LETTER **Communicated by Theoden Netoff**

Fragility in Dynamic Networks: Application to Neural Networks in the Epileptic Cortex

Duluxan Sritharan
dsritha1@jhu.edu
Sridevi V. Sarma
sree@jhu.edu
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, U.S.A.

Epilepsy is a network phenomenon characterized by atypical activity at the neuronal and population levels during seizures, including tonic spiking, increased heterogeneity in spiking rates, and synchronization. The etiology of epilepsy is unclear, but a common theme among proposed mechanisms is that structural connectivity between neurons is altered. It is hypothesized that epilepsy arises not from random changes in connectivity, but from specific structural changes to the most fragile nodes or neurons in the network. In this letter, the minimum energy perturbation on functional connectivity required to destabilize linear networks is derived. Perturbation results are then applied to a probabilistic nonlinear neural network model that operates at a stable fixed point. That is, if a small stimulus is applied to the network, the activation probabilities of each neuron respond transiently but eventually recover to their baseline values. When the perturbed network is destabilized, the activation probabilities shift to larger or smaller values or oscillate when a small stimulus is applied. Finally, the structural modifications to the neural network that achieve the functional perturbation are derived. Simulations of the unperturbed and perturbed networks qualitatively reflect neuronal activity observed in epilepsy patients, suggesting that the changes in network dynamics due to destabilizing perturbations, including the emergence of an unstable manifold or a stable limit cycle, may be indicative of neuronal or population dynamics during seizure. That is, the epileptic cortex is always on the brink of instability and minute changes in the synaptic weights associated with the most fragile node can suddenly destabilize the network to cause seizures. Finally, the theory developed here and its interpretation of epileptic networks enables the design of a straightforward feedback controller that first detects when the network has destabilized and then applies linear state feedback control to steer the network back to its stable state.

Neural Computation **26, 2294–2327** (2014)
doi:10.1162/NECO_a_00644

1 Introduction

Epilepsy is a neurological disease that affects over 50 million individuals worldwide (de Boer, Mula, & Sander, 2008), manifesting in recurrent seizures. During seizures, there is evidence of abnormal activity in cortical neurons, including more heterogeneity in firing (Truccolo et al., 2011), transient increases or decreases in firing rate (Truccolo et al., 2011), and synchronization in population oscillations (Uhlhaas & Singer, 2003). The etiology of epilepsy is unclear, and as a highly heterogeneous disorder, a variety of proposed mechanisms acting at different scales for different variants have been put forth. Studies have implicated genetic anomalies (Jeffrey, 2003), axo-axonic gap junctions (Traub et al., 2001), neurotransmitter imbalance (Bradford, 1995), loss of inhibitory chandelier cells in cortex (DeFelipe, 1999), atypical axonal sprouting from layer V pyramidal cells (Jin, Prince, & Huguenard, 2006), infection (Lancman & Morris, 1996), or physical injury (Annegers & Coan, 2000).

In the case of physical injury, one hypothesis is that posttraumatic epilepsy arises because of homeostatic synaptic scaling. In healthy cortex, homeostatic plasticity mechanisms dynamically adjust synaptic strengths to promote stability in average neuronal activity levels (Turrigiano & Nelson, 2004), but in damaged cortex, this same mechanism is posited to cause epileptiform activity. For example, experimental evidence suggests that glial TNFα, which has been shown to facilitate homeostatic synaptic scaling (Stellwagen & Malenka, 2006), enhance expression of AMPA receptors, and maintain synaptic strength at excitatory synapses, may modulate response to neural injury as well (Beattie et al., 2002). Modeling studies seem consistent with this hypothesis. In an isolated neocortex model of epileptogenesis, Houweling, Bazhenov, Timofeev, Steriade, & Sejnowski (2005) found that homeostatic synaptic plasticity following complete cortical deafferentation, as may occur in trauma, produces synchronized bursting that appears epileptiform in nature. Volman, Bazhenov, and Sejnowski (2011) determined that homeostatic synaptic plasticity produces a more pronounced posttraumatic elevation of paroxysmal activity if the deafferentation is not topologically diffuse and that the geometric organization of a small population of trauma-surviving neurons is important in determining the properties of posttraumatic interictal epileptiform discharges (Volman, Sejnowski, & Bazhenov, 2011).

Although the details vary, a common thread among all etiologies is that the effective coupling between neurons or populations of neurons is altered, resulting in aberrant and perhaps *unstable* epileptiform spiking. For this reason, epilepsy is understood to be a network phenomenon, and network models are often used to study the relationship between structural connections and functional activity.

Network models consist of a set of nodes with internal dynamics, connected by edges that define the nature of nodal interactions. For neural

networks, nodes are neurons or populations of neurons, and edge weights represent synaptic connection strengths. Negative outgoing edge weights from a node indicate that neurons in that node are inhibitory, while positive outgoing weights from a node indicate that the neurons within are excitatory. Figure 1 shows how two possible mechanisms of epilepsy operate at the level of cortical columns. Chandelier cell loss (DeFelipe, 1999) and abnormal axonal sprouting from layer V pyramidal cells (Jin et al., 2006), can be modeled by altering specific edge weights in the graphical representation of the network. For example, abnormal axonal sprouting of an excitatory cell can be modeled in two ways: (1) by increasing its outbound edge weights or its recurrent loop edge weight (structural features), which increases the firing activity (a functional feature) of its downstream targets, or (2) by decreasing inbound inhibitory edge weights or increasing inbound excitatory edge weights, which bolsters the cell's excitatory influence on its targets. Similarly, the loss of an inhibitory chandelier cell can be modeled in two ways: (1) by decreasing its outbound edge weights, which disinhibits its downstream targets thereby increasing their firing activity, or (2) by decreasing inbound excitatory edge weights or increasing inbound inhibitory edges weights, which reduces the cell's inhibitory influence on its targets. Linking such structural changes to corresponding functional activity in neural networks is nontrivial and contingent on the dynamics and topology of the network.

Recent theoretical studies have attempted to link structure to function. Zhao, Beverlin, Netoff, and Nykamp (2011) considered two-edge structural motifs as the basic building blocks of a network and investigated how the relative concentration and coupling of these motifs affect functional features such as synchrony. The observability and controllability (functional features) of neuronal network motifs have also been analyzed as a first step toward reconstructing networks and controlling their dynamics (Whalen, Brennan, Sauer, & Schiff, 2012). Kramer and Cash (2012) have considered epilepsy as a disorder of cortical networks and have closely examined functional features without linking these back to structural features. They found that functional cortical networks fragment and coalesce during the course of seizures (Kramer et al., 2010), with characteristic topological features and dynamical mechanisms in seizure onset (Kramer, Kolaczyk, & Kirsch, 2008), propagation (Kramer, Szeri, Sleigh, & Kirsch, 2007), and termination (Kramer et al., 2012). This study proposes a link between structure and function in more general epileptic networks (as opposed to motifs) using matrix theory and linear systems theory.

We hypothesize that aberrant functional activity observed during seizures arises not from random changes in network structure but from the disruption of the most fragile nodal connections in the network. Fragility of a network node is defined as the minimum energy perturbation (on functional connections to its neighbors) required to destabilize the network. To test this hypothesis, a minimum energy perturbation theory on linear

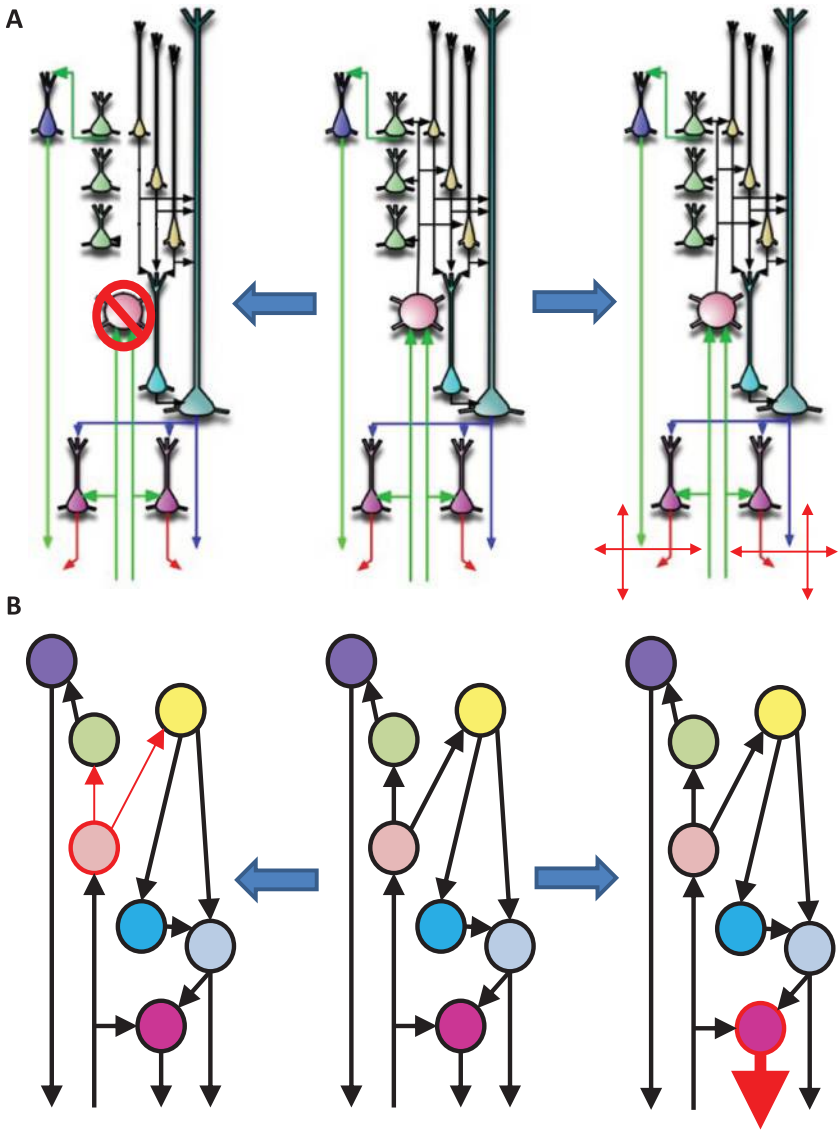

Figure 1: Graph representation of cortical column pathologies. (A) Schematic of an idealized cortical column and (B) corresponding graph representation with cells of the same type grouped into a single node. Assume the light pink cell in the healthy column (center) is a chandelier cell. Chandelier cell loss (left) can be represented by reducing the weights on outbound edges. Similarly, abnormal axonal sprouting (right) from deep pyramidal cells can be represented as a strengthened edge weight. (Adapted from George & Hawkins, 2009, Creative Commons Attribution License.)

networks is developed for physiologically relevant classes of perturbations that affect different subnetwork topologies. In particular, the fragility of each node in a network is computed when only its outgoing connections can change and only its incoming connections can change. The theory is then applied to a probabilistic nonlinear neural network model with nodes that operate at a stable fixed point. That is, if a small stimulus is applied to this stable network, the activation probability of each neuron responds transiently but eventually returns to the fixed point. When the perturbed network is destabilized, the activation probabilities shift to larger or smaller values or oscillate indefinitely. Finally, the structural modifications, or the changes in the synaptic weights of the most fragile node that achieve the functional perturbation, are derived.

