National Institute of Child Health and Human Development NIH K12HD073945 (AY), the National Institute of Neurological Disorders and Stroke of the National Institute of Health R01NS116464 (AY, JZ), the University of Washington Royalty Research Fund (AY, KK), the National Science Foundation Graduate Research Fellowship Program (KK), the Graduate Education for Minorities Fellowship (JB), and the Weill Neurohub (JZ).

Methods

Graph Diffusion Autoregressive (GDAR) Model

Derivation and algebraic representation: The starting point for deriving the GDAR model is to describe the spatiotemporal dynamics of the neural activity s as a heat diffusion process

$$\dot{s} = w\Delta s, \tag{3}$$

where temporal changes in activity ($\dot{s}$) are driven by spatial activity gradients (Δs, where Δ is the surface Laplacian) multiplied by the diffusion rate w. The right-hand side of (3) is equivalent to current source density (CSD) which is a common technique for analyzing neurophysiological recordings. In practice, we have given a finite set of spatially distributed and discrete LFP measurements recorded from an N-channel electrode array (see Fig. 1a left). Thus, we can denote the LFPs recorded at time t as a vector $s[t] \in \mathbb{R}^N$. The surface Laplacian Δ is equivalent to the second spatial derivative and thus describes local interactions within the brain network. In a discrete measurement setup, this can be encoded by constructing a locally connected graph from the locations of the electrodes within the recording array[63]. Thereby, each electrode corresponds to a node in the graph and edges connect neighboring electrodes such as illustrated in Fig. 1a. The resulting unweighted graph consisting of N vertices and E edges can be represented algebraically using the node-to-edge incident matrix $\mathbf{B} \in \mathbb{R}^{N\times E}$, where the e^{th} column $b^{(e)}$ corresponds to the e^{th} edge in the graph. Each edge is defined by a tail node n_i and head node n_j such that $b^{(e)}_{n_i} = -b^{(e)}_{n_j} = -1$ and all other entries $b^{(e)}_{n_k} = 0$ for $k \neq i, j$. For each edge it is thereby arbitrary which incident node is defined as head and tail node. Using $\mathbf{B}$, the continuous surface Laplacian Δ can be approximated using the negative of the graph Laplacian operator $\mathbf{B}\mathbf{B}^T$. Furthermore, the first temporal derivative $\dot{s}$ can be approximated by the first temporal difference $s[t] - s[t-1]$. Thus, (3) can be approximated by

$$s[t] = (\mathbf{I}_N - w\mathbf{B}\mathbf{B}^T)s[t-1] + u[t] = \mathbf{I}_N s[t-1] - \mathbf{B}w\mathbf{B}^T s[t-1] + u[t], \tag{4}$$

Where, $\mathbf{I}_N$ is the $N \times N$ identity matrix and $u[t]$ is a white noise term. Previously, it has been shown that the matrices $\mathbf{B}^T$ and $\mathbf{B}$ can be interpreted as discrete approximations of the gradient and divergence operators, respectively[35]. Thus, the term $\mathbf{B}w\mathbf{B}^T s[t-1]$ has a clear physical meaning in the context of LFP recordings as elaborated in the steps below:

1. $\mathbf{B}^T s[t-1] = \nabla s[t-1]$: computes the voltage gradient for each node in the graph.
2. $w\nabla s[t-1] = f[t]$: In analogy to resistive circuits and CSD analysis, w can be interpreted as a conductivity such that conductivity times voltage gradient yields a current flow $f[t]$.
3. $\mathbf{B}f[t]$: For each node, the net flow, i.e., the sum of all in flows minus the sum of all outflows, is computed. This is equivalent to computing the current sources and sinks in CSD analysis.

Equation (4) effectively expresses CSD analysis as a first order vector autoregressive (VAR) model. However, the model in Eq. (4) has limited expressivity as the only learnable parameter is the conductivity w. Thus, to improve its expressivity, we can 1) add parameterized node dynamics, 2) assume a spatially varying conductivity, and 3) extend the model order to a pth order VAR model. The resulting graph diffusion autoregressive (GDAR) model is given by

$$s[t] = \sum_{k=1}^{p} (\mathbf{M}_k - \mathbf{B}\mathbf{W}_k\mathbf{B}^T)s[t-k] + u[t], \tag{5}$$

where $\mathbf{M}_k = \mathrm{diag}(\boldsymbol{m}_k) \in \mathbb{R}^{N\times N}$ and $\mathbf{W}_k = \mathrm{diag}(\boldsymbol{w}_k) \in \mathbb{R}^{E\times E}$ are diagonal matrices containing the node and edge parameters $\boldsymbol{m}_k \in \mathbb{R}^N$ and $\boldsymbol{w}_k \in \mathbb{R}^E$ of the k^{th} lag, respectively. The term $\mathbf{B}\mathbf{W}_k\mathbf{B}^T$ can also be regarded as a weighted graph Laplacian matrix. The GDAR flow is defined as

$$f[t] = \sum_{k=1}^{p} \mathbf{W}_k\mathbf{B}^T s[t-k] \tag{6}$$

Representation as constrained VAR model: The GDAR model in (5) can be related to the standard notation of a VAR model

$$s[t] = \sum_{k=1}^{p} \mathbf{A}_k s[t-k] + u[t] \tag{7}$$

where $\mathbf{A}_k \in \mathbb{R}^{N\times N}$ contains the VAR model parameters and is generally a dense matrix. It can be shown that if $\mathbf{A}_k$ has the same sparsity structure as the graph Laplacian $\mathbf{B}\mathbf{B}^T$ and is symmetric, Eq. (7) is equivalent to Eq. (5) with $(\mathbf{A}_k)_{i,j} = (\mathbf{A}_k)_{j,i} = (\mathbf{W}_k)_{l,l}$ if l corresponds to the edge between node i and j and $(\mathbf{M}_k)_{i,i} = (\mathbf{A}_k)_{i,i} + \sum_{j\in\mathcal{N}(i)}(\mathbf{A}_k)_{i,j}$, where $\mathcal{N}(i)$ are the set of neighbors of node i. The representation of the GDAR model as a constrained VAR model is useful for fitting the model to neural data.

Model fitting: Using the VAR representation in Eq. (7), the model parameters $\mathbf{M}_k$ and $\mathbf{W}_k$ can be estimated using least squares regression following the procedure described by Lütkepohl[64]. Given $T + p$ snapshots of neural activity by an N-channel recording array (T is the number of samples used for model fitting), the predicted neural activity can be collected in the data matrix $\mathbf{Y} = \big[s[p+1], \ldots, s[p+T]\big] \in \mathbb{R}^{N\times T}$ and its vectorized version $\gamma = \mathrm{vec}(\mathbf{Y})$. The regressors can be

expressed as $\mathbf{S} = [S_1, \dots, S_T] \in \mathbb{R}^{Np \times T}$, where $S_t = [s[t+p-1]^T, \dots, s[t]^T]^T \in \mathbb{R}^{Np \times 1}$. The coefficients $\mathbf{A}_k$ can be expressed as $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_p] \in \mathbb{R}^{N \times Np}$ and $\alpha = \mathrm{vec}(\mathbf{A})$. As shown in the previous section, $\mathbf{A}_k$ is spare and symmetric. Therefore, there exist a matrix $\mathbf{R}$ such that $\alpha = \mathbf{R}\tilde{\alpha}$ and $\tilde{\alpha}$ only contains the non-zero entries of the upper triangle of $\mathbf{A}$. Now (7) can be written as

$$\gamma = (\mathbf{S}^T \otimes \mathbf{I}_N)\mathbf{R}\tilde{\alpha} + u, \tag{8}$$

where $\otimes$ is the Kronecker product. Furthermore, we assume that u is white noise with covariance matrix $\mathbf{\Sigma}_u$. Eq. (8) can be solved in close form by minimizing $u^T(\mathbf{I}_T \otimes \mathbf{\Sigma}_u^{-1})u$, where $\mathbf{I}_T$ is the $T \times T$ identity matrix, to obtain the optimal parameters $\tilde{\alpha}^*$:

$$\tilde{\alpha}^* = [\mathbf{R}^T(\mathbf{S}\mathbf{S}^T \otimes \mathbf{\Sigma}_u^{-1})\mathbf{R}]^{-1}\mathbf{R}^T(\mathbf{S} \otimes \mathbf{\Sigma}_u^{-1})\gamma. \tag{9}$$

Eq. (9) is the solution to the generalized least squares (GLS) estimator, which is generally different from the ordinary least squares (OLS) estimator due to the sparsity and symmetry contraints[64]. However, it requires knowledge of the noise covariance matrix $\mathbf{\Sigma}_u^{-1}$, which is unknown in practice. Therefore, we first estimate the $\mathbf{\Sigma}_u^{-1}$ by solving the OLS estimator $u^T u$ to compute $\hat{\tilde{\alpha}}$ as

$$\hat{\tilde{\alpha}} = [\mathbf{R}^T(\mathbf{S}\mathbf{S}^T \otimes \mathbf{I}_N)\mathbf{R}]^{-1}\mathbf{R}^T(\mathbf{S} \otimes \mathbf{I}_N)\gamma \tag{10}$$

and denote $\hat{\alpha} = R\hat{\tilde{\alpha}}$. The corresponding coefficient matrix is $\widehat{\mathbf{A}}$ with $vec(\widehat{\mathbf{A}}) = \hat{\alpha}$. Then we estimate $\mathbf{\Sigma}_u$ as

$$\mathbf{\Sigma}_u = \frac{1}{T}(\mathbf{Y} - \widehat{\mathbf{A}}\mathbf{S})(\mathbf{Y} - \widehat{\mathbf{A}}\mathbf{S})^T \tag{11}$$

It is also noted that (5) can be directly casted as a least squares minimization problem. However, we found that it is more efficient to compute the optimal parameters according to (8).

Power spectrum of GDAR flow: If the model is applied to resting state neural activity, the GDAR flow signal may exhibit a similar oscillatory behavior as the neural activity. Therefore, it may be reasonable to compute its power spectrum to study frequency specific communication patterns. Using Eq. (6) and recognizing that it expresses the GDAR flow as the convolution between the model parameters $\mathbf{W}_k$ and the activity gradients $\mathbf{B}^T s[t]$, the GDAR flow power spectrum between nodes i and j is given by

$$\left|F^{\{i,j\}}(\omega)\right|^2 = \left|W^{\{i,j\}}(\omega)\right|^2 \cdot \left|S_j(\omega) - S_i(\omega)\right|^2, \tag{12}$$

where $W^{\{i,j\}}(\omega)$, $S_i(\omega)$, and $S_j(\omega)$ are the Fourier transforms of the parameters, as well as the neural activity of the two channels, and ω is the frequency variable. An interesting case occurs when the spectra of both channels have the same magnitude for a given frequency. Assuming $|S_i(\omega)| = |S_j(\omega)| = 1$, Eq. (12) can be simplified to

$$\frac{\left|F^{\{i,j\}}(\omega)\right|^2}{2} = \left|W^{\{i,j\}}(\omega)\right|^2 \cdot [1 - \cos(\phi_j - \phi_i)], \tag{13}$$

where ϕ_i and ϕ_j are the phase of the $S_i(\omega)$, and $S_j(\omega)$, respectively. That is, in this case the communication dynamics are driven only by phase differences between connected nodes. In general, however, communication dynamics will be determined by differences in magnitude and phase modulated by $W^{\{i,j\}}(\omega)$, which was estimated with the objective of improving the prediction of future neural activity.

Decomposition into gradient and rotational flow spectra: Similar to the classical Fourier transform for time series, where a signal can be decomposed into a series of oscillatory components of increasing frequency, a flow signal can be decomposed into a set of spatial components (flow signals) with increasing spatial frequency. Furthermore, a flow signal can be decomposed into gradient (directional) components, which have non-zero divergence (sum of in-flow minus out-flow) for some or all nodes of the graph, and rotational components, which have zero divergence for all graph nodes. This can be achieved via the Hodge-decomposition that defines two orthogonal sets of spatial basis functions (defined on the edge domain) for a given graph[33–35]. Each GDAR flow snapshot can then be projected onto these sets of basis functions to obtain the gradient and rotational flow spectrum.

To obtain the gradient basis, we first compute the eigenvectors $\widetilde{\mathbf{V}}_{\mathbf{grad}} \in \mathbb{R}^{N\times N}$ of the graph Laplacian $\mathbf{B}\mathbf{B}^T$. The orthonormal gradient flow basis $\mathbf{V}_{\mathbf{grad}} \in \mathbb{R}^{E\times N}$ is then obtained by

$$\mathbf{V}_{\mathbf{grad}} = \frac{\mathbf{B}^T\widetilde{\mathbf{V}}_{\mathbf{grad}}}{\left|\left|\mathbf{B}^T\widetilde{\mathbf{V}}_{\mathbf{grad}}\right|\right|_F}. \tag{14}$$

The eigenvalues $\lambda_{\mathbf{grad}}$ associated with each eigenvector define a natural ordering of the eigenvectors in terms of spatial frequency. Specifically, if we compute the divergence of the eigenvectors $\mathbf{V}_{\mathbf{grad}}$, we find that eigenvectors corresponding to small eigenvalues have small divergence, whereas eigenvectors associated with large eigenvectors have large divergence. Small-divergence eigenvectors correspond to flow signals that are *smooth* (or low-frequency) across the graph, that is flow signals where the direction of flow is largely preserved or only slowly changes within a local neighborhood (also see Extended Fig. 5c for an example). High-divergence eigenvectors on the other hand correspond to flow patterns that rapidly change direction within a local neighborhood and can therefore be considered as *non-smooth* or being high-frequency. We can now obtain a gradient flow spectrum for each flow snapshot by projecting $f[t]$ onto $\mathbf{V}_{\mathbf{grad}}$:

$$F_{\mathbf{grad}}[\lambda_{\mathbf{grad}}, t] = \mathbf{V}_{\mathbf{grad}}^T f[t]. \tag{15}$$

To obtain the rotational basis, we first have to define a set of triangles in the graph, which can be obtained, for example, via Delaunay triangulation. Mathematically, the triangle set is captured by the edge-to-triangle incident matrix $\mathbf{B}_{\mathbf{tri}} \in \mathbb{R}^{E\times T}$, where T is the number of triangles and where the t^{th} column $b_{\mathbf{tri}}^{(t)}$ corresponds to the t^{th} triangle in the graph. Each triangle is defined by three edges e_i, e_j, and e_k and an arbitrarily chosen reference direction. If the edge direction across e_i (as defined in $\mathbf{B}$) aligns with that reference direction $b_{\mathbf{tri}}^{(t)}(e_i) = 1$. Otherwise $b_{\mathbf{tri}}^{(t)}(e_i) = -1$ (the same logic applies to e_j, and e_k). For edges not involved in the triangle we have $b_{\mathbf{tri}}^{(t)}(\cdot) = 0$. To compute the rotational basis, we then follow the same procedure as above. That is, we first compute the eigenvectors $\widetilde{\mathbf{V}}_{\mathbf{rot}} \in \mathbb{R}^{T\times T}$ of the Laplacian $\mathbf{B}_{\mathbf{tri}}^T\mathbf{B}_{\mathbf{tri}}$ and then project $\widetilde{\mathbf{V}}_{\mathbf{rot}}$ onto $\mathbf{B}_{\mathbf{tri}}$ and normalize:

$$\mathbf{V}_{\mathbf{rot}} = \frac{\mathbf{B}_{\mathbf{tri}}\widetilde{\mathbf{V}}_{\mathbf{rot}}}{\left|\left|\mathbf{B}_{\mathbf{tri}}\widetilde{\mathbf{V}}_{\mathbf{rot}}\right|\right|_F}. \tag{16}$$

Similar to the gradient flow, the eigenvalues $\lambda_{\mathbf{rot}}$ corresponding to the eigenvectors $\mathbf{V}_{\mathbf{rot}}$ can be used to define an ordering in terms of spatial frequency. Specifically, eigenvectors with small eigenvalues correspond to global rotational flows (akin to global currents) across the graph that

maintain or only slowly change orientation between local neighborhoods. On the other hand, eigenvectors with small eigenvalues exhibit localized rotational flows (akin to local eddy currents) that rapidly change orientation across local neighborhoods (see Extended Fig. 5e for an example). Finally, we can obtain a rotational flow spectrum for each flow snapshot by projecting $f[t]$ onto $\mathbf{V}_{\text{rot}}$:

$$F_{\text{rot}}[\lambda_{\text{rot}}, t] = \mathbf{V}_{\text{rot}}^T f[t]. \tag{17}$$

Wilson-Cowan Simulations

Simulating neural activity: We simulated neural activity using various networks of Wilson-Cowan oscillators[65,66] shown in Fig. 2 and Extended Fig. 1. Each node consists of an excitatory and inhibitory subpopulation whose dynamics are governed by the following differential equations:

$$\tau_e \frac{de_i(t)}{dt} = -e_i(t) + S\left(c_{ee}e_k(t) + c_{ie}i_k(t) + P + \xi(t) + \sum_{j\in\mathcal{N}(i)} w_{j\to i}e_j\big(t-\tau_{jk}\big)\right) \tag{18}$$

$$\tau_i \frac{di_i(t)}{dt} = -i_i(t) + S\big(c_{ei}e_i(t) + \xi(t)\big) \tag{19}$$

where S is the sigmoid function:

$$S(x) = \frac{1}{1 + e^{\frac{x-\mu}{\sigma}}} \tag{20}$$

The description of the parameters and their values are listed in Table 1. The values are based on previous work by Abeysuriya et al.[67] and Deco et al.[68] and result in a power spectrum with a pronounced beta oscillation around 18 Hz and a $1/\omega$ slope for higher frequencies. Coupling between excitatory populations of neighboring nodes is determined by the parameter $w_{j\to i}$ where each edge in the graph has two coupling parameters ($w_{j\to i}$ and $w_{i\to j}$) resulting in bidirectional coupling. For the 16-node random graphs, we simulated 10 independent trials per graph, resulting in a total of 100 trials for 10 graphs, where for each trial the values of the edge weights $w_{j\to i}$ are randomly sampled from a uniform distribution (see Table 1 for range of $w_{j\to i}$). For the 7-node, and 16-node grid graph, we simulated 100 independent trials respectively. The ranges of $w_{j\to i}$ were chosen such that neural activity whose power spectrum resembles realistic local field potential signals was generated by the network. We integrated the system with a time step of 1e-4 seconds using a 4th order Runge-Kutta scheme for 20 seconds and discarded the first 15 seconds to eliminate transient effects of the simulation. The resultant 5 seconds of excitatory activity $e[t]$ was then downsampled to 1 kHz using an 8th order Chebyshev type I anti-aliasing filter and denoted as the simulated neural activity. Power spectral density (PSD) estimates of the simulated activity and ground truth flow for the 16-node random graphs averaged over all trials, graphs, and edges are shown in Extended Fig. 1b and c.

Table 1: Simulation parameters of Wilson-Cowan model adapted from Abeysuriya et al.[67] and Deco et al.[68]

Parameter	Description	Value
τ_e	Excitatory time constant	0.002
τ_i	Inhibitory time constant	0.004
c_{ee}	Local excitatory to excitatory coupling	3.5
c_{ie}	Local inhibitory to excitatory coupling	-2.5
c_{ei}	Local excitatory to inhibitory coupling	3.75
P	Constant excitatory input	0.31
μ	Firing response threshold	1
σ	Firing threshold variability	0.25
ξ	Random noise input	$\mathcal{N}(0, 0.05)$
$w_{i\to j}$	Excitatory to excitatory connectivity (16 node random graphs)	0.05, ..., 0.3
$w_{i\to j}$	Excitatory to excitatory connectivity (7 node graph)	0.05, ..., 0.55
$w_{i\to j}$	Excitatory to excitatory connectivity (16 node grid graph)	0.1, ... 0.5

Simulating ground truth neural flow: We simulated the ground truth flow by calculating the moment-to-moment influence that each excitatory node exerts on its neighbors. To do so, we first executed each integration step with the full set of parameters to obtain $e[t]$. Then, for each excitatory coupling parameter $w_{j\to i}$, we repeated the integration step with $w_{j\to i} = 0$ to obtain $e_i[t]|_{w_{j\to i}=0}$, which denotes the activity at node i in the absence of an influence from node j at time t. The flow from node j to node i was then defined as $f_{j\to i}[t] = e_i[t] - e_i[t]|_{w_{j\to i}=0}$. This second step is repeated for all excitatory coupling parameters. The full two-step procedure is repeated for each integration step. The resulting bidirectional ground truth flow $f^{\mathrm{gt,b}}[t]$ was downsampled using the same anti-aliasing filter as used for the simulated neural activity. As our GDAR model only produced a unidirectional flow at each point in time, we define the unidirectional ground truth flow $f^{\mathrm{gt}}[t]$ between node i and j as $f^{\mathrm{gt}}_{i,j}[t] = f^{gt,b}_{i\to j}[t] - f^{gt,b}_{j\to i}[t]$. It is noted that while $f^{\mathrm{gt}}[t]$ is unidirectional at each time point, the flow direction across each edge can change over time.

GDAR flow: For each trial, we used the last 5 seconds of simulated neural activity to estimate the parameters of the GDAR model for varying model orders as described in (Graph Diffusion Autoregressive (GDAR) Model). The graph used for fitting the model is equal to the graph used for the simulations. The estimated model parameters were then used to transform the simulated neural activity into an estimated flow signal according to (6).

VAR flow: For comparison, we also estimated the neural flow using a classical VAR model (Eq. (7)). To do so, we first estimated the VAR model parameters using the same data as for fitting the GDAR model. The directional flow across edge $j \to i$ is then computed as $f^{\mathrm{VAR,b}}_{j\to i}[t] = \sum_{k=1}^{p}(A_k)_{i,j}s_i[t]$. Similar to the ground truth flow, the unidirectional flow is defined as $f^{\mathrm{VAR}}_{i,j}[t] = f^{VAR,b}_{i\to j}[t] - f^{VAR,b}_{j\to i}[t]$. The VAR model assumes a fully connected network resulting into non-zero flow signals across connections that are not part of the network. To compare the VAR flow with

the ground truth flow, we therefore only extract the VAR flow for edges that exists in the ground truth network.

eVAR flow: For a fair comparison with the GDAR model, we test a second autoregressive model that has access to the ground truth graph when estimating the VAR model coefficients. That is, we enforce $(A_k)_{i,j} = (A_k)_{j,i} = 0$ if node i and j are not connected. Using this eVAR models, we also computed a bidirectional flow $f^{\mathrm{VAR,b}}[t]$ and compare it to the ground truth bidirectional flow $f^{\mathrm{gt,b}}[t]$ for the 7-node graph. However, we found that this does not result in higher correlation coefficients than the unidirectional flow (see Extended Fig. 1f).