Results suggest that the most fragile nodes in the neural network are either excitatory neurons that become more active or inhibitory neurons that become less active. This is consistent with abnormal axonal sprouting of excitatory neurons and loss of inhibitory cells observed in epileptic cortical tissue. Furthermore, simulation of the unperturbed and perturbed network reflects neuronal activity observed in epilepsy patients implanted with electrodes before and during seizure, respectively. In particular, ictal firing rates of neurons either increase, decrease or oscillate.

Examining the topology of these fragile subnetworks, the structure of the perturbations required to disrupt them, and the nature of the resulting network instability may elucidate the relation between structure and function in the epileptic brain. Furthermore, the qualitative changes in network dynamics due to destabilizing perturbations, including the emergence of an unstable manifold or a stable limit cycle, may be indicative of neuronal or population dynamics during seizure. That is, the epileptic cortex is always on the brink of instability, and minute changes in synaptic weights can suddenly destabilize the network to cause seizures. Finally, the theory developed here and its interpretation of epileptic networks enables the design of a straightforward feedback controller that first detects when the network has destabilized and then applies linear state feedback control to steer the network back to its stable state (Ehrens, Sritharan, & Sarma, 2014).

2 Methods

In this section, network connectivity is first described, followed by a brief overview of linear networks and notions of stability from linear systems theory. Network fragility is then defined as the minimum energy perturbation required to cause instability in a linear network. The fragility of nodes is derived using matrix theory for two perturbation structures that affect different subnetwork topologies. Next, a probabilistic rate model of a neuronal network that links functional activity to structural connectivity is introduced, and a procedure is developed for determining the structural

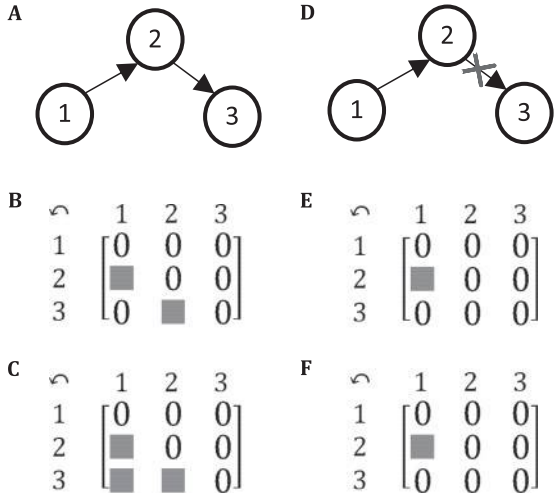

Figure 2: Connectivity of a sample network. (A/D) Graph representation, (B/E) structural connectivity, and (C/F) functional connectivity of a sample network before and after a modification removing the edge between the second and third nodes. Shaded entries represent nonzero entries. The removal of a single edge corresponds to the zeroing of a single structural connection and two functional connections.

disruptions required to reproduce the functional instability corresponding to the minimum energy perturbation.

2.1 Structural Versus Functional Connectivity. The relation between structure and function in a network was previously introduced, and here a formalism is presented to capture these notions more explicitly. Consider a network of three nodes as shown in Figure 2A. Each node has internal dynamics that are modulated by external inputs to produce outputs propagated in the direction of the arrows. For the moment, the internal dynamics will be ignored, and only the relationship between the nodes will be considered. Node 2 receives direct input from node 1, and node 3 receives direct input from node 2. The structural connectivity, the physical connections that define adjacent nodes, can be represented as an adjacency matrix, as shown in Figure 2B.

From a functional point of view, however, it is clear that while the activity of the second node depends only on the activity of the first, the activity of node 3 depends on the activity of both nodes 1 and 2—the activity of node 1 is relayed through node 2 to affect node 3. Therefore, the functional connectivity, which is the effect of one node's activity on another's, is different from the structural connectivity, which simply reflects the existence of direct

connections. The functional connectivity of this network can also be represented as an adjacency matrix, as shown in Figure 2C.

Now consider a modification to the existing network structure in which the arrow from node 2 to node 3 is removed as in Figure 2D. The corresponding modification to the adjacency matrix of the structural connectivity is the zeroing of a single element, as shown in Figure 2E. The functional connectivity matrix changes by two elements however; without the connection between nodes 2 and 3, neither node 2 nor node 1 can affect the activity of node 3 (see Figure 2F). It is clear from this example that local changes in structural connectivity can have more far-reaching effects in terms of functional connectivity. Multiple structural configurations may also give rise to the same functional behavior.

In the context of neurological diseases, what is of interest is not the structural connectivity per se but the interdependence of the activity of nodes, i.e., the functional connectivity that arises from the structural connectivity. In epilepsy, the obvious defect is in the functional connectivity, the aberrant behavior of neurons and populations that manifests during seizures, and the goal is to link this back to some abnormality in the structural connectivity. This is a nontrivial problem that depends strongly on both network topology and internal nodal dynamics.

2.2 Stability of Linear Networks. First, consider a linear system that evolves in continuous time as equation 2.1 with state vector $\mathbf{x}(t) \in \mathbb{R}^N$ and state matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ with eigenvalues $\lambda_{1\ldots N} \in \mathbb{C}$ where $Re\{\lambda_1\} \geq \ldots \geq Re\{\lambda_N\}$.

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t). \tag{2.1}$$

The state matrix can be viewed as an adjacency matrix representation of the functional connectivity of a network of N nodes, whose dynamics are linear and captured in the evolution of the state vector. The elements in the state vector are some metric of the activity of each node. Element A_{ij} indicates how the activity of node j, $x_j(t)$, affects the future activity of node i, $\dot{x}_i(t)$. Element A_{ii} is an autofeedback term, representing a first-order approximation to the internal dynamics of node i. More generally, the ith row of $\mathbf{A}$ dictates the network's cumulative functional effect on node i, while the jth column captures the functional effect of the activity of node j on the entire network.

From linear systems theory, the system is said to be asymptotically stable about a fixed point, $\widehat{\mathbf{x}}$, if $\mathbf{x}(t)$ converges to $\widehat{\mathbf{x}}$ as $t \to \infty$ for all initial conditions. This implies that the activity of the nodes remains at a baseline value and responds transiently to external inputs before recovering. In terms of the matrix representation, stability is equivalent to the eigenvalues of $\mathbf{A}$ being in the open left-hand plane, that is, $Re\{\lambda_i\} < 0$ for all $i = 1, 2, \ldots, N$.

This is more evident when equation 2.1 is cast in modal form, as shown in equation 2.2, where $c_i \in \mathbb{C}$ are set according to initial conditions and $\mathbf{v_i} \in \mathbb{C}^N$ are the eigenvectors of $\mathbf{A}$ corresponding to λ_i. Assume the system has already been transformed so the fixed point is at the origin ($\widehat{\mathbf{x}} = \mathbf{0}$) and that there are n distinct eigenvalues of $\mathbf{A}$. Then the solution to the linear system can be written as

$$\mathbf{x}(t) = \sum_{i=1}^{N} c_i e^{\lambda_i t} \mathbf{v_i}. \tag{2.2}$$

If $Re\{\lambda_1\} < 0$, the real components of the other eigenvalues are also negative by virtue of their ordering, and all the terms in equation 2.2 decay so that the system settles to its baseline activity (transformed to be at the origin). If $\lambda_1 = 0$, the other eigenvalues have negative real components and their corresponding terms vanish over time, so the state settles somewhere along the first eigenvector based on initial conditions: $\mathbf{x}(t) \to c_1 \mathbf{v_1}$. Therefore, the network gets stuck in some pattern of activity instead of decaying to its baseline. If the elements in the state vector represent spiking rates, this may be analogous to tonic spiking or silenced neurons. Finally, if $\lambda_{1,2} = \pm j\omega$, then for large t, the activity of each node oscillates around its baseline without ever decaying. This may be a suitable representation for spiking rates entrained to an oscillation.

A linear system becomes unstable if there is a change to the state matrix so that the conditions on the eigenvalues (real parts being negative) are no longer met. This change can be modeled as an additive perturbation, $\mathbf{\Delta}$, to the state matrix so that $\mathbf{A} + \mathbf{\Delta}$ replaces $\mathbf{A}$ in the state equation, equation 2.1, to give equation 2.3:

$$\dot{\mathbf{x}}(t) = (\mathbf{A} + \mathbf{\Delta})\mathbf{x}(t). \tag{2.3}$$

This notion of instability can be adopted for networks whose dynamics are linear or can be approximated as linear in some regime. For such networks, the eigenvalues of the functional connectivity matrix determine the stability properties.