CSD flow: The last approach for estimating the neural flow is through CSD analysis. Since CSD is the second spatial derivative, which, for a given graph, can be approximated as the graph Laplacian operator $\mathbf{B}\mathbf{B}^T$, the CSD flow is simply the gradient between the simulated neural activity at connected nodes in the network: $f_{i,j}^{CSD}[t] = s_i[t] - s_j[t]$. This is equivalent to a first order GDAR model with spatially constant conductivity.

Comparing ground truth and estimated neural flow: The ground truth and estimated flow signals are first z-scored independently for each trial and model, and then compared using the Pearson correlation coefficient (CC) computed independently for each edge in the graph. The CC distributions obtained by pooling CCs from all edges and 100 trials for each model are compared using a Wilcoxon rank-sum test. Furthermore, we computed the error between magnitude and phase spectrum of ground truth and estimated flow for each graph edge and trial (Fig. 2c). To do so, the power spectral density (PSD) of the flow across each edge (5s for each trial) was estimated using Welch's method[69] with a Hann window of size 256 samples and 50% overlap. Then the absolute difference (in dB) between ground truth and estimated flow was computed. To compare the phases, the 5s of neural flow obtained for each trial were first divided into 19 non-overlapping segments of 256 samples and then the discrete Fourier transform for each segment was computed. Afterwards, the phase difference between ground truth and estimated phase was computed and mapped into the range from 0 to π for each segment before being averaged over all 19 segments. Fig. 2c shows the median, first, and third quartile of the magnitude and phase difference using data from all edges and trials. The PSDs of the ground truth and estimated flow signals were used to compute the PSD correlations in Fig. 2d. Finally, we compared the dynamics of the estimated flow with the dynamics of the ground truth flow using dynamical similarity analysis (DSA) in Fig. 2e[36]. To do so, Hankel dynamic mode decomposition (DMD) models are first independently fitted to the high-dimensional ground truth and estimated flow signal and the resultant DMD matrices A_{est} and A_{gt} are compared using a modified version of Procrustes analysis. To fit the Hankel-DMD models we used 15 delay time steps to construct the Hankel matrices and full rank regression. Optimization during the Procrustes analysis used 1000 iterations at a learning rate of 10^{-2}.

Optogenetic Stimulation Experiment

One adult male rhesus macaque (monkey G: 8 years old, 17.5 kg) was used in this experiment. All procedures were performed under the approval of the University of California, San Francisco

Institutional Animal Care and Use Committee and were compliant with the Guide for the Care and Use of Laboratory Animals.

Neural stimulation and recording interface: In this study, we used a subset of neural data recorded by a large-scale optogenetic neural interface[11] that has previously been utilized to study changes in network functional connectivity due to cortical stimulation[37,38]. The interface was composed of several key components: a semi-transparent micro-electrode array, a semi-transparent artificial dura, a titanium implant, and a laser system for delivering optical stimulation. First, neurons in the primary sensorimotor cortex were rendered light-sensitive through a viral-mediated expression of the C1V1 opsin. To do so, $200\,\mu$L of the viral cocktail AAV5-CamKIIa-C1V1(E122T/E162T)-TS-eYFP-WPRE-hGH (2.5×10^{12} virus molecules/mL; Penn Vector Core, University of Pennsylvania, PA, USA, Addgene number: 35499) was administered across four sites into the primary somatosensory (S1) and primary motor (M1) cortices of the left hemisphere using convection-enhanced delivery[11,37,70]. Next, the chronic neural interface was surgically implanted by performing a 25mm craniotomy over the primary sensorimotor cortex, replacing the dura mater beneath the craniotomy with a chronic transparent artificial dura housed in a titanium chamber. During each experimental session, the artificial dura was removed and a custom 96 channel micro-electrocorticography array consisting of platinum-gold-platinum electrodes and traces encapsuled in Parylene-C[12] was placed on the cortical surface. Optical stimulation was performed by two 488 nm lasers (PhoxX 488-60, Omicron-Laserage, Germany) connected to a fiber optic cable (core/cladding diameter: 62.5/125 µm, Fiber Systems, TX, USA) and positioned above the array such that the tip of the fiber-optic cable touched the array. Neural data in the form of local field potentials was recorded by the micro-ECoG array at a sampling frequency of 24 kHz using a Tucker-Davis Technologies system (FL, USA). It was verified that evoked neural responses were due to optogenetic activation and not other effects such as photoelectric artifacts or heating[11,12,38].

Stimulation protocol: The data analyzed in this study stems from three experimental sessions all performed on the same day. The only difference between the sessions was the location of stimulation, which is depicted in Fig. 3b. As the micro-ECoG array was not removed between sessions its location on the cortex remains unchanged. Each experimental session consists of 5 stimulation blocks during which two lasers alternatingly and repeatedly stimulate. Each stimulation block lasts approximately 7 min and is intermittent by shorter resting state blocks during which no stimulation is performed. The stimulation pulse width for both lasers was 5 ms with a delay of 10 ms between stimulation by lasers 1 and 2. This paired stimulation is repeated at a frequency of 7 Hz (143 ms) resulting in a total of approximately 2970 pulse pairs for each stimulation block. All stimulation parameters (except for stimulation locations) are identical for the three sessions analyzed in this study.

Signal preprocessing: First, bad channels were identified as 1) electrodes with high impedance and 2) channels with a low signal-to-noise ratio, and excluded from the analysis[38]. The location of the remaining 67 good channels was used to construct a sparse and locally connected graph, where each electrode corresponds to a node in the graph and each node is connected approximately to it 8 nearest neighbors (see Fig. 3c top). The raw time series data was downsampled to 1017.25 Hz using a low-pass Chebyshev anti-aliasing filter and the mean activity within each channel was subtracted from the respective time series.

GDAR model fitting: The preprocessed LPFs during each stimulation block were divided into segments of 10004 samples (approximately 10 s) with 4 samples overlap between segments and a 5^{th} order GDAR model was fitted to each segment as described in (Graph diffusion autoregressive (GDAR) model). The estimated model parameters were used to transform the neural activity into a the GDAR flow signal according to equation (6). The overlap between segments was chosen such that a continuous GDAR flow signal was obtained from the segmented LFPs without relying on zero padding. A model order of 5 was chosen for this application due to the short (10 ms) delay between stimulation by lasers 1 and 2. For larger model orders, the GDAR flow evoked by the second laser would increasingly be influenced by the neural activity evoked by the first laser resulting in a mixing of the neural responses to both stimulation pulses. Flow dynamics akin to the plots in Fig. 3e for a model order $p = 10$ are shown in Extended Fig. 2c and d.

Visualizing flow dynamics: To visualize the flow dynamics evoked by paired cortical stimulation, we have to project the high dimensional flow signals ($f[t] \in \mathbb{R}^E$, where E is the number of edges in the graph) onto a lower dimensional subspace. To so so, we first pooled the first 25 flow snapshots from the onset of stimulation by the first laser from all sessions, blocks, and pulse pairs in a single data matrix $F \in \mathbb{R}^{E \times M}$, where $M \approx 3 \cdot 5 \cdot 2970 \cdot 25$ (3 sessions, 5 blocks per session, approximately 2970 pulse pairs per block, 25 flow snapsots per pulse pair). Afterwards we performed principle component analysis (PCA) and projected F onto its first two principal components (PCs) to obtain $\tilde{F} \in \mathbb{R}^{2 \times M}$. Fig. 3e shows the PCA reduced GDAR flow dynamics where each trace illustrates a 25 snapshot long flow trajectory from a single pulse pair. For better visualization only 250 individual trajectories per stimulation block selected at random are plotted. Fig. 3g shows the same dynamics but instead using the VAR and CSD flow, respectively. Since the number of edges for the VAR model is very large, computing the PCs of the associated matrix M was not feasible. Therefore we first averaged the flow snapshots over 20 consequitive trials before computing the PCs. For comparison, we performed the same trial averaing for the GDAR flow and recomputed the flow trajectories (Extended Fig. 2e). The averaging does not seem to have a negative effect on the discriminability of the trajectories between different sessions.

Modeling increased delay across sulcus: The GDAR model can easily be augmented to model variable delay across different edges. For example, it is reasonable to assume that signals that travel across the sulcus between M1 and S1 experience larger delays than signals traveling within each cortex. Larger delays in the GDAR model across an edge between node i and j can be incorporated by constraining edge coefficients $w_k^{\{i,j\}} = 0$ for small delays (i.e., $k = 1, 2, \dots$), which can be achieved by augmenting the matrix $\mathbf{R}$ in equation (8). We have used this approach to model larger delays across the sulcus by setting $w_k^{\{i,j\}} = 0$ for edges that connect nodes in M1 to nodes in S1 for $k = 1, 2, 3$. That is, the minimum delay across each sulcus edge is constrained to be 4 (see Extended Fig. 2f and Supplementary Video 2 for corresponding GDAR flow dynamics).

Changes in resting-state communication due to electrical stimulation:

To demonstrate the GDAR model's ability to uncover changes in communication during resting state, we analyze data from two distinct experiments that were conducted using a 96-channel microelectrode array (Utah array) and two 32-channel ECoG arrays.

Utah array experimental procedure: One adult rhesus macaque (*Macaca mulatta*, 12 kg, 11 years, male) was used in this study. All procedures were performed under the approval of the University of California, San Francisco Institutional Animal Care and Use Committee and were compliant with the Guide for the Care and Use of Laboratory Animals. The experimental procedure was previously described by Bloch et al.[71]. A 96-channel Utah array was implanted in S1 and LFPs were recorded at a sampling frequency of 24 kHz before being downsampled to a frequency of 1017 Hz (8^{th} order Chebychev anti-aliasing filter). The dataset consists of resting state recordings intermitted by five 10 minute stimulation blocks that contain repeated single site or paired electrical stimulation. For the single site stimulation session, stimulation is performed in in the form of five pulses (1 kHz burst frequency) that are repeated every 200 ms. The paired stimulation sessions use the same stimulation patterns for each stimulation site. For session *paired-stim 1*, electrode B stimulated 100 ms after electrode A. For session *paired-stim 2*, the delay between stimulation sites A and B is chosen uniformly at random between -100 ms and 100 ms for each paired stimulation trial.