2.3 Perturbations and Network Fragility. A variety of perturbation matrices in equation 2.3 can push the original network in equation 2.1 into instability. Based on the structure of the perturbation (e.g., $\mathbf{\Delta} \in \mathbb{R}^{N \times N}$ or $\mathbf{\Delta} \in diag\{\mathbb{R}^N\}$ as in Figure 3A) and which elements are preferentially affected, different perturbation strengths (measured by a matrix norm) are required to cause the perturbed system to become unstable. Network fragility is defined here as the magnitude of the minimum energy perturbation required to push the network to the brink of instability. If a large-magnitude perturbation is required, the network is more robust, while small energy perturbations correspond to a fragile network. The elements that are

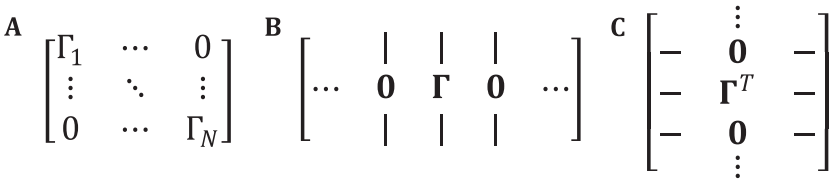

Figure 3: Three perturbation topologies in $\mathbb{R}^{N\times N}$. (A) Diagonal perturbations only disrupt autofeedback terms. (B) Column perturbations have nonzero entries in a single column, while (C) row perturbations have a single row with nonzero entries.

modified by the minimum energy perturbation define the edges of the most fragile subnetwork.

Whole network perturbations or diagonal perturbations will not be considered because they are biologically unlikely. It is improbable that a whole host of functional network connections needs to be modified to cause aberrant behavior or that changes in autofeedback occur in different neurons independently and in isolation without affecting functional connectivity in any other way.

Network fragility will therefore be derived for row perturbations (see Figure 3C) corresponding to changes in the effect of incoming projections to a single neuron or population. Row perturbations may be a plausible perturbation structure to reflect anomalies in a neuron or population that affects how it integrates inputs. The column perturbation case can be derived similarly (Sritharan, 2013).

2.4 Structured Perturbation Problem. In this section, the structured row perturbation problem is formulated and solved.

2.4.1 Theorem. Suppose $\mathbf{A} \in \mathbb{R}^{N\times N}$ *represents the state transition matrix of a stable linear network. Then the minimum 2-induced norm additive row perturbation,* $\widehat{\mathbf{\Delta}} \in \Lambda_r = \mathbf{e_k}\mathbf{\Gamma}^T$ and $\mathbf{\Gamma} \in \mathbb{R}^N$, *which will destabilize the linear network (or, more precisely, place an eigenvalue of* $\mathbf{A} + \widehat{\mathbf{\Delta}}$ *at* $\lambda = j\omega, 0 < \omega \le \omega_0$*), is given by*

$$\widehat{\mathbf{\Delta}} = \mathbf{e}_{\widehat{\mathbf{k}}}\big[\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{b}\big]^T \tag{2.4}$$

where

$$\mathbf{B}(\omega, \widehat{k}) = \begin{bmatrix} Im\left\{\mathbf{e}_{\widehat{\mathbf{k}}}^T(\mathbf{A} - j\omega\mathbf{I})^{-T}\right\} \\ Re\left\{\mathbf{e}_{\widehat{\mathbf{k}}}^T(\mathbf{A} - j\omega\mathbf{I})^{-T}\right\} \end{bmatrix} \tag{2.5a}$$

$$\mathbf{b} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \tag{2.5b}$$

and $\widehat{k}$ is the row where the perturbation is applied. $\mathbf{e_k} \in \mathbb{R}^N$ *is the kth elementary basis vector.*

2.4.2 Proof.

$$\widehat{\mathbf{\Delta}} = \underset{\mathbf{\Delta}\in\Lambda_r,\omega}{\operatorname{argmin}}\{\|\mathbf{\Delta}\|_2 \mid \exists i : \lambda_i(\mathbf{A}+\mathbf{\Delta}) = j\omega, \forall i : Re\left\{\lambda_i(\mathbf{A})\right\} < 0, i \in 1\ldots N, \mathbf{A} \in \mathbb{R}^{N\times N}\} \tag{2.6}$$

Since $j\omega$ is an eigenvalue of $\mathbf{A}+\mathbf{\Delta}$, $\exists \mathbf{v} \in \mathbb{C}^N$, $\mathbf{v} \neq \mathbf{0}$ such that

$$(\mathbf{A}+\mathbf{\Delta})\mathbf{v} = j\omega\mathbf{v}, \tag{2.7a}$$

$$(\mathbf{A}+\mathbf{e_k}\mathbf{\Gamma}^T)\mathbf{v} = j\omega\mathbf{v}, \tag{2.7b}$$

$$(\mathbf{A}+\mathbf{e_k}\mathbf{\Gamma}^T - j\omega\mathbf{I})\mathbf{v} = \mathbf{0}. \tag{2.7c}$$

Let $|\bullet|$ denote the determinant. From the characteristic equation,

$$|\mathbf{A} - j\omega\mathbf{I} + \mathbf{e_k}\mathbf{\Gamma}^T| = 0. \tag{2.8}$$

Since $\lambda_i(\mathbf{A}) \neq j\omega$, $|\mathbf{A} - j\omega\mathbf{I}| \neq 0$ and $(\mathbf{A} - j\omega\mathbf{I})^{-1}$ exists, so equation 2.8 can be factored as

$$|(\mathbf{A} - j\omega\mathbf{I})(\mathbf{I} + (\mathbf{A} - j\omega\mathbf{I})^{-1}\mathbf{e_k}\mathbf{\Gamma}^T)| = 0. \tag{2.9}$$

Recall that $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$:

$$|\mathbf{A} - j\omega\mathbf{I}||\mathbf{I} + (\mathbf{A} - j\omega\mathbf{I})^{-1}\mathbf{e_k}\mathbf{\Gamma}^T| = 0 \tag{2.10a}$$

$$|\mathbf{I} + (\mathbf{A} - j\omega\mathbf{I})^{-1}\underline{e}_k\mathbf{\Gamma}^T| = 0 \tag{2.10b}$$

For conformable matrices $\mathbf{C}$ and $\mathbf{D}$, $|\mathbf{I} - \mathbf{CD}| = |\mathbf{I} - \mathbf{DC}|$ so

$$|1 + \mathbf{\Gamma}^T(\mathbf{A} - j\omega\mathbf{I})^{-1}\mathbf{e_k}| = 0, \tag{2.11a}$$

$$\mathbf{\Gamma}^T(\mathbf{A} - j\omega\mathbf{I})^{-1}\mathbf{e_k} = -1. \tag{2.11b}$$

Note that equation 2.11b also holds if $j\omega$ is replaced by $-j\omega$ (by taking the complex conjugate of both sides) as expected. Since $\mathbf{A}+\mathbf{\Delta}$ is a real matrix, its complex eigenvalues must exist in conjugate pairs with the same $\mathbf{\Gamma}$ necessarily producing both.

Define

$$\mathbf{a}(\omega, k) = (\mathbf{A} - j\omega\mathbf{I})^{-1}\mathbf{e_k}, \tag{2.12a}$$

$$\mathbf{a_i}(\omega, k) = Im\{\mathbf{a}(\omega, k)\}, \tag{2.12b}$$

$$\mathbf{a_r}(\omega, k) = Re\{\mathbf{a}(\omega, k)\}. \tag{2.12c}$$

By definition of the induced 2-norm,

$$\|\mathbf{\Delta}\|_2 = \max\left\{\|\mathbf{\Delta x}\|_2 \mid \mathbf{x} \in \mathbb{R}^N, \|\mathbf{x}\|_2 = 1\right\}. \tag{2.13}$$

Since $\|\mathbf{\Delta}\|_2 = \|\mathbf{\Delta}^T\|_2$ and only the kth column of $\mathbf{\Delta}^T$ is nonzero, the maximum is achieved for $\mathbf{x} = \mathbf{e_k}$:

$$\left\|\mathbf{\Delta}^T\right\|_2 = \left\|\mathbf{\Delta}^T\mathbf{e_k}\right\|_2 \tag{2.14a}$$

$$= \left\|\mathbf{\Gamma}\mathbf{e_k}^T\mathbf{e_k}\right\|_2 \tag{2.14b}$$

$$= \|\mathbf{\Gamma}\|_2. \tag{2.14c}$$

Then, when equations 2.11b, 2.12b, and 2.12c are used, an equivalent minimization is

$$\min_{\mathbf{\Delta}\in\Lambda_r,\omega} \|\mathbf{\Delta}\|_2 \iff \min_{0<\omega\le\omega_o,k} \left\{\|\mathbf{\Gamma}\|_2 \mid \begin{bmatrix} \mathbf{a_i}^T(\omega, k) \\ \mathbf{a_r}^T(\omega, k) \end{bmatrix} \mathbf{\Gamma} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}\right\}. \tag{2.15}$$

This is just an underdetermined least squares problem over ω and k, where ω_o is the upper bound of the frequency range within which to search. Then the minimum magnitude $\mathbf{\Gamma}$ is achieved at $(\widehat{\omega}, \widehat{k})$ using the definitions from equation 2.5a, and $\widehat{\mathbf{\Delta}} = \mathbf{e_{\widehat{k}}}\widehat{\mathbf{\Gamma}}^T$:

$$\widehat{\mathbf{\Gamma}} = \mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1}\mathbf{b}. \tag{2.16}$$

If $\omega = 0$, then $\mathbf{a_i}(\omega, k) = \mathbf{0}$ so the first condition in equation 2.15 needs to be omitted and $\mathbf{\Gamma}$ simplifies to the expected form:

$$\mathbf{\Gamma} = \frac{-\mathbf{a_r}}{\mathbf{a_r}^T\mathbf{a_r}} = \frac{-\mathbf{a}}{\mathbf{a}^T\mathbf{a}}. \tag{2.17}$$