ECoG array experimental procedure: One adult macaque (*Macaca nemestrina*, 14.6 kg, 7 years, male) was used in this study. All procedures were performed under the approval of the University of Washington Institutional Animal Care and Use Committee and were compliant with the Guide for the Care and Use of Laboratory Animals. The experimental procedure was previously described elsewhere[13,72,73]. The animal was first anesthetized with isoflurane and a craniotomy with 25mm diameter was performed in each hemisphere over the sensorimotor cortex. A focal ischemic lesion in the left hemisphere was created by photo-activation of a previously injected light-sensitive dye (Rose Bengal). Following illumination, the dye causes platelet aggregation, thrombi formation, and interruption of local blood flow, leading to local neural cell death near the illuminated area. The location and extent of the lesion were estimated through post-mortem histological analysis of coronal slices and is illustrated as a black patch in Fig. 5. Electrical activity was recorded before, during, and after lesion induction simultaneously in the ipsi- and contralesional hemisphere using two 32-channel ECoG arrays (Fig. 5c)[74,75]. Approximately 60 min after the end of lesioning, repeated electrical stimulation was performed 8 mm away from the lesion center. 1 kHz stimulation charge-balanced pulses (60 µA, 450 µs pulse width, 50 µs interphase interval) were given in 5 Hz bursts (5 pulses per burst) consecutively for 10 minutes, where each stimulation block was followed by 2 min of baseline recording. The experiments included a total of six 10 min stimulation blocks. We used the 60 min of neural recording after lesion induction but before stimulation (pre stim), as well as the 2 min blocks of baseline recording in between the stimulation blocks (post stim). In total we used 4 blocks of post stim recordings for each hemisphere as the recordings in the other blocks were corrupted.

Signal preprocessing: The preprocessing for both datasets was performed akin to the optogenetic stimulation experiment. The location of the ECoG channels was used to construct a sparse and locally connected graph (for the ECoG data, this was done separately for each hemisphere), where each node (electrode) is connected approximately to its 8 nearest neighbors (no bad channels were identified). The raw time series data was downsampled to 1 kHz using low-pass Chebyshev anti-aliasing filter and the mean was removed from each channel. Additionally, artifacts – defined as signal values that deviate by ten or more standard deviations from the mean

simultaneously for all channels – were removed by linearly interpolating between the sample immediately before and after the artifact.

Model fitting and postprocessing: The preprocessed LFPs for both datasets during each block were divided into segments of length 10009 samples (approximately 10 s) with 9 samples overlap between segments and a 10^{th} order GDAR model was fitted to each segment as described in Graph diffusion autoregressive (GDAR) model. The estimated model parameters were used to transform the neural activity into the GDAR flow signal according to equation (6), where each segments contains 10000 samples. To assess changes in neural communication due to stimulation in different frequency bands, we then computed the GDAR flow power spectral density (PSD) using Welch's method[69] (Hann window of size 1000 samples with 50% overlap) for each segment, and stored the average flow PSD within the gamma band ($30-70$ Hz). Finally, we computed the change in average flow PSD from before to after stimulation. Specifically, if we denote $F_k^{\text{pre stim}}$ and $F_k^{\text{post stim}}$ as the average gamma flow PSD before and after stimulation of the k^{th} segment, the relative change in GDAR flow magnitude Δ_{stim} is given by

$$\Delta_{\text{stim}} = \frac{\langle F_k^{\text{post stim}} \rangle_k - \langle F_k^{\text{pre stim}} \rangle_k}{\langle F_k^{\text{pre stim}} \rangle_k}, \tag{21}$$

where $\langle \cdot \rangle_k$ denotes the average over all segments. We assessed the statistical significance of Δ_{stim} for each edge by forming sample distributions for pre- and post-stim communication from all pre- and post-stim segments and compared the distributions using a two-sample Kolmogorov-Smirnov test. If the distributions for a given edge differ with a significance level of $p \leq 0.01$, the edge is plotted in the graph.

For the Utah array data, changes in gamma GDAR flow power due to stimulation (Fig. 4b) were computed using all data before stimulation (pre stim) as well as the five 2-minute resting state blocks following the stimulation blocks (post stim) for each session. To compute the temporal evolution of the normalized and averaged gamma GDAR flow power (Fig. 4c-e), the GDAR flow power in the gamma band was first averaged over all edges connected to the stimulation node. Then GDAR flow and LFP power were z-scored using the mean and standard deviation from the pre stim period for each session independently. We then computed the best linear fit between the z-scored LFP and average GDAR flow power $\bar{F}_{\text{GDAR}}$ using all segments (pre and post stim)

$$\bar{F}_{\text{GDAR}} = s \cdot LFP + o. \tag{22}$$

The goal is to test whether the GDAR flow power changes beyond what can be linearly explained by changes in LFP power. Hence, we subtract the linear regression line from the average GDAR flow power

$$\bar{F}_{\text{GDAR,corrected}} = \bar{F}_{\text{GDAR}} - s \cdot LFP. \tag{23}$$

and plotted the result in Fig. 4c-e.

For the ECoG data, we additionally computed the change in flow power Δ_{stim} using the CSD approach and a 10^{th} order VAR model (CSD and VAR flow were computed as described in Wilson-

Cowan Simulations). The 10th order VAR model was also used to compute changes in coherence, partial directed coherence (PDC)[22] and directed transfer function (DTF)[21]. PDC and DTF for each directed edge pair $j \to i$ were calculated using the following equations:

$$PDC_{j\to i} = \frac{\left|A_{i,j}(\omega)\right|^2}{\sum_{l=1}^{N}\left|A_{l,j}(\omega)\right|^2}, \tag{24}$$

$$DTF_{j\to i} = \frac{\left|H_{i,j}(\omega)\right|^2}{\sum_{l=1}^{N}\left|H_{i,l}(\omega)\right|^2}, \tag{25}$$

where $A_{i,j}(\omega)$ is the Fourier transform of $(A_k)_{i,j}$ (note that here k is the time variable), and $H_{i,j}(\omega)$ is the (i,j) entry of the Fourier transform of the inverse of A_k. To obtain a unidirectional communication signal (Fig. 5d), we calculated the average between $i \to j$ and $j \to i$ for each pair of edges.

Center-out Reach Task

One adult male rhesus macaque (7 years old, 16.5 kg) was used in this study. All procedures were performed under the approval of the University of California, San Francisco Institutional Animal Care and Use Committee and were compliant with the Guide for the Care and Use of Laboratory Animals. Surgical procedure, neural interface, and signal preprocessing are the same as described in Optogenetic Stimulation Experiment. However, for the center-out reach task, the ECoG was placed fully over the primary motor (M1) cortex. Channels with persistent distortions were identified and excluded from the analysis resulting in 77 good channels used to construct a sparsely connected nearest neighbor graph as described previously. The animal performed a total of 200 successful reach trial, 25 for each of the eight directions (see Fig. 6a). Each individual reach trial is divided into start, instructed delay, and reach phase. During the start phase, the monkey places its hand on the center of the screen. After that the instructed delay phase begins where first the target direction is presented before a randomly selected delay period terminated by a go-tone is introduced. The reach phase starts once the go-tone appears and ends when the monkey touches the target. The finger position of the monkey was tracked throughout the experiment using and electromagnetic position sensor (Polhemus Liberty, Colchester, VT) at 240 Hz[76].

GDAR model fitting and post-processing: To ensure accurate model fitting, recorded LFPs from all three phases were used to estimate the parameters of the GDAR model. The model order was set to $p = 5$ to ensure enough independent samples for each parameter. After the model parameters have been estimated, the GDAR flow is computed according to Eq. (6) and filtered into the high-gamma band using a 3rd order Butterworth filter with cutoff frequencies of 70 and 200 Hz. The high-gamma GDAR flow signal $f[t]$ is then decomposed into its gradient and rotational flow spectrum $F_{\text{grad}}[\lambda_{\text{grad}}, t]$ and $F_{\text{rot}}[\lambda_{\text{rot}}, t]$ according to Eq.(14)-(17). To obtain the flow power spectrogram in Fig. 6b, we compute the magnitude square of $F_{\text{grad}}[\lambda_{\text{grad}}, t]$ and $F_{\text{rot}}[\lambda_{\text{rot}}, t]$. Flow power spectra as well as reach velocities are temporally smoothed using a 51 sample 3rd order Savitzky-Golay filter. To account for the time delay between motor commands observable in M1 and actual movement onset[77], we calculated the median correlation across all spatial frequencies λ_{grad} between $F_{\text{grad}}[\lambda_{\text{grad}}, t-d]$ and the reach velocity for varying delays d

(Extended Fig. 5f). We found a maximum correlation for a delay of 104 ms, which we corrected for in all subsequent analysis.

To quantify the extend to which the gradient flow spectrum is dominated by low frequencies during reaching, we defined the alignment index as

$$\text{aligment index} = \frac{\sum_t \sum_{i=1}^{15} F_{\text{grad}}[\lambda_{\text{grad},i}, t]}{\sum_t \sum_{i=1}^{15} F_{\text{grad}}[\lambda_{\text{grad,N}-i}, t]}, \tag{26}$$

where N is the total number of gradient frequencies. The temporal averaging is performed over all time points where the reach velocity is above a threshold of 0.1 for the tuning curve analysis (Fig. 6d) and over the last 100 ms prior to the go-tone for the reaction time analysis (Fig. 6f). For the rotational flow spectrum, we do not observe spectral changes during reaching that are strongly dependent on the spatial frequency. Therefore, we simply use the average over all spatial frequencies in Fig. 6e.

Generalization performance

According to Eq. (5), the GDAR model can predict the neural activity at the current time step using the past p samples. To assess the generalization performance of the model, we computed the normalized root mean square error (RMSE) between the observed neural activity $s[t]$ and predicted neural activity $\hat{s}[t]$ as follows:

$$RMSE = \frac{\sqrt{\sum_{t,n}(\hat{s}_n[t] - s_n[t])^2}}{\sqrt{\sum_{t,n}(s_n[t])^2}} \tag{27}$$

The summation is performed over all time points t within a segment, as well as over all channels n of the recording array. To compute the train RMSEs, the predictions $\hat{s}[t]$ are computed for the same time points that were used for model fitting. For the test RMSEs, the models are applied to data that were not used for fitting the model. To compute the test RMSEs for the optogenetic stimulation, stroke, and Utah array datasets, the prediction RMSEs are computed for the 10 s segment that immediately follows the segment used for model fitting. That is, if the models have been fitted using segment i, the test RMSEs are computed using segment $i + 1$. For Fig. 7d, where the generalization gap over larger time scales is assessed, the prediction RMSEs are computed for segments further away from the segment used for model fitting. For the reach dataset, the models are tested on the subsequent reach trial in the same direction. That is if the models have been fitted using trial i from direction d then the test RMSE is computed for trial $i + 1$ from direction d.

Data availability:

Data will be made available upon reasonable request from the authors.

Code Availability:

Source code for the GDAR model will be made available prior to publication.

887

888 ## References:

889 1. Zeki, S. & Shipp, S. The functional logic of cortical connections. *Nature* **335**, 311–317 (1988).

890 2. Bressler, S. L. Large-scale cortical networks and cognition. *Brain Res. Rev.* **20**, 288–304 (1995).

891 3. Friston, K. Beyond Phrenology: What Can Neuroimaging Tell Us About Distributed Circuitry? *Annu.*
892 *Rev. Neurosci.* **25**, 221–250 (2002).

893 4. McIntosh, A. R. Towards a network theory of cognition. *Neural Netw.* **13**, 861–870 (2000).

894 5. Duncan, J., Humphreys, G. & Ward, R. Competitive brain activity in visual attention. *Curr. Opin.*
895 *Neurobiol.* **7**, 255–261 (1997).

896 6. Bressler, S. L. & Menon, V. Large-scale brain networks in cognition: emerging methods and
897 principles. *Trends Cogn. Sci.* **14**, 277–290 (2010).

898 7. Voytek, B. & Knight, R. T. Dynamic Network Communication as a Unifying Neural Basis for Cognition,
899 Development, Aging, and Disease. *Biol. Psychiatry* **77**, 1089–1097 (2015).

900 8. He, B. J., Shulman, G. L., Snyder, A. Z. & Corbetta, M. The role of impaired neuronal communication
901 in neurological disorders. *Curr. Opin. Neurol.* **20**, 655 (2007).

902 9. Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L. & Greicius, M. D. Neurodegenerative Diseases
903 Target Large-Scale Human Brain Networks. *Neuron* **62**, 42–52 (2009).

904 10. Shing, N., Walker, M. C. & Chang, P. The role of aberrant neural oscillations in the hippocampal-
905 medial prefrontal cortex circuit in neurodevelopmental and neurological disorders. *Neurobiol. Learn.*
906 *Mem.* **195**, 107683 (2022).

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

909 12. Ledochowitsch, P. *et al.* Strategies for optical control and simultaneous electrical readout of
910 extended cortical circuits. *J. Neurosci. Methods* **256**, 220–231 (2015).

911 13. Khateeb, K. *et al.* A versatile toolbox for studying cortical physiology in primates. *Cell Rep. Methods*
912 **2**, 100183 (2022).

913 14. Griggs, D. J. *et al.* Optimized large-scale optogenetic interface for non-human primates. in
914 *Optogenetics and Optical Manipulation 2019* (eds. Mohanty, S. K. & Jansen, E. D.) 3 (SPIE, San
915 Francisco, United States, 2019). doi:10.1117/12.2511317.

15. Vázquez-Guardado, A., Yang, Y., Bandodkar, A. J. & Rogers, J. A. Recent advances in neurotechnologies with broad potential for neuroscience research. *Nat. Neurosci.* **23**, 1522–1536 (2020).

16. Chang, E. F. Towards Large-Scale, Human-Based, Mesoscopic Neurotechnologies. *Neuron* **86**, 68–78 (2015).

17. Bastos, A. M. & Schoffelen, J.-M. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Front. Syst. Neurosci.* **9**, (2016).

18. He, B. *et al.* Electrophysiological Brain Connectivity: Theory and Implementation. *IEEE Trans. Biomed. Eng.* **66**, 2115–2137 (2019).

19. Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37**, 424–438 (1969).

20. Geweke, J. F. Measures of Conditional Linear Dependence and Feedback between Time Series. *J. Am. Stat. Assoc.* **79**, 907–915 (1984).

21. Kaminski, M. J. & Blinowska, K. J. A new method of the description of the information flow in the brain structures. *Biol. Cybern.* **65**, 203–210 (1991).

22. Baccalá, L. A. & Sameshima, K. Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* **84**, 463–474 (2001).

23. Breakspear, M. Dynamic models of large-scale brain activity. *Nat. Neurosci.* **20**, 340–352 (2017).

24. Bassett, D. S., Zurn, P. & Gold, J. I. On the nature and use of models in network neuroscience. *Nat. Rev. Neurosci.* **19**, 566–578 (2018).

25. Rubino, D., Robbins, K. A. & Hatsopoulos, N. G. Propagating waves mediate information transfer in the motor cortex. *Nat. Neurosci.* **9**, 1549–1557 (2006).

26. Muller, L., Reynaud, A., Chavane, F. & Destexhe, A. The stimulus-evoked population response in visual cortex of awake monkey is a propagating wave. *Nat. Commun.* **5**, 3675 (2014).

27. Avena-Koenigsberger, A., Misic, B. & Sporns, O. Communication dynamics in complex brain networks. *Nat. Rev. Neurosci.* **19**, 17–33 (2018).

28. Horvát, S. *et al.* Spatial Embedding and Wiring Cost Constrain the Functional Layout of the Cortical Network of Rodents and Primates. *PLOS Biol.* **14**, e1002512 (2016).

29. Schwock, F., Bloch, J., Atlas, L., Abadi, S. & Yazdan-Shahmorad, A. Estimating and Analyzing Neural Information flow using Signal Processing on Graphs. in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1–5 (IEEE, 2023).

30. Coifman, R. R. & Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006).

31. Carvalhaes, C. & de Barros, J. A. The surface Laplacian technique in EEG: Theory and methods. *Int. J. Psychophysiol.* **97**, 174–188 (2015).

32. Mitzdorf, U. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* **65**, 37–100 (1985).

33. Barbarossa, S. & Sardellitti, S. Topological Signal Processing Over Simplicial Complexes. *IEEE Trans. Signal Process.* **68**, 2992–3007 (2020).

34. Schaub, M. T. & Segarra, S. FLOW SMOOTHING AND DENOISING: GRAPH SIGNAL PROCESSING IN THE EDGE-SPACE. in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* 735–739 (IEEE, Anaheim, CA, USA, 2018). doi:10.1109/GlobalSIP.2018.8646701.

35. Schaub, M. T., Zhu, Y., Seby, J.-B., Roddenberry, T. M. & Segarra, S. Signal processing on higher-order networks: Livin' on the edge... and beyond. *Signal Process.* **187**, 108149 (2021).

36. Ostrow, M., Eisen, A., Kozachkov, L. & Fiete, I. Beyond Geometry: Comparing the Temporal Structure of Computation in Neural Circuits with Dynamical Similarity Analysis. *Adv. Neural Inf. Process. Syst.* **36**, 33824–33837 (2023).

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

38. Bloch, J. *et al.* Network structure mediates functional reorganization induced by optogenetic stimulation of non-human primate sensorimotor cortex. *iScience* **25**, 104285 (2022).

39. Yang, Y. *et al.* Modelling and prediction of the dynamic responses of large-scale brain networks during direct electrical stimulation. *Nat. Biomed. Eng.* **5**, 324–345 (2021).

40. Sun, S. *et al.* Human intracortical responses to varying electrical stimulation conditions are separable in low-dimensional subspaces. in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* 2663–2668 (2022). doi:10.1109/SMC53654.2022.9945369.

41. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).

42. Schwartz, A. B., Kettner, R. E. & Georgopoulos, A. P. Primate motor cortex and free arm movements to visual targets in three- dimensional space. I. Relations between single cell discharge and direction of movement. *J. Neurosci.* **8**, 2913–2927 (1988).

43. Moran, D. W. & Schwartz, A. B. Motor Cortical Representation of Speed and Direction During Reaching. *J. Neurophysiol.* **82**, 2676–2692 (1999).

44. Miller, K. J. *et al.* Spectral Changes in Cortical Surface Potentials during Motor Movement. *J. Neurosci.* **27**, 2424–2432 (2007).

45. Rickert, J. *et al.* Encoding of Movement Direction in Different Frequency Ranges of Motor Cortical Local Field Potentials. *J. Neurosci.* **25**, 8815–8824 (2005).

46. Heldman, D. A., Wang, W., Chan, S. S. & Moran, D. W. Local field potential spectral tuning in motor cortex during reaching. *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**, 180–183 (2006).

47. Seguin, C., Sporns, O. & Zalesky, A. Brain network communication: concepts, models and applications. *Nat. Rev. Neurosci.* **24**, 557–574 (2023).

48. Abdelnour, F., Voss, H. U. & Raj, A. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *NeuroImage* **90**, 335–347 (2014).

49. Seguin, C. *et al.* Communication dynamics in the human connectome shape the cortex-wide propagation of direct electrical stimulation. *Neuron* **111**, 1391-1401.e5 (2023).

50. Jirsa, V. K. & Haken, H. Field Theory of Electromagnetic Brain Activity. *Phys. Rev. Lett.* **77**, 960–963 (1996).

51. Ng, B., Varoquaux, G., Poline, J.-B. & Thirion, B. A Novel Sparse Graphical Approach for Multimodal Brain Connectivity Inference. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012* (eds. Ayache, N., Delingette, H., Golland, P. & Mori, K.) 707–714 (Springer, Berlin, Heidelberg, 2012). doi:10.1007/978-3-642-33415-3_87.

52. Hinne, M., Ambrogioni, L., Janssen, R. J., Heskes, T. & van Gerven, M. A. J. Structurally-informed Bayesian functional connectivity analysis. *NeuroImage* **86**, 294–305 (2014).

53. Pineda-Pardo, J. A. *et al.* Guiding functional connectivity estimation by structural connectivity in MEG: an application to discrimination of conditions of mild cognitive impairment. *NeuroImage* **101**, 765–777 (2014).

54. Pascucci, D. *et al.* Structure supports function: Informing directed and dynamic functional connectivity with anatomical priors. *Netw. Neurosci.* **6**, 401–419 (2022).

55. Shakil, S., Lee, C.-H. & Keilholz, S. D. Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states. *NeuroImage* **133**, 111–128 (2016).

56. Scheid, B. H. *et al.* Time-evolving controllability of effective connectivity networks during seizure progression. *Proc. Natl. Acad. Sci.* **118**, e2006436118 (2021).

57. Allen, E. A., Damaraju, E., Eichele, T., Wu, L. & Calhoun, V. D. EEG Signatures of Dynamic Functional Network Connectivity States. *Brain Topogr.* **31**, 101–116 (2018).

58. Milde, T. *et al.* A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials. *NeuroImage* **50**, 960–969 (2010).

59. Trongnetrpunya, A. *et al.* Assessing Granger Causality in Electrophysiological Data: Removing the Adverse Effects of Common Signals via Bipolar Derivations. *Front. Syst. Neurosci.* **9**, (2016).

60. Pesaran, B. *et al.* Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation. *Nat. Neurosci.* **21**, 903–919 (2018).

61. Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P. & Brown, E. N. A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. *J. Neurophysiol.* **93**, 1074–1089 (2005).

62. Schwock, F., Nordgren, D., Atlas, L., Yazdan-Shahmorad, A. & Hesamoddin, J. Integrating Structural and Functional Connectivity for Dynamic fMRI Modeling Via Graph Diffusion Autoregression. in *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Scociety (EMBC)* (2025).

63. Burago, D., Ivanov, S. & Kurylev, Y. A graph discretization of the Laplace–Beltrami operator. *J. Spectr. Theory* **4**, 675–714 (2015).

64. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*. (New York : Springer, Berlin, 2005).

65. Wilson, H. R. & Cowan, J. D. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophys. J.* **12**, 1–24 (1972).

66. Wilson, H. R. & Cowan, J. D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **13**, 55–80 (1973).

67. Abeysuriya, R. G. *et al.* A biophysical model of dynamic balancing of excitation and inhibition in fast oscillatory large-scale networks. *PLOS Comput. Biol.* **14**, e1006007 (2018).

68. Deco, G., Jirsa, V., McIntosh, A. R., Sporns, O. & Kötter, R. Key role of coupling, delay, and noise in resting brain fluctuations. *Proc. Natl. Acad. Sci.* **106**, 10302–10307 (2009).

69. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoustics* **15**, 70–73 (1967).

70. Khateeb, K., Griggs, D. J., Sabes, P. N. & Yazdan-Shahmorad, A. Convection Enhanced Delivery of Optogenetic Adeno-associated Viral Vector to the Cortex of Rhesus Macaque Under Guidance of Online MRI Images. *JoVE J. Vis. Exp.* e59232 (2019) doi:10.3791/59232.