2.4.3 Corollary. For $\omega = 0$, $\mathbf{\Gamma}$ is itself the corresponding eigenvector of $\mathbf{A} + \mathbf{\Delta}$. This can be shown by substituting equation 2.17 a for $\mathbf{\Gamma}$ in equation 2.18a and 2.12a for $\mathbf{a}$ in equation 2.18c:

$$\left(\mathbf{A} + \mathbf{e_k}\mathbf{\Gamma}^T\right)\mathbf{\Gamma} = \mathbf{A}\mathbf{\Gamma} + \mathbf{e_k}\mathbf{\Gamma}^T\mathbf{\Gamma} \tag{2.18a}$$

$$= -\mathbf{A}\frac{\mathbf{a}}{\mathbf{a}^T\mathbf{a}} + \mathbf{e_k}\frac{\mathbf{a}^T\mathbf{a}}{\left(\mathbf{a}^T\mathbf{a}\right)^2} \tag{2.18b}$$

$$= \frac{-\mathbf{A}\mathbf{a} + \mathbf{e_k}}{\mathbf{a}^T\mathbf{a}} \tag{2.18c}$$

$$= \frac{-\mathbf{A}\mathbf{A}^{-1}\mathbf{e_k} + \mathbf{e_k}}{\mathbf{a}^T\mathbf{a}} \tag{2.18d}$$

$$= 0\mathbf{\Gamma}. \tag{2.18e}$$

More generally, these results can be extended to ensure certain elements remain unperturbed within a row or column by including additional orthogonality constraints in equation 2.15 to the corresponding elementary basis vectors.

2.5 Probabilistic Neural Network Model. The structured perturbation results shown above will now be applied to destabilize a neural network model used in Benayoun, Cowan, van Drongelen, and Wallace (2010), described briefly below. The network model consists of a set of nodes, connected by weighted edges, signifying synaptic strengths. The internal activity of a node is dependent on synaptic inputs and its current state. For simplicity, assume for now that each node is a single neuron. The interpretation of each node as a population of neurons is a straightforward extension and described in section A.5. This model is attractive because it includes parameters analogous to physiological features like refractory periods, synaptic strengths, and firing rates without involving the complexity of biophysical models.

Consider node i to be a single neuron that exists in one of two states at some time, active ($x_i(t) = 1$) or quiescent ($x_i(t) = 0$). The state probability evolves as a Markov process, with rate constants as given below, for a small time interval dt:

$$\Pr\{x_i(t + dt) = 0 \mid x_i(t) = 1\} = \alpha dt, \tag{2.19a}$$

$$\Pr\{x_i(t + dt) = 1 \mid x_i(t) = 0\} = f(s_i(t))dt. \tag{2.19b}$$

The active state represents the action potential duration of a neuron, including its refractory period. On average, a neuron is active for a period of α^{-1}. The input to neuron i is given by $s_i(t)$. The spiking propensity of a neuron therefore depends on the level of input it receives, but the inactivation

propensity is fixed. $f(\bullet)$ is a nonlinear response function, representing the expected firing rate when quiescent. The derivation places no assumption on the form of this function, but a clamped hyperbolic tangent was used in simulations:

$$f(s) = \max\{0, \tanh(s)\} \tag{2.20}$$

When equations 2.19a and 2.19b are combined, the nodal probability, $p_i(t) = \langle x_i(t) \rangle = Pr\{x_i(t) = 1\}$, that is, the probability of a neuron being active at some time, evolves according to the following nonlinear rate equation:

$$\dot{p}_i(t) = -\alpha p_i(t) + f(s_i(t))(1 - p_i(t)). \tag{2.21}$$

Consider a network of N such nodes, connected according to a structural connectivity matrix, $\mathbf{W} = [w_{ij}]$, whose elements are positive if node j excites node i, negative if node j inhibits node i, and zero if nodes i and j are physically unconnected:

$$\mathbf{W} = \begin{bmatrix} - & \mathbf{w_1}^T & - \\ & \vdots & \\ - & \mathbf{w_N}^T & - \end{bmatrix}, \mathbf{w_i} \in \mathbb{R}^N. \tag{2.22}$$

Row i stores the weights of inbound connections to node i, while column j stores the weights of outbound connections from node j.

The input to node i, $s_i(t)$, is contingent on the state of the nodes that connect to it, the weights of these connections, and some constant external input, h_i, as shown. If node j is active, node i feels an incremental synaptic drive of w_{ij}:

$$s_i(t) = \sum_{j=1}^{N} w_{ij} x_j(t) + h_i. \tag{2.23}$$

The state of all the nodes in the network at any time, the network state, is given by aggregating the nodal states, $x_i(t)$, into the Boolean vector $\mathbf{x}(t) \in \{0, 1\}^N$. Then equation 2.23 can be rewritten as

$$s_i(t) = \mathbf{w_i}^T \mathbf{x}(t) + h_i. \tag{2.24}$$

Similarly, the nodal inputs, $s_i(t)$, can be aggregated to form the synaptic input vector $\mathbf{s}(t) \in \mathbb{R}^N$. It can be expressed in matrix form as follows, where $\mathbf{h} \in \mathbb{R}^N$ is the aggregate vector of external inputs:

$$\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t) + \mathbf{h}. \tag{2.25}$$

The network probability, $\mathbf{p}(t) \in \mathbb{R}^N$, formed by aggregating nodal probabilities, $p_i(t)$, therefore evolves according to a nonlinear rate equation analogous to the nodal rate equation in equation 2.21:

$$\begin{aligned}\dot{\mathbf{p}}(t) &= -\alpha \mathbf{p}(t) + diag\left\{\mathbf{f}(\mathbf{W}\mathbf{p}(t) + \mathbf{h})\right\}\left(\mathbf{1} - \mathbf{p}(t)\right) \\ &\triangleq \mathbf{g}\left(\mathbf{p}(t); \mathbf{W}\right). \end{aligned} \tag{2.26}$$

The state vector is replaced by the network probability in the expansion of the synaptic drive term in the response function, since the mean behavior is being described.

2.5.1 Determining Functional Connectivity. Note that equation 2.26 explicitly forecasts the functional activity of the network, $\dot{\mathbf{p}}(t)$, given some network structure, $\mathbf{W}$, and so the notation $\mathbf{g}\left(\mathbf{p}(t); \mathbf{W}\right)$ is used. Assume that the initial structural connectivity of the network gives rise to functional stability, so that there exists a unique stable fixed point, $\widehat{\mathbf{p}} \in \mathbb{R}^N$. Details on the existence of a stable fixed point are given in section A.1. This fixed point is a steady-state probability that satisfies equation 2.27 and therefore represents the baseline behavior of the network. The fixed point also depends on the network structure:

$$\mathbf{g}(\widehat{\mathbf{p}}; \mathbf{W}) = \mathbf{0}. \tag{2.27}$$

Equation 2.26 can be linearized around this fixed point, where $\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})$ is the Jacobian evaluated at the fixed point (Strogatz, 1994):

$$\delta\dot{\mathbf{p}}(t) \approx \left(\frac{\partial \mathbf{g}(\mathbf{p}; \mathbf{W})}{\partial \mathbf{p}}\right)_{\mathbf{p}=\widehat{\mathbf{p}}} \delta\mathbf{p}(t) \tag{2.28a}$$

$$= \mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})\delta\mathbf{p}(t). \tag{2.28b}$$

The general form of the Jacobian is as follows. δ_{ij} is the Kronecker delta:

$$J_{ij}(\mathbf{p}; \mathbf{W}) = \frac{\partial g_i(\mathbf{p}; \mathbf{W})}{\partial p_j} \tag{2.29a}$$

$$= f'\left(\mathbf{w_i}^T\mathbf{p} + h_i\right) w_{ij}\left(1 - p_i\right) - \delta_{ij}\left(\alpha + f\left(\mathbf{w_i}^T\mathbf{p} + h_i\right)\right). \tag{2.29b}$$

Since $\widehat{\mathbf{p}}$ is a stable fixed point, $Re\left\{\lambda_i(\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W}))\right\} < 0$, so $\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})$ is the stable adjacency matrix for a continuous time (CT) system operating about $\widehat{\mathbf{p}}$. $\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})$ is therefore also the functional connectivity of a linearized network, with a given structural connectivity, $\mathbf{W}$. Row i of $\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})$ captures how the probability of any node being active affects the probability of node i being

active. Column j indicates how the active probability of node j affects the active probability of other nodes.

2.5.2 Determining the Baseline Behavior. The fixed-point $\widehat{\mathbf{p}}$ represents the baseline behavior of the network. Solving for $\widehat{\mathbf{p}}$ analytically using equations 2.26 and 2.27 is not generally possible. However, it can be computed using a gradient descent algorithm. (See section A.2 for more details.)

2.5.3 Row Perturbation on Neural Network Model. The goal is to determine the minimum energy functional perturbation required to produce instability and then decipher what structural changes would manifest with this functional perturbation. It is important to note that in equation 2.29b, the entries in row i of the functional connectivity matrix depend only on entries in row i of the structural connectivity matrix. That means that the effect of inbound activity is facilitated at the structural level wholly by the weight of inbound connections. Therefore, if a row perturbation is applied to $\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})$, this can be accounted for solely through modifications to the corresponding row of $\mathbf{W}$, interpreted as modifications to incoming synaptic strengths.