71. Bloch, J. A. *et al.* Cortical Stimulation Induces Network-Wide Coherence Change In Non-Human Primate Somatosensory Cortex. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 6446–6449 (2019). doi:10.1109/EMBC.2019.8856633.

72. Khateeb, K. *et al.* A Practical Method for Creating Targeted Focal Ischemic Stroke in the Cortex of Nonhuman Primates. in *2019 41st Annual International Conference of the IEEE Engineering in*

Medicine and Biology Society (EMBC) 3515–3518 (IEEE, Berlin, Germany, 2019). doi:10.1109/EMBC.2019.8857741.

73. Zhou, J. *et al.* Neuroprotective Effects of Electrical Stimulation Following Ischemic Stroke in Non-Human Primates. in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 3085–3088 (2022). doi:10.1109/EMBC48229.2022.9871335.

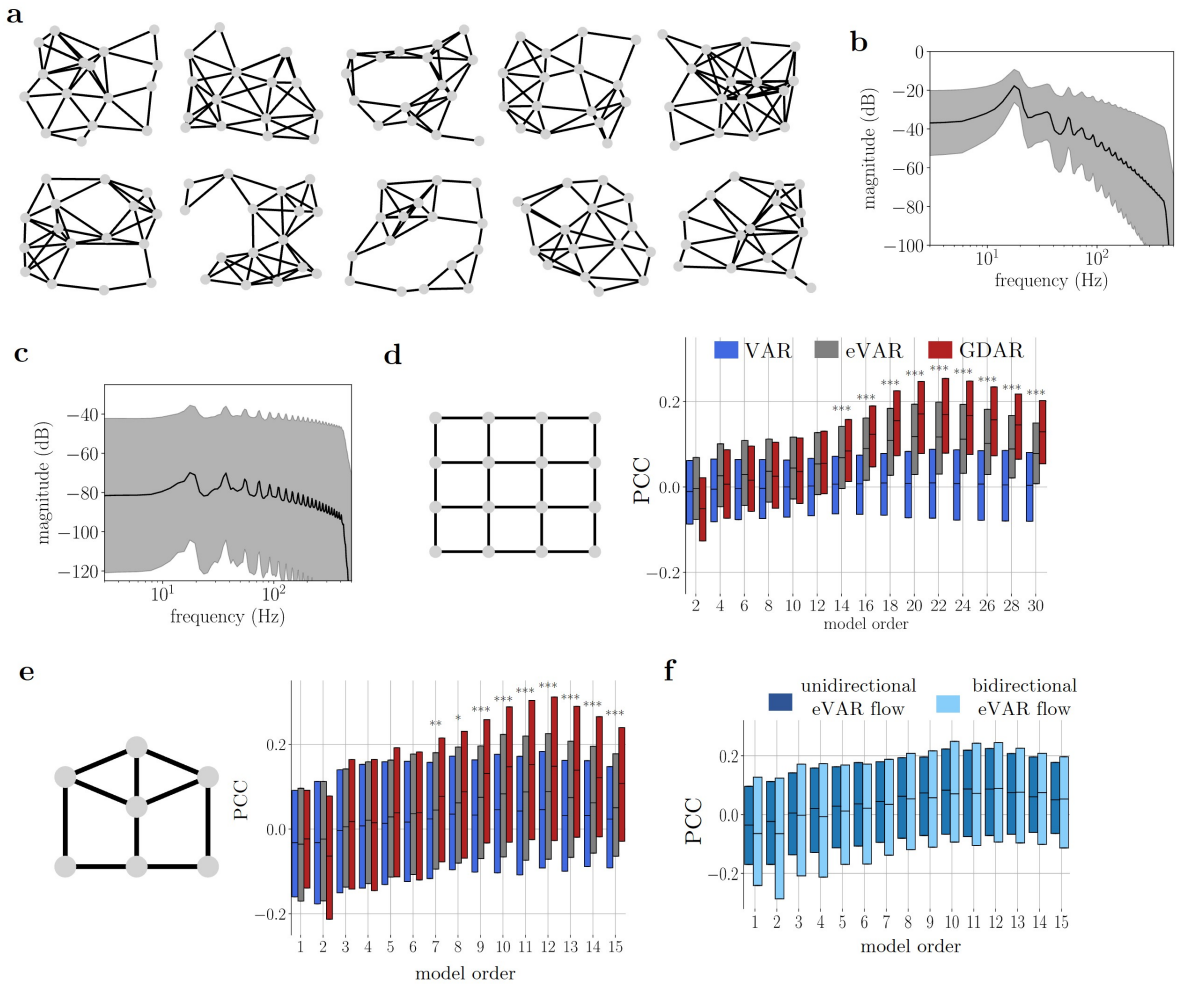
74. Griggs, D. J. *et al.* Multi-modal artificial dura for simultaneous large-scale optical access and large-scale electrophysiology in non-human primate cortex. *J. Neural Eng.* **18**, 055006 (2021).

75. Griggs, D. J. *et al.* Demonstration of an Optimized Large-scale Optogenetic Cortical Interface for Non-human Primates. in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 3081–3084 (2022). doi:10.1109/EMBC48229.2022.9871332.

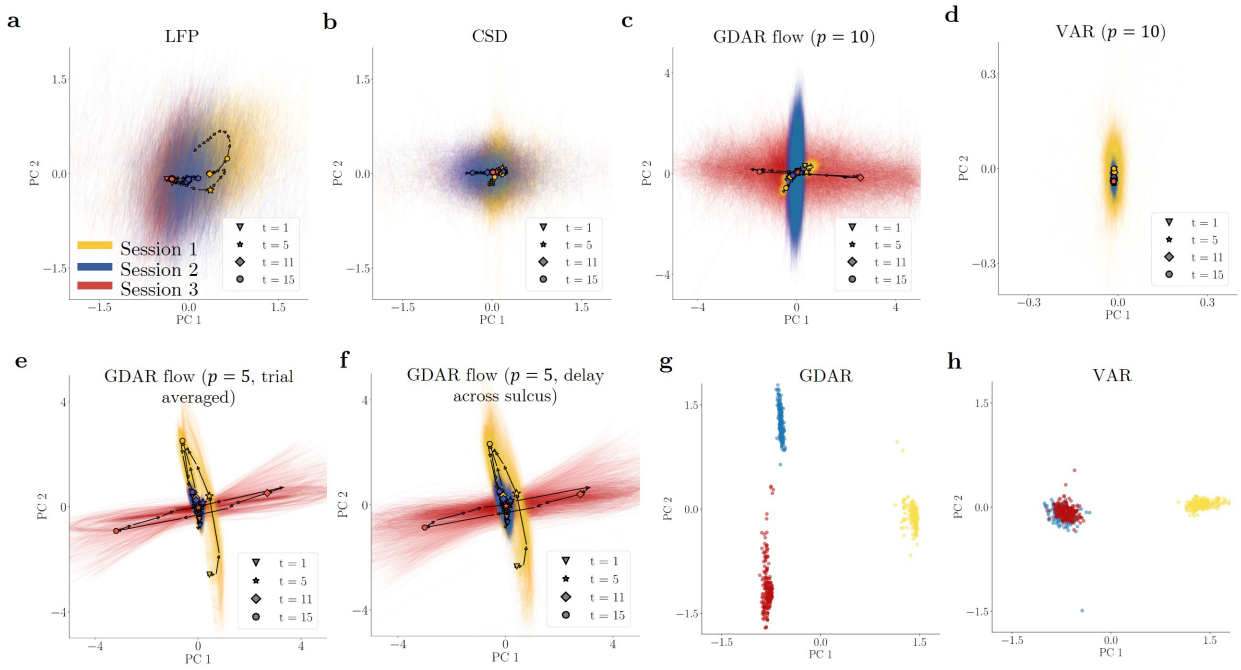
76. Dadarlat, M. C., O'Doherty, J. E. & Sabes, P. N. A learning-based approach to artificial sensory feedback leads to optimal integration. *Nat. Neurosci.* **18**, 138–144 (2015).

77. Van Acker, G. M., III, Luchies, C. W. & Cheney, P. D. Timing of Cortico-Muscle Transmission During Active Movement. *Cereb. Cortex* **26**, 3335–3344 (2016).

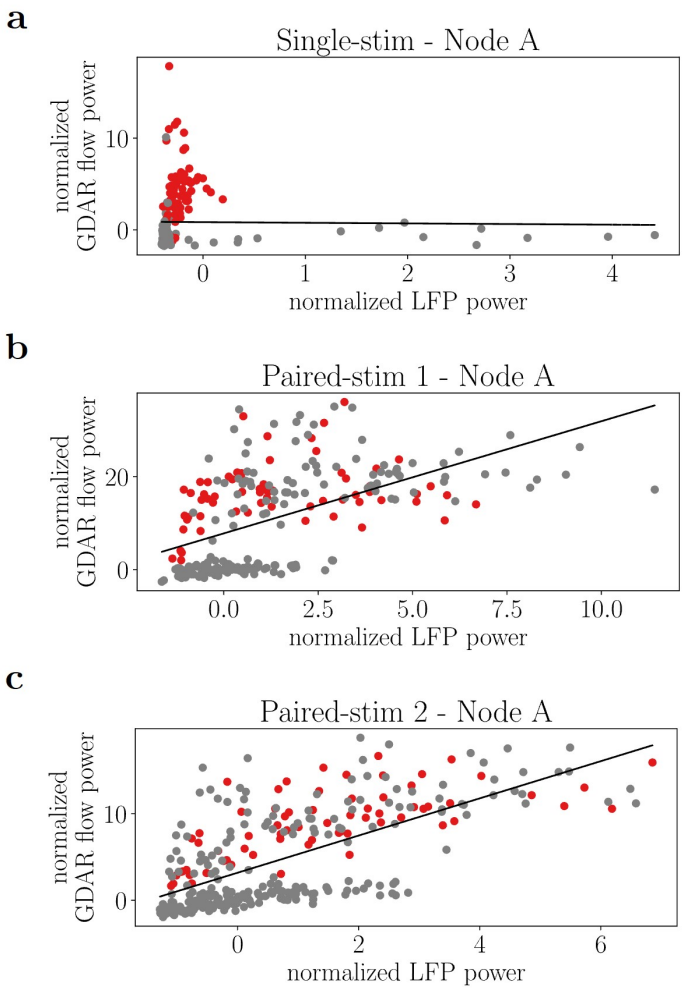
Extended data figures



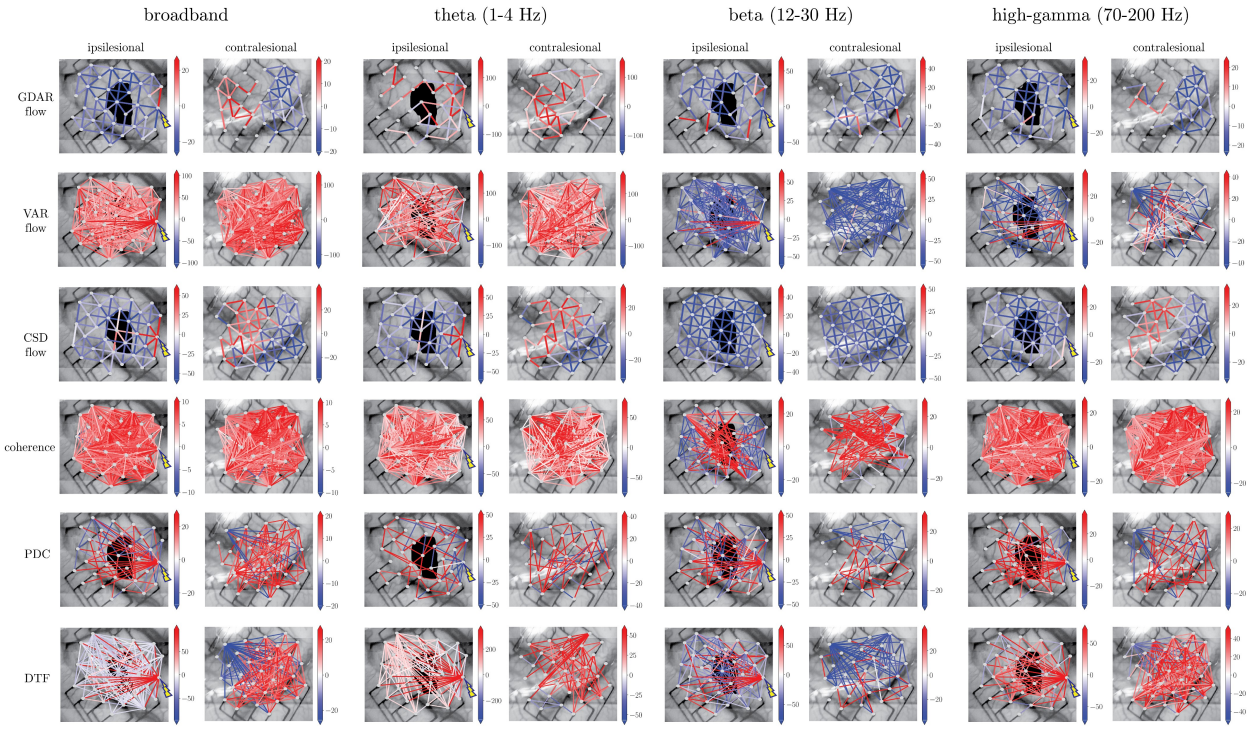
Extended Fig. 1: (a) 10 randomly connected 16-node networks used for conducting the simulations in Fig. 2. Each graph was used to generate 10 independent simulation trials. (b), (c) Power spectral densities of simulated field potentials and ground truth flow, respectively. The black line shows the average across all edges and trials. The gray shaded area indicated one half of the standard deviation. The simulation parameters produce a strong oscillation around 18Hz. The steep drop-off above 400 Hz can be attributed to the 8^{th} order Chebyshev filter that was used for downsampling the data to a sampling frequency of 1 kHz. (d), (e) Pearson correlation coefficient (CC) of GDAR, VAR, and eVAR model on 16-node grid graph and 7-node locally connected graph for various mode orders. The CC is pooled from 100 simulation trials with varying excitatory coupling parameters (see Methods) for each graph. Markers indicate whether the GDAR model significantly outperforms the eVAR model (Wilcoxon ranked-sum test, $p \leq 0.001$). The GDAR model significantly outperforms the other two models for all tested graphs given a sufficiently high model order. (f) CC between unidirectional ground truth and eVAR flow, as well as bidirectional ground truth and eVAR flow for various model orders. There is no significant performance difference between unidirectional and bidirectional eVAR flow for any model order. For most model orders, the unidirectional eVAR flow yields slightly higher median CCs than the bidirectional eVAR flow.



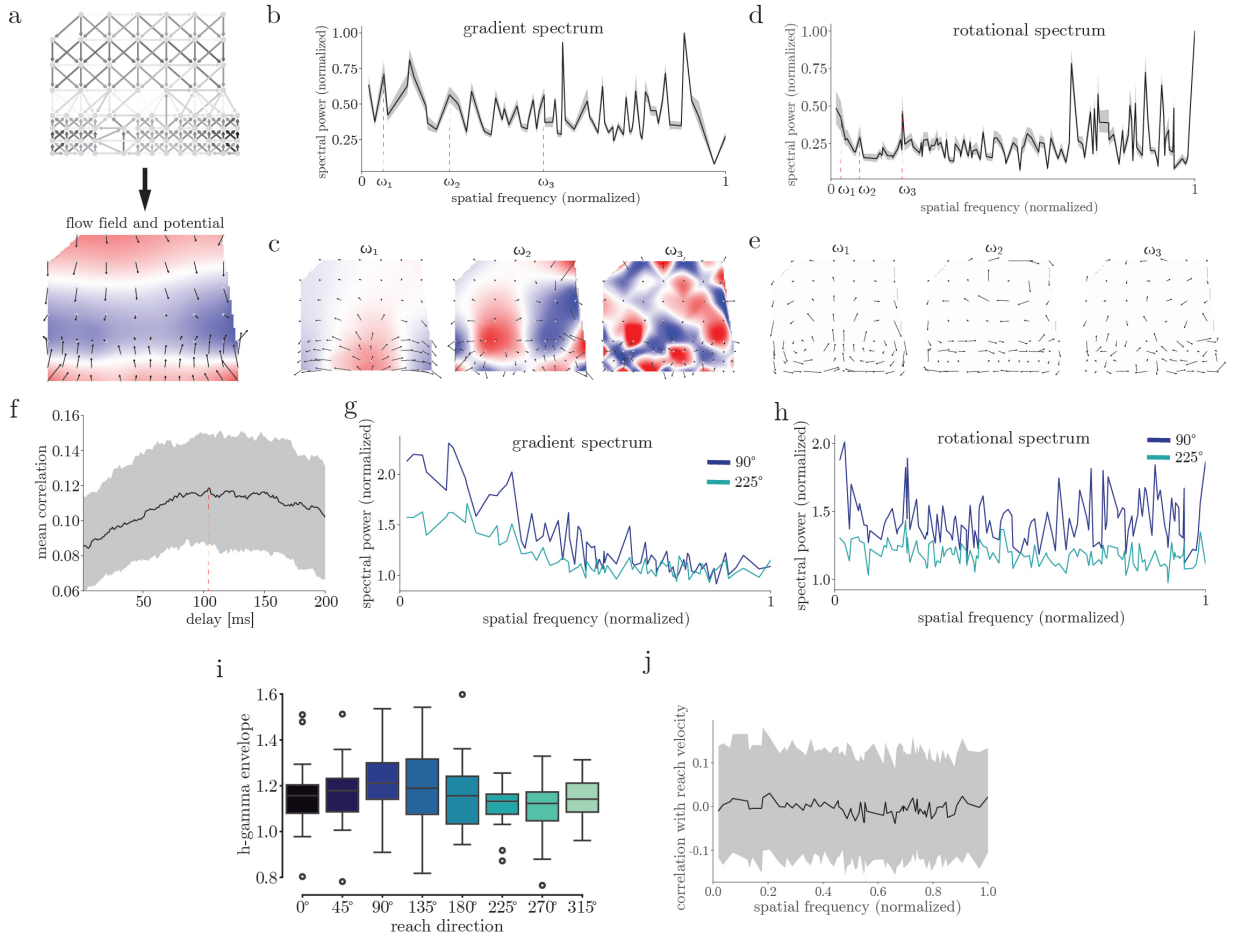
Extended Fig. 2: (a)-(f) Stimulation evoked dynamics using different signals and modeling approaches similar to the plots in Fig. 3e. Neither LFP (a) nor classical CSD (b), i.e., the second spatial derivative computed via the graph Laplacian, show significant temporal dynamics that are distinct between the sessions. Using a 10^{th} order GDAR model (c) results in reduced separability between sessions when compared to a 5^{th} order model shown in Fig. 3e (similarly for a 10^{th} order VAR model (d)). This is likely a result of the time scale of the paired stimulation (the onset of the second laser pulse occurs 5 ms after the offset of the first pulse) causing a mixing of the effects of the two lasers when fitting the models. (e) Averaging the flow signal over 20 consecutive trials before computing the low dimensional embedding does not a have negative affect on the dynamics. (f) PC reduced GDAR flow dynamics when constraining edges crossing the sulcus between M1 and S1 to exhibit a minimum signal propagation delay of 4ms. The dynamics are almost identical to the dynamics from the unconstrained GDAR model (Fig. 3e), suggesting that the stimulation induced communication dynamics uncovered by our model are robust to such constraints. (g), (h) The parameters of the 5^{th} order GDAR and VAR model for all segments, blocks, and sessions were stacked into a single matrix and projected onto its first two PCs. Each dot represents the PC reduced parameters of a single 10 s segment used for model fitting. The parameters from the three different sessions are well separated in this low dimensional subspace for the GDAR model, but not for the VAR model where Session 2 and 3 are not separable.



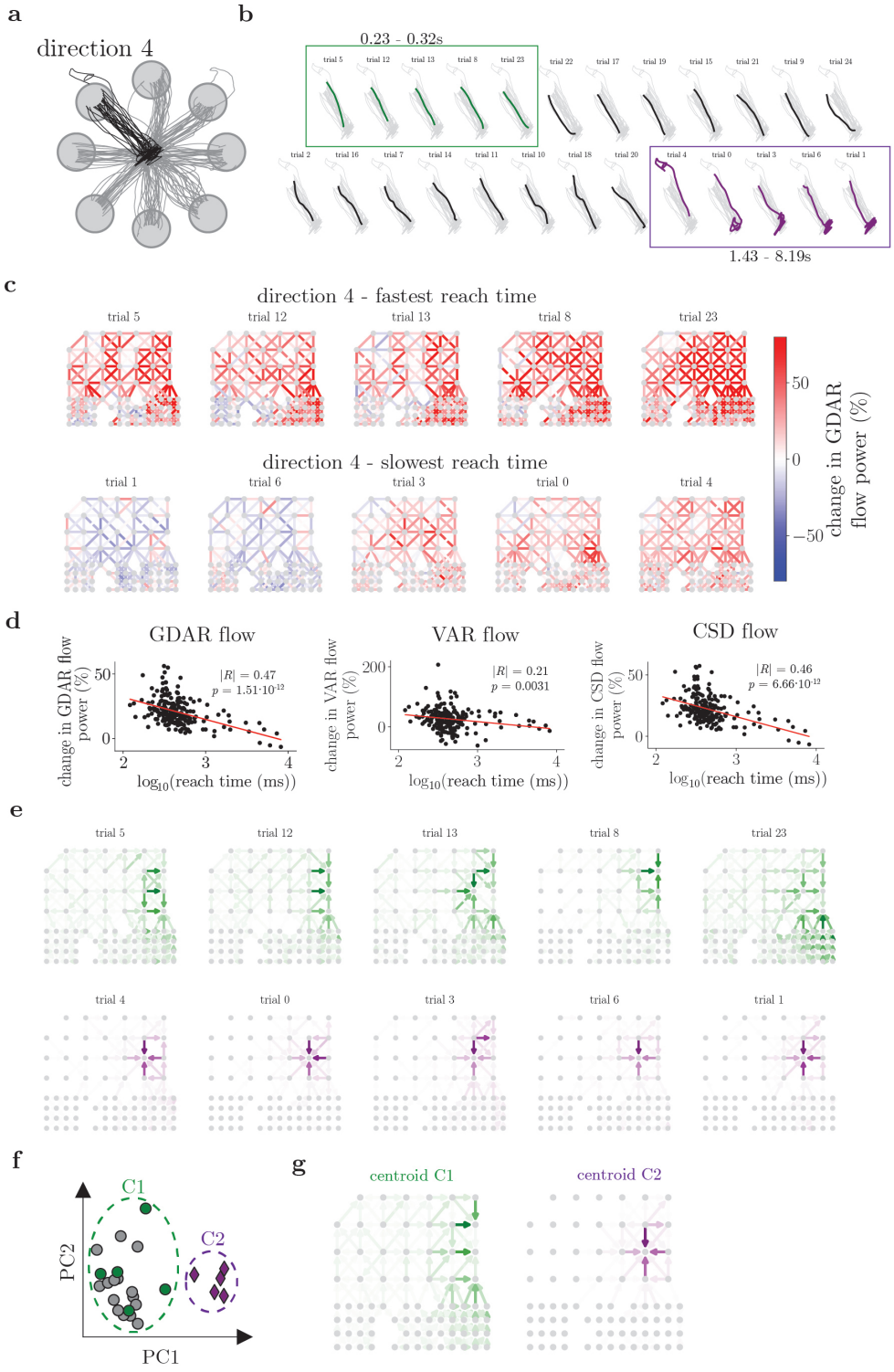
Extended Fig. 3: Relation between average gamma GDAR flow and LFP power for the stimulation electrodes in the Utah array dataset. The best linear regression lines are also shown (Single-stim: $slope = -0.0692$, $p = 0.827$; Paired-stim 1: $slope = 2.41$, $p = 1.05e-19$; Paired-stim 2: $slope = 2.15$, $p = 2.1e-43$). The two paired stim sessions show a strong linear relation between LFP and average GDAR flow power. Nevertheless, the GDAR flow power increases due to stimulation beyond what can be explained by linear changes in LFP power.