Since the network has been linearized about $\widehat{\mathbf{p}}$, the minimum row perturbation, $\mathbf{\Delta}_{\mathbf{J}}(\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W}))$, required to cause CT instability is given by the result in section 2.4. The unstable functional connectivity, $\tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W})$, therefore has the following form, where $r \in \{1, 2, \ldots, N\}$ is the perturbed row:

$$\tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W}) = \mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W}) + \mathbf{\Delta}_{\mathbf{J}}(\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})) \tag{2.30a}$$

$$= \mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W}) + \mathbf{e}_{\mathbf{r}}\mathbf{\Gamma}_{\mathbf{J}}{}^{T}(\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})). \tag{2.30b}$$

2.5.4 Disrupting Structure to Destabilize the Network. As described in section 2.5.3, the perturbed structural connectivity required to produce the unstable functional connectivity, is necessarily of the following form:

$$\tilde{\mathbf{W}} = \mathbf{W} + \mathbf{\Delta}_{\mathbf{W}} \tag{2.31a}$$

$$= \mathbf{W} + \mathbf{e}_{\mathbf{r}}\mathbf{\Gamma}_{\mathbf{W}}{}^{T}. \tag{2.31b}$$

This perturbed structural connectivity should give rise to a new functional connectivity, $\mathbf{J}(\widehat{\mathbf{p}}; \tilde{\mathbf{W}})$, identical to the unstable one computed from the original network configuration as shown in Figure 4:

$$\mathbf{J}(\widehat{\mathbf{p}}; \tilde{\mathbf{W}}) = \tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W}). \tag{2.32}$$

Note, however, that changes in the structural connectivity ($\mathbf{W} \to \tilde{\mathbf{W}}$) may change the fixed point, which alters all rows of the functional connectivity

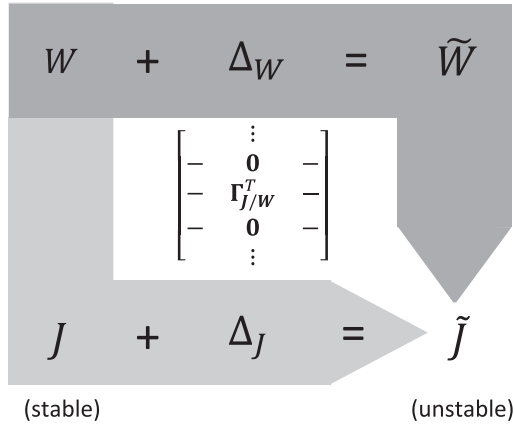

Figure 4: Procedure for producing network instability. In the first step (shown in light gray), the minimum energy row perturbation ($\mathbf{\Delta_J}$) theoretically required to cause functional instability is applied to the functional connectivity ($\mathbf{J}$) arising from the original network structure ($\mathbf{W}$), to compute the theoretical unstable functional connectivity ($\tilde{\mathbf{J}}$). In the second step (shown in dark gray), synaptic weights are perturbed ($\mathbf{\Delta_W}$) in the corresponding row of $\mathbf{W}$ to produce a new structural connectivity matrix ($\tilde{\mathbf{W}}$), which manifests a functional connectivity matrix equal to $\tilde{\mathbf{J}}$. Both functional and structural perturbations have the form shown in the middle matrix with nonzero entries along the same single row.

so that $\mathbf{J}(\widehat{\mathbf{p}}; \tilde{\mathbf{W}})$ cannot reproduce $\tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W})$, which differs only by a single row. This is to be expected, since $\mathbf{\Delta_J}$ is computed for a functional connectivity at a given fixed point. Therefore, $\tilde{\mathbf{W}}$ must be restricted to be invariant on the fixed point of the unperturbed system so that

$$\widehat{\mathbf{p}}(\tilde{\mathbf{W}}) = \widehat{\mathbf{p}}(\mathbf{W}) = \widehat{\mathbf{p}}. \tag{2.33}$$

This can be accomplished only if the probability flux in equation 2.26 remains at zero while the structural connectivity is changed so that the fixed point does not move:

$$\mathbf{g}(\tilde{\mathbf{W}}; \widehat{\mathbf{p}}) = \mathbf{g}(\mathbf{W}; \widehat{\mathbf{p}}) = \mathbf{0}. \tag{2.34}$$

Note that $\mathbf{g}(\mathbf{W}; \widehat{\mathbf{p}})$ is used instead of $\mathbf{g}(\widehat{\mathbf{p}}; \mathbf{W})$ as before, since the fixed point is now the constant and structural connectivity the variable. Inspection of equation 2.26 reveals conditions required to satisfy equation 2.34 element-wise:

- For $i \neq r$:

$$\mathbf{e_i}^T \tilde{\mathbf{W}} = \mathbf{e_i}^T \mathbf{W} \Rightarrow g_i(\mathbf{W}; \widehat{\mathbf{p}}) = g_i(\tilde{\mathbf{W}}; \widehat{\mathbf{p}}) = 0 \text{ always.} \tag{2.35}$$

- For $i = r$:

$$g_r(\tilde{\mathbf{W}}; \widehat{\mathbf{p}}) = 0 \iff \mathbf{w_r}^T \widehat{\mathbf{p}} = f^{-1}\left(\frac{\alpha \widehat{p}_r}{1 - \widehat{p}_r}\right) - h_r = \tilde{\mathbf{w}}_r^T \widehat{\mathbf{p}} \tag{2.36a}$$

$$\Rightarrow \left(\tilde{\mathbf{w}}_r - \mathbf{w_r}\right)^T \widehat{\mathbf{p}} = 0. \tag{2.36b}$$

The modified row of the structural connectivity, $\tilde{\mathbf{w}}_r^T$, can be computed using a constrained gradient descent in a manner analogous to the fixed-point discovery. (See section A.3 for details.)

2.5.5 Model Simulation. The network model is a probabilistic one, and so far its average behavior has been described. To simulate realizations of the stochastic process, the Gillespie algorithm (Gillespie, 1977) can be used as in Benayoun et al. (2010). It is an exact algorithm that generates stochastic realizations faithful to the underlying distribution. (The steps are described in section A.4.)

3 Results

In this section, simulation results are presented for a sample network constructed with certain symmetry properties to guarantee a stable baseline behavior (see section A.1). The network was designed with only a few nodes to allow for easy visualization. The structure and function of the network in its stable state are described first, followed by its properties after destabilization.

3.1 Stable Network. A schematic of the original, stable network is shown in Figure 5A. The network has 6 nodes, each a single neuron, and 14 connections. The decay rate, α, is set to 100 Hz, which caps the neuronal firing rates at that value. This network was simulated using Matlab with the same initial conditions for 100 trials, each 10 seconds in duration.

The functional connectivity, computed by evaluating equation 2.29b, at the fixed point, is shown in Figure 6A. The negative elements along the diagonal arise partly because of the decay constant intrinsic to the nodal dynamics. Generally the more attractive the fixed point, that is, the more negative the eigenvalues of the functional connectivity, the faster the theoretical and empirical results converge with respect to the number of trials. The eigenspectrum of the unperturbed network is presented in Figure 7A.

Raster plots of 1 second segments from 10 stochastic realizations are shown in Figure 8A. The spike times are marked as the instant when a node

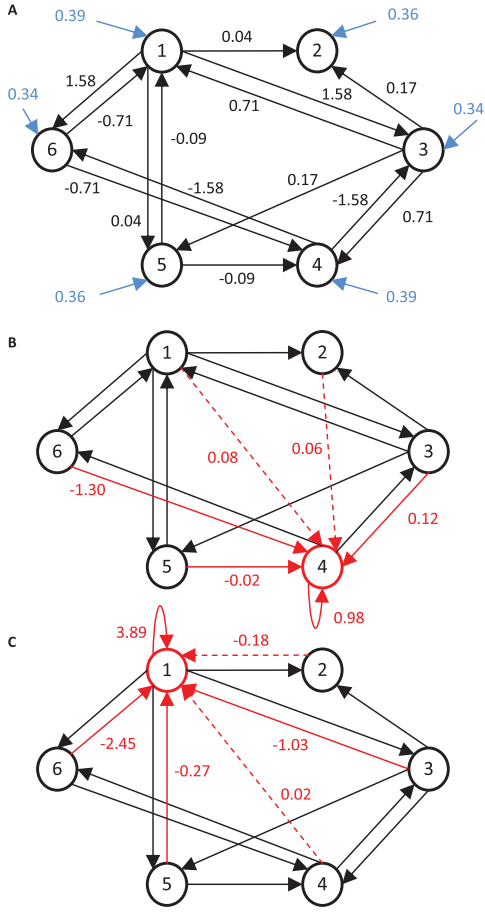

Figure 5: Graph representation of a neuronal network with six nodes. (A) The original structural connectivity of the network. The gray arrows indicate the external inputs, h_i. Black arrows represent synaptic weights, w_{ij}, between nodes. (B) The perturbed structure of the network after a row perturbation is applied at 0 Hz. The most fragile node and subnetwork is highlighted in gray. Solid gray arrows indicate existing connections with modified strengths. Dotted gray arrows indicate new connections. The black arrows have weights identical to the original network structure. The external input is also the same and omitted for clarity. (C) The perturbed structure of the network after a row perturbation is applied at 40 Hz.

switches from quiescent to active. Firing rates are therefore closely related to the active state probabilities at the fixed point. All nodes fire fairly regularly in the raster plot.

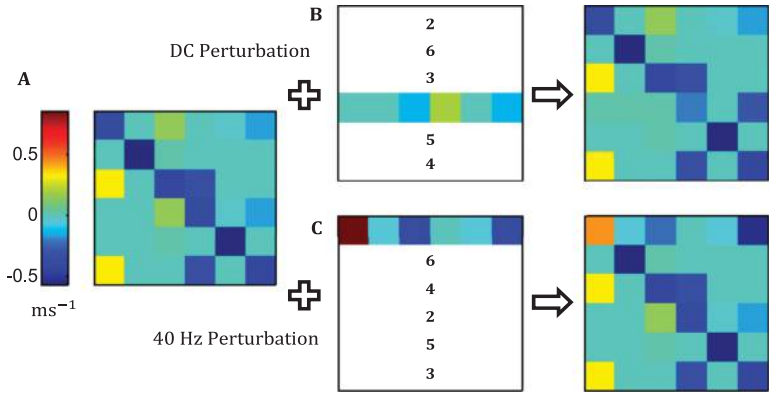

Figure 6: Functional connectivity of neuronal network. The functional connectivity of (A) the stable network, $\mathbf{J}(\widehat{\mathbf{p}}; \mathbf{W})$. (center) the applied perturbation, $\mathbf{\Delta}_{\mathbf{J}}$, and (right) the unstable functional connectivity, $\tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W})$, for (B) a DC perturbation and (C) a perturbation at 40 Hz. The DC perturbation is the minimum energy perturbation across all frequencies. The numbers in the perturbation matrix show the fragility rank of each node for perturbations at the given frequency, with larger ranks being less fragile. Nodes 2 and 5, an excitatory-inhibitory pair, have the weakest incoming synaptic weights and are the most robust to both perturbations.

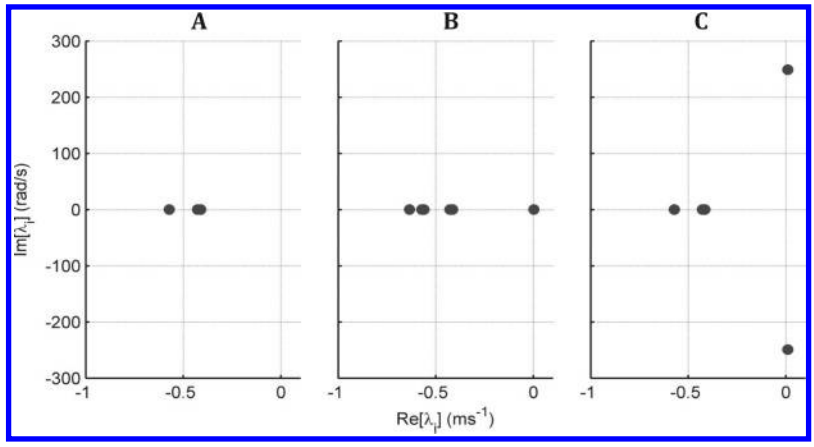

Figure 7: Eigenspectrum of functional connectivity. Eigenspectrum of the functional connectivity for (A) the original stable network, (B) the unstable network perturbed at 0 Hz, and (C) the unstable network perturbed at 40 Hz. In the unstable cases, the eigenspectrum shown is computed on the functional connectivity arising from the perturbed structure, $\mathbf{J}(\tilde{\mathbf{W}}; \widehat{\mathbf{p}})$, not simply the theoretical value, $\tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W})$.

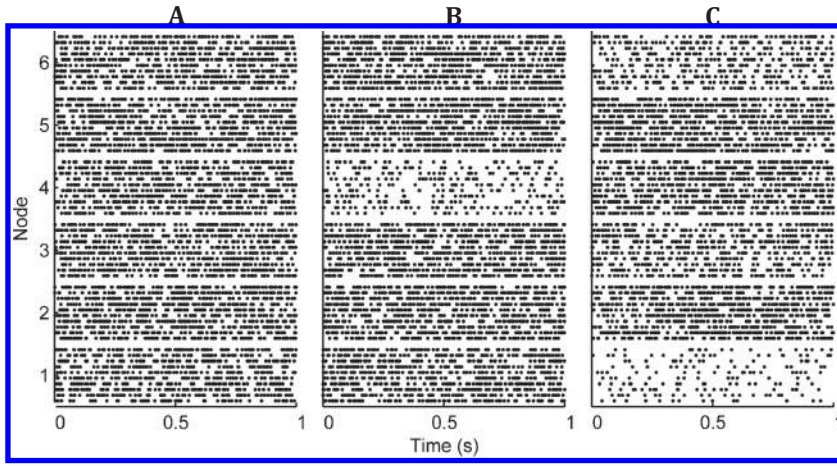

Figure 8: Raster plots for stable and marginally stable (perturbed) networks. Ten stochastic realizations of neural activity from (A) the original stable network, (B) the unstable network perturbed at 0 Hz, and (C) the unstable network perturbed at 40 Hz. There is a single neuron in each node.

3.2 Perturbed Network: DC Instability. The behavior of the perturbed network is now highlighted. Using result 2, the minimum energy row perturbation for the system is computed and shown in Figure 6B. In this case, the minimum perturbation occurs at a DC frequency ($\lambda = 0$), and the eigenspectrum of the functional connectivity has an eigenvalue shifted to the origin, as illustrated in Figure 7B. The structural perturbation that would manifest as this specific functional instability is computed as described in section 2. The structure of the perturbed network is shown in Figure 5B. The most fragile subnetwork corresponds to the incoming edges to node 4, an inhibitory neuron. The empirical changes in the synaptic input-output properties across the network due to the DC perturbation applied at node 4 are shown in Figure 9A. The changes in synaptic properties reflect the combined effect of a deviation from the baseline active probabilities and altered synaptic weights.

Figure 8B shows a segment of 10 stochastic realizations of network activity after the structural perturbation is applied. The network is operating away from the baseline behavior at its fixed point with node 4 spiking less frequently.

The nodes with the largest deviation in active probabilities from their stable fixed-point values are those with the largest magnitude elements in the eigenvector of the perturbed functional connectivity corresponding to $\lambda = 0$. As proved in section 2.4.3, for DC perturbations, this eigenvector is collinear to the row perturbation itself. This is confirmed by comparing

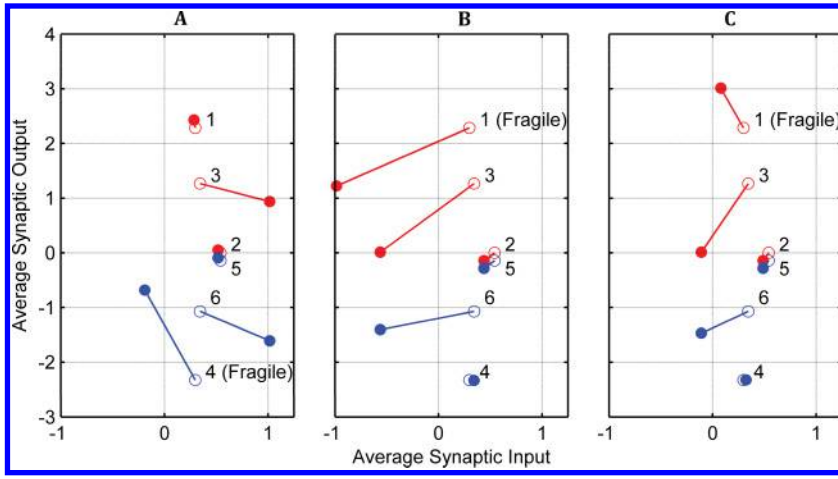

Figure 9: Changes in synaptic input and output to each node due to perturbation. The open and solid circles mark the nodal synaptic properties before and after perturbation, with excitatory/inhibitory nodes marked in light/dark gray. The synaptic input is calculated as $\mathbf{Wp} + \mathbf{h}$. The synaptic output is computed by multiplying a node's active state probability by the sum of its outgoing weights. This metric captures the contribution of a node to the synaptic input across all its targets. The active state probability $\mathbf{p}$ is computed empirically using the simulation protocol described for each of the different network configurations. The changes in synaptic properties are shown for (A) the DC perturbation, (B) the 40 Hz perturbation with a single neuron per node, and (C) the 40 Hz perturbation with 10 neurons per node. The DC perturbation causes inhibitory node 4 to receive less synaptic input and produce less inhibitory output.

Figure 6B to the raster plot in Figure 8B: a node's deviation from the theoretical stable fixed point value is proportional to its element in the row perturbation vector.

When the linearized system is perturbed so that an eigenvalue is moved to the origin, a stable manifold is on the cusp of becoming unstable. In a deterministic system, the emergence of an unstable manifold implies that for almost all initial conditions, the trajectory will grow without bound along the unstable manifold at a rate proportional to the eigenvalue.

In this system, since probabilities are being modeled, there are necessarily upper and lower bounds. The firing rates too are therefore capped. Furthermore, since the system is stochastic, trajectories do not indefinitely follow the unstable manifold but are kicked away and then recycled to the saddle point by the stable manifolds. This mechanism may explain the heterogeneity in firing rates seen during seizure activity as trajectories wander around a fixed point.

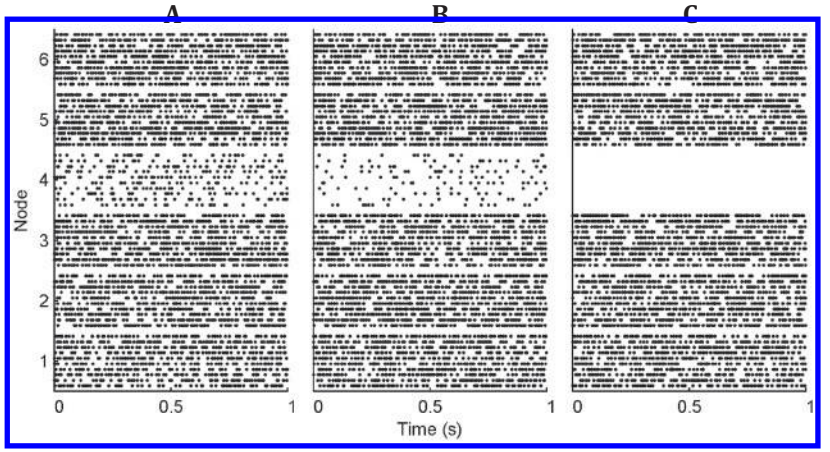

Figure 10: Raster plots for network destabilized at DC. Ten stochastic realizations of neural activity from a network progressively destabilized at DC with (A) $\lambda = 0\ \mathrm{ms}^{-1}$, (B) $\lambda = 0.2\ \mathrm{ms}^{-1}$, and (C) $\lambda = 1\ \mathrm{ms}^{-1}$. There is a single neuron in each node. Node 4 is eventually silenced.