Extended Fig. 4: Replication of Fig. 5c and d for different frequency bands. Changes in communication are strongly dependent on frequency, as well as the methods used for estimating it. However, some features such an increase in communication in areas near the stimulation location in the ipsilesional hemisphere can be observed across multiple frequency bands and methods.



Extended Fig. 5: Extended plots for reach data shown in Fig. 6. (a) To better visualize the spatial frequency characteristic of the gradient and rotational components, we compute the flow field (net-flow magnitude and direction at each node) and potential field (divergence of flow at each node) (bottom) from the GDAR flow basis vectors V_{grad} and V_{rot} (top). The example in (a) shows the gradient flow component corresponding to the 6^{th} lowest frequency. (b) Average spectrum of the gradient flow component (median and interquartile range) across all time points, trials and directions. Note that the spectrum largely resembles white noise with a few stronger spectral components at the low, mid, and high frequencies. (c) Flow and potential fields for three spectral components marked in (b). As the frequency increases, the flow becomes more disorganized with an increasing number of local sources and sinks and the overall divergence (measure of spatial frequency) increases. (d) and (e) Same as (b) and (c), but for rotational flow. The average rotational flow spectrum is primarily marked by an increase in power for some high-frequency components. (f) Correlation (median and interquartile range) between average gradient flow power (averaged over all spatial frequencies) and reach velocity for different delays between the neural signal and the recorded velocity. The maximum correlation occurs at 104 ms, which we assume to be transmission delay between motor commands in the brain and observable movements. (g) and (h) Average gradient and rotational power spectra for the 90° (up) and 225° (bottom-left) directions. The 90° direction shows a substantially larger increase in the low frequency gradient flow power, as well as rotational flow power across almost all frequencies than the 225° direction, highlighting the directional tuning of the GDAR flow. (i) Directional tuning curves (quartiles, 1.5 times interquartile range, and outliers) for envelope of high gamma ($70-200$ Hz) filtered local field potential signal averaged over all recording electrodes. The same trend as for the GDAR flow alignment index (Fig. 6e) and average rotational flow power (Fig. 6f) can be observed, however differences between directions are not significant. (j) Correlation (median and interquartile range) between average gradient flow power and reach velocity during the last 100 ms prior to the go-cue. The median correlation of zero suggest that there is no residual movement occurring during that period that could explain the strong correlation between the alignment index and the reaction time in Fig. 6d. The monkey was instructed to hold his finger still on the center of the screen.



Extended Fig. 6: Applying the GDAR model to center-out reach data to study trial-by-trial variability of neural communication. (a) To analyze the trial-by-trial variability, we mainly focus on direction 4 (top-left) as it showed a greater reach variability than other directions. (b) The 25 reach trajectories of direction 4 were grouped into the five fastest, 15 normal, and five slowest trials. The reach times during the fastest (slowest) trials range from $0.23 - 0.31$ *s* $(1.43 - 8.19$ *s). The five slowest trajectories are very jagged either at the beginning or end of the reach. (c) Similar to*

the procedure described in Fig. 5, changes in GDAR flow power during reach compared to baseline were computed for the high-gamma band ($70-200$ Hz) and are plotted for the five fastest, and slowest reach trials, respectively. Fast reach t rials show a significantly stronger increase in GDAR flow power than slow trials. (d) The average change in high-gamma GDAR, VAR and CSD flow power is plotted against the reach time for each of the 200 trials (all directions). While all metrics show a significant negative relation between the log of the reach time and the change in average flow power, the GDAR model shows the highest $|R|$, closely followed by the CSD flow. (e) Dominant GDAR flow patterns (obtained via principal component analysis) for the five fastest and slowest reach trials of direction 4. (f) These GDAR flow patterns naturally form two clusters where C1 contains fast and average trials and C2 the five slowest trials. (g) Cluster centroids of C1 and C2. Fast reach trials correspond to a more coordinated GDAR flow pattern that involves larger parts of the network, whereas slow reach trials exhibit a flow pattern that is mainly centered around a single node.

Supplementary Material:

Supplementary Table 2: p-values comparing GDAR and eVAR model for differences in Pearson correlation coefficients (PCC) for the raw flow, PCC of the power spectral density (PSD) of the raw flow, and dynamical similarity analysis (DSA) dissimilarity score for 16-node random graphs in Fig. 2b, d, and e, as well as PCC of the raw flow for the 16-node grid graph in Extended Fig. 1d. We used one-sided Wilcoxon rank-sum tests with the alternative hypothesis that the GDAR model outperforms the eVAR model.

Model order	PCC raw flow – 16-node random graphs	PCC PSD of raw flow – 16-node random graphs	DSA dissimilarity – 16-node random graphs	PCC raw flow – 16-node grid graph
2	0.999999988	0.74	0.373	1.0
4	0.93	0.51	0.157	1.0
6	0.617	0.167	0.397	1.0
8	0.509	0.201	0.728	1.0
10	0.179	0.132	0.916	0.955
12	0.0104	0.0693	0.99	0.19
14	$6.17 \cdot 10^{-8}$	0.0133	0.99939	$8.05 \cdot 10^{-8}$
16	$4.32 \cdot 10^{-19}$	$4.41 \cdot 10^{-4}$	0.999995	$2.35 \cdot 10^{-23}$
18	$1.17 \cdot 10^{-27}$	$1.12 \cdot 10^{-5}$	0.9999988	$8.21 \cdot 10^{-40}$
20	$1.41 \cdot 10^{-28}$	$1.53 \cdot 10^{-6}$	0.9999967	$3 \cdot 10^{-44}$
22	$2.13 \cdot 10^{-25}$	$1.12 \cdot 10^{-7}$	0.987	$2.33 \cdot 10^{-42}$
24	$9.06 \cdot 10^{-23}$	$4.88 \cdot 10^{-7}$	0.806	$4.59 \cdot 10^{-43}$
26	$1.47 \cdot 10^{-22}$	$2.88 \cdot 10^{-6}$	0.536	$1.25 \cdot 10^{-46}$
28	$1.2 \cdot 10^{-21}$	$1.03 \cdot 10^{-5}$	0.147	$6.56 \cdot 10^{-50}$
30	$9.54 \cdot 10^{-19}$	$1.5 \cdot 10^{-5}$	0.134	$4.58 \cdot 10^{-49}$

Supplementary Table 3: p-values for differences in Pearson correlation coefficients shown in Extended Fig. 1e and f. For "GDAR vs. eVAR", the alternative hypothesis is that the GDAR model outperforms the eVAR model. For "eVAR bidirectional vs. eVAR unidirectional", the alternative hypothesis is that the bidirectional outperforms the unidirectional VAR flow. We used one-sided Wilcoxon rank-sum tests.

Model order	GDAR vs. eVAR	eVAR bidirectional vs. eVAR unidirectional
1	0.118	0.985
2	0.9997	0.99904
3	0.176	0.8
4	0.719	0.987
5	0.121	0.954
6	0.457	0.961
7	0.00497	0.871
8	0.0161	0.834
9	$8.26 \cdot 10^{-8}$	0.793

10	$8.24 \cdot 10^{-8}$	0.619
11	$3.06 \cdot 10^{-10}$	0.662
12	$6.31 \cdot 10^{-9}$	0.577
13	$2.1 \cdot 10^{-10}$	0.602
14	$8.03 \cdot 10^{-10}$	0.508
15	$5.28 \cdot 10^{-9}$	0.491

Supplementary Table 4: p-values (one-sided Wilcoxon rank-sum test) for tuning curves in Fig. 6e and f. Only significant comparisons are shown. All p-values are corrected for multiple comparisons via the Bonferroni method. 0° and 90° correspond to the right and top direction, respectively.

Gradient spectrum alignment index (Fig. 6e)	Rotational spectrum average spectral power (Fig. 6f)
90° - 0°: 0.0165	45° - 225°: 0.0165
90° - 225°: $4.42 \cdot 10^{-5}$	90° - 225°: $7.04 \cdot 10^{-6}$
90° - 270°: 0.00111	135° - 225°: $4.42 \cdot 10^{-5}$
	180° - 225°: 0.00455
	315° - 225°: $4.42 \cdot 10^{-5}$

Supplementary Table 5: p-values (one-sided Wilcoxon rank-sum test) for differences in generalization gap between GDAR and VAR model for four electrophysiological datasets in Fig. 7c.

Opto	Reach	Rest Utah	Rest ECoG
$8.27 \cdot 10^{-16}$	$1.76 \cdot 10^{-60}$	$4.05 \cdot 10^{-40}$	$2.37 \cdot 10^{-5}$

Supplementary Notes:

Modeling signal propagation delay across sulcus for optogenetic stimulation experiment: The GDAR model can also be adapted to model longer signal propagation paths between specific nodes in the network. For example, this may be the case for connections across the sulcus that separates M1 and S1. To model this, we have constrained the GDAR model to enforce a minimum propagation delay of 4 ms for all edges that connect nodes in M1 with nodes in S1 (see Methods). The evolution of the GDAR flow for Session 1 averaged over all trials is shown in Supplementary Video 2 and is similar to the dynamics observed in Fig. 3d and Supplementary Video 1, but with noticeably less flow across the sulcus. The PC reduced flow dynamics across all sessions are shown in Extended Fig. 2f and are almost identical to the dynamics shown in Fig. 3e, suggesting that adding signal propagation constraints to the model does not negatively impact the sensitivity of the GDAR flow signal to the stimulation parameters.