As the eigenvalue is set to be more positive, the trajectories follow the unstable manifold more strongly, the noisy resetting mechanism occurs less frequently, and the active probabilities saturate at the boundaries more consistently during simulations, so that the nodes either approach tonic spiking or become silent. A raster plot of the progressive change in spiking activity is shown in Figure 10. Inhibitory node 4 (where the perturbation is applied) is eventually silenced, and there is increased activity in nodes 3 and 6, which originally felt the strongest inhibition from node 4 prior to the perturbation. Cortical neurons from epilepsy patients show a similar increase or decrease in firing rates during seizure. Compare Figures 8A and 10 to the preictal and ictal spiking patterns, respectively, in Figure 11.

3.3 Perturbed Network: Gamma Frequency Instability. The original structure is now perturbed specifically using the minimum energy perturbation at a gamma frequency of 40 Hz ($\lambda = j\omega = j80\pi$). This frequency was chosen because structural perturbations that could achieve a desired functional perturbation existed only for a narrow band around 40 Hz in this network (apart from DC).

It is interesting that the network is susceptible to perturbations only at specific frequencies: the EEG signals of different epilepsy patients exhibit increased power in different frequency bands during seizure, and this may reflect the vulnerability of each patient's brain networks to oscillations in a specific band.

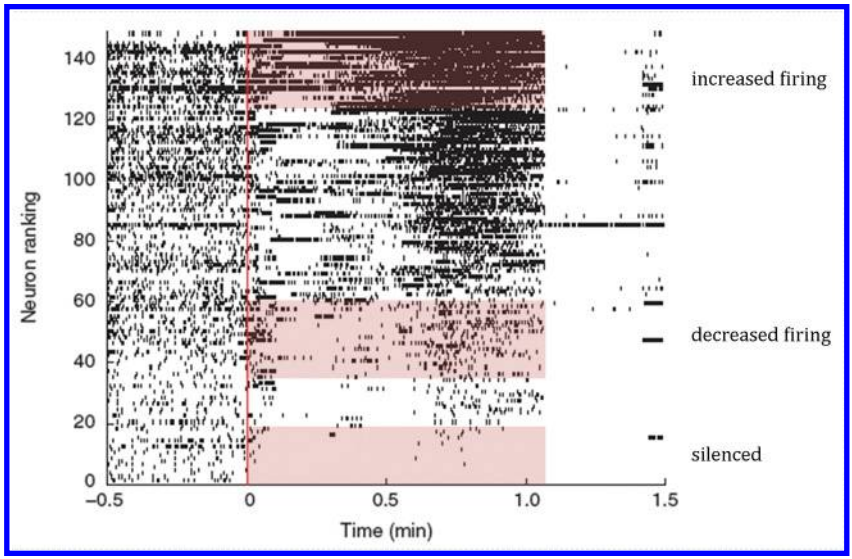

Figure 11: Raster plot of cortical neurons sorted by mean spiking rate during seizure. The spiking rate of any given neuron is fairly uniform before seizure onset (marked by the vertical gray line) suggesting a stable baseline behavior. During seizure, some neurons are silenced, a few fire almost tonically throughout, and others fire in bursts that appear as short streaks in the raster plot. (Reprinted by permission from Macmillan Publishers Ltd: Nature Neuroscience (Truccolo et al., 2011), copyright 2011.)

Generally it also appears as if smaller networks can be destabilized only at higher frequencies for a given set of parameter values. However, biological systems necessarily have thermodynamic constraints that place an upper bound on how fast network activity can oscillate, a fact worth considering when setting ω_o, the maximum frequency within which to search, in equation 2.15.

The perturbed functional connectivity at 40 Hz is presented in Figure 6C and its corresponding eigenspectrum in Figure 7C. The complex conjugate eigenvalues on the imaginary axis correspond to the perturbation frequency. The most fragile node at this frequency is node 1, the excitatory counterpart to the node that is most fragile in the DC case. The structure of the perturbed network required to produce this instability is shown in Figure 5C. The altered synaptic properties are shown in Figure 9B. Nodes 3 and 6, which normally feel the strongest excitatory effects from node 1, are quieter due to the reduced activity in node 1.

A perturbation that causes eigenvalues to move to the imaginary axis in the linearized system corresponds to a Hopf bifurcation in the nonlinear system, with the emergence of a stable limit cycle. When the time course

of the probability rate equation in equation 2.26 is simulated after perturbing the structure, the components of the probability vector fluctuate, indicating the presence of a closed orbit. The oscillations settle to the same envelope for different initial conditions, confirming that this orbit is isolated.

The imaginary component of the eigenvalue is the frequency of limit cycle oscillations at the birth of the limit cycle, when it is of infinitesimal radius (Strogatz, 1994). Since trajectories spiral inward toward this small-radius limit cycle, the oscillation frequencies are expected to be different from the nominal frequency. Furthermore, if a stronger perturbation is applied so that the real part of the eigenvalue becomes positive, the radius of the limit cycle increases and the nominal frequency varies proportionally.

It is important to note that the active probability itself is oscillating over time at the perturbation frequency. The actual incidence of spiking, which may occur at other frequencies, is modulated by this probability wave. In this example, the spiking does not appear as obvious bursts in Figure 8C. Stochastic systems follow the deterministic mean equations more closely for larger system sizes (neurons per node). In the neuronal simulation paradigm used in this example, the effect of intrinsic noise (from stochastic realizations) overrides the predicted dynamics. This is discussed in more detail in the following section.

3.4 Perturbed Network: Population Oscillations. The network is simulated using the same structure and 40 Hz perturbation results as in section 3.3, but now with 10 neurons per node, according to the extension described in section A.5. Equation 2.26 now represents the evolution of the active proportion of neurons in each node. Figure 12 shows a raster plot of a single trial with the activity of the neurons grouped by node. There are clearly population oscillations occurring, with the neurons in a population firing in bursts. These occur at somewhat irregular intervals and this is partly because the frequency of the perturbation is fairly high with respect to the maximum neuronal firing rate, so that the neurons are unable to stay entrained with the probability wave calculated analytically. However, the nodal activity is clearly oscillating, which was not the case when a single neuron occupied a node. In fact, the oscillations in node 1, the perturbed node, appear to be driving oscillations in nodes 3 and 6, the downstream targets on which it synapses most strongly. The synaptic properties, when considered by node, are also different compared to the single-neuron case (see Figure 9C).

As the number of neurons in a node is increased, the lattice formed by the Markovian network becomes higher in resolution, and the system is able to track the deterministic equation more closely. Since the evolution of the active proportion in equation 2.26 has no history dependence, every instance of a state transition during stochastic realizations effectively resets the equation to a new initial condition. That is why in the single-neuron case, there is no spiking coincidence across trials. Each transition resets the state

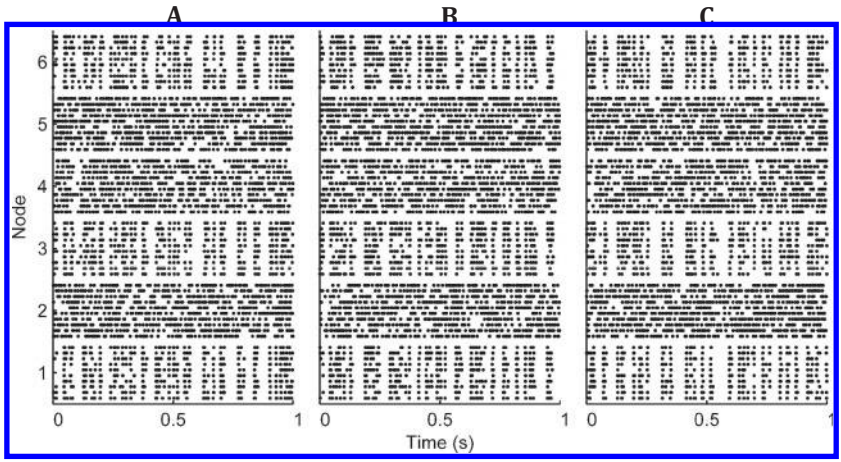

Figure 12: Raster plots for network destabilized at 40 Hz. A single stochastic realization of neural activity from a network progressively destabilized at 40 Hz with (A) $Re\{\lambda\} = 0$ ms^{-1}, (B) $Re\{\lambda\} = 0.05$ ms^{-1}, and (C) $Re\{\lambda\} = 0.1$ ms^{-1}. There are 10 neurons in each node.

to a new corner of the six-dimensional hypercube formed by the Markovian lattice, which effectively introduces stochastic phase shifts within a trial that are not consistent across trials.

As the population size increases, the lattice becomes finer, and noise (from any specific neuronal transition realization) has a diminishing effect on the active proportion, so that the network is able to track the analytical probability oscillations with greater fidelity. In the thermodynamic limit, as the number of neurons in a node approaches infinity, the empirical active proportion effectively becomes a continuous variable and is able to track the analytical proportion exactly.

In the DC perturbation case, the dynamics are phase independent, so even single-neuron nodes exhibit the expected dynamics. Even there, though, increasing the number of neurons per node will result in a faster convergence to the analytical steady-state proportions. (Refer to Goutsias & Jenkinson, 2013, for a complete treatment of how network size affects stochastic solutions of a probability master equation for a complex network with Markovian dynamics.)

A more rigorous analysis of the oscillatory behavior of a multinode network is beyond the scope of this letter, but Wallace, Benayoun, van Drongelen, and Cowan (2011) offer a comprehensive study of the behavior of a two-node network governed by the same dynamics, consisting of an excitatory and inhibitory population. The authors characterize the network's fidelity to the deterministic equations along with other emergent phenomena,

such as noisy limit cycles and quasi-cycles. Their conclusions can likely be extrapolated to the more complex multinode networks studied here.

4 Discussion

This study explored the relation between structure and function in a neuronal network model. The concept of network fragility was introduced, and the minimum energy perturbation required to destabilize a linear network was derived for the row perturbation topology. A procedure for linking structural connectivity to functional connectivity was developed for a probabilistic neuronal network so that a given functional instability could be reproduced by appropriate modifications to synaptic weights. These results were used to perturb a neuronal network with a stable baseline behavior to cause atypical functional activity, including the silencing of nodes, and population oscillations. Simulation results suggest that the most fragile nodes in the neural network are either excitatory neurons that become more active or inhibitory neurons that become less active. This is consistent with abnormal axonal sprouting of excitatory neurons and loss of inhibitory chandelier cells observed in epileptic cortical tissue. Furthermore, simulation of the unperturbed and perturbed network reflects neuronal activity observed in epilepsy patients implanted with electrodes before and during seizure, respectively. The mechanisms by which aberrant activity was produced, namely, wandering trajectories about a fixed point with an unstable manifold and a noisy limit cycle, may be indicative of neuronal or population dynamics in seizure.

4.1 Interpreting Recordings in Epileptic Cortex. In microelectrodes recordings, different electrodes record from single units or clusters of neurons, corresponding to nodes in the network model. In the microelectrode recordings from Truccolo et al. (2011), some neurons fire at increased rates during seizure, others at decreased rates, and some are silenced altogether. These behaviors are consistent with the dynamical mechanism of an unstable manifold in the probability space suggested here. Our findings indicate that the epileptic cortex is on the brink of instability and that minute changes in synaptic weights can suddenly destabilize the network to cause seizures.

Furthermore, in EEG data, during seizure, there is increased synchronous, oscillatory behavior. Our findings here suggest that when the sample size in a node is small, oscillatory behavior does not arise empirically, but in a survey of a large sample of neurons from a single recording site, as with an EEG electrode, the oscillatory signal collected corresponds to the mean behavior of many neurons, whose intrinsic activity may not be inherently oscillatory.

4.2 Toward a Cortical Network Model. In order to explicitly link functional and structural connectivity, a simplistic network model was used.

However, in order to capture more detailed dynamics, complex biophysical models may be necessary, and a procedure for fitting the average firing behavior of biophysical models into a probabilistic framework may be a worthwhile endeavor. In this model, uniform decay rates were used for all nodes, but these could be adapted for each neuronal type. The response function could also be selectively scaled for each neuron to reflect its spiking propensity. The synaptic strengths connecting the neurons could be informed by actual synaptic conductances and receptor densities.

4.2.1 Structural Connectivity Perturbations. For a given perturbation, it is not always realistic to expect that a plausible change in structural connectivity can produce the given functional instability. In the example explored here, with very few nodes and hence few free parameters, in order to produce a certain functional activity, the required changes to structural connectivity caused some nodes to excite particular neighbors while inhibiting others. At the level of populations this may not be problematic. However, future work may entail computing the minimum energy functional perturbation that can be realized using realistic modifications to synaptic weights.

Multiple changes to structural connectivity may also give rise to a particular functional activity profile, so a more refined procedure that involves first running a gradient descent, followed by individually adjusting each synaptic weight to avoid these inconsistencies, may yield more realistic results.

4.2.2 Incorporating Synaptic Plasticity. In this work, the unperturbed structural connectivity is treated as a fixed entity. However, synaptic plasticity is another mechanism by which structural and functional connectivity change in cortical networks. This could perhaps be incorporated into the model by having the structural connectivity be time varying on a scale much slower than the nodal dynamics. A naturally arising constraint is that the corresponding functional connectivity due to synaptic plasticity-induced changes alone should continue to be stable. The effect of incorporating synaptic plasticity would be to change the form of the minimum energy perturbation over time, but the dynamical mechanisms by which instability arise would remain the same.

4.2.3 Effects of External Inputs. This letter was motivated by the idea that structural changes can bring about the functional instability observed during seizures. However, if structural pathologies are a property of the epileptic brain, it is reasonable to ask why cortical networks are unstable only in the ictal state. It may be the case that compensatory mechanisms exist, such as modulations in the external input to fragile subnetworks from other brain areas, so that changes in network dynamics produced by the structural pathology are offset to maintain the physiological functional

profile. Network instability may manifest only when there are transient failures in the compensatory modulations of external inputs. This is supported by Fröhlich, Sejnowski, and Bazhenov (2010) who found that sparse physiological activity and tonic-clonic pathological activity can coexist in a computational network model with identical model parameters and afferent inputs, with transient perturbations in the input being sufficient to switch the network between these two states.

In this letter, external inputs were taken to be constant, which necessitated the fairly strict condition of fixed-point invariance with respect to the structural connectivity. Allowing for modulated external inputs relaxes this constraint, so that the fixed point is invariant only on the synaptic input, while simultaneously providing new conditions on the values of the external input that would trigger network instability. Permitting external inputs to be variable also allows more tractable analysis of perturbation structures beyond just row perturbations.

Destabilizing perturbations were treated as a static entity here, whereas the progression of seizures is a dynamical process through marked seizure states (Burns et al., 2012). This dynamical evolution could perhaps be reproduced by tuning the external input to achieve a progression of functional connectivities (for a fixed perturbed structural connectivity), with stability properties matching the time course of EEG signals or neural data collected during seizure activity.

Finally, this framework may also be applicable to the physiological state of the healthy brain. There are many processes, such as learning, where there is evidence of entrainment to oscillations, which could potentially be modeled this way.

4.3 Toward Closed-Loop Control of Seizures. This framework naturally lends itself to a two-state hidden Markov model (HMM) of a neuronal network (Santaniello, Sherman, Mirski, Thakor, & Sarma, 2011; Santaniello, Sherman, Thakor, Eskandar, & Sarma, 2012). The network is in either a stable hidden state or an unstable hidden state. The latter is described as the perturbed network. One can then assume that the state changes randomly from stable to unstable and vice versa to create a two-state Markov chain. The output observations of the HMM are the spikes from each node. With this HMM, one can construct a feedback controller that detects the underlying state from sequential output observations and, if the network is detected as unstable, applies a state feedback gain to move from unstable to stable. Detection of the underlying state can be done in real time using Bayesian estimation or the quickest detection algorithm, which has been applied to various neural data sets (Santaniello, Burns, Golby, Singer, Anderson, & Sarma, 2011; Santaniello et al., 2012). The feedback gain can be designed by applying state-feedback principles to the linearized system, equation 2.28a to move the eigenvalues back to a stable regime. This is left for future work.

Appendix: Additional Details on the Method

A.1 Existence of a Stable Fixed Point. The derivation assumes that a stable fixed point exists with meaningful elements (in the closed interval $\in [0, 1]$), but this is not guaranteed for all structural configurations of the network. Biologically, it is reasonable to assume a stable fixed point exists because neurons exhibit some baseline firing rate. It is worth noting that networks of N excitatory and N inhibitory nodes with the symmetry properties described in Benayoun et al. (2010) will have a stable fixed point. This provides a convenient way of generating stable networks.

In this class of network structures, each inhibitory node shares common network and external inputs with an excitatory node. The projections of the nodes in each pair, however, can be different.

A.2 Fixed-Point Discovery. The fixed point can be determined using a gradient descent algorithm that iterates on candidate solutions, $\widehat{\mathbf{p}}^{\mathbf{n}}$, with step size γ, to minimize a cost function, $C(\widehat{\mathbf{p}}^{\mathbf{n}})$:

$$\widehat{\mathbf{p}}^{\mathbf{n+1}} = \widehat{\mathbf{p}}^{\mathbf{n}} - \gamma \nabla C(\widehat{\mathbf{p}}^{\mathbf{n}}). \tag{A.1}$$

Since equation 2.27 is a property of the fixed point, a natural cost function is

$$C(\widehat{\mathbf{p}}^{\mathbf{n}}) = \frac{1}{2}\mathbf{g}^T(\widehat{\mathbf{p}}^{\mathbf{n}}; \mathbf{W})\mathbf{g}(\widehat{\mathbf{p}}^{\mathbf{n}}; \mathbf{W}). \tag{A.2}$$

Then the gradient is given by

$$\nabla C(\widehat{\mathbf{p}}^{\mathbf{n}}) = \mathbf{J}^T(\widehat{\mathbf{p}}^{\mathbf{n}}; \mathbf{W})\mathbf{g}(\widehat{\mathbf{p}}^{\mathbf{n}}; \mathbf{W}). \tag{A.3}$$

The initial seed probabilities are drawn from a uniform distribution:

$$\widehat{p}_i^{\mathbf{1}} \sim U(0, 1). \tag{A.4}$$

The halting condition is triggered when the cost function falls below some threshold:

$$C(\widehat{\mathbf{p}}^{\mathbf{n}}) \to 0 \Rightarrow \widehat{\mathbf{p}}^{\mathbf{n}} \to \widehat{\mathbf{p}}. \tag{A.5}$$

A.3 Gradient Descent for Perturbing Structural Connectivity. The initial conditions, iteration scheme and halting criterion for a typical gradient descent are given by equations A.6a to A.6c, respectively:

$$\widetilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{1}} = \mathbf{w}_{\mathbf{r}}, \tag{A.6a}$$

$$\tilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{n+1}} = \tilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{n}} - \gamma \mathbf{K}^T(\tilde{\mathbf{W}}^{\mathbf{n}}; \widehat{\mathbf{p}})[\mathbf{J}(\tilde{\mathbf{W}}^{\mathbf{n}}; \widehat{\mathbf{p}}) - \tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W})]^T \mathbf{e}_{\mathbf{r}}, \tag{A.6b}$$

$$\|\mathbf{J}(\tilde{\mathbf{W}}^{\mathbf{n}}; \widehat{\mathbf{p}}) - \tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W})\|_2 \to 0 \Rightarrow \tilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{n}} \to \tilde{\mathbf{w}}_{\mathbf{r}} \Rightarrow \tilde{\mathbf{W}}^{\mathbf{n}} \to \tilde{\mathbf{W}}. \tag{A.6c}$$

The initial condition is the original structural connectivity. The formulation of the cost function is analogous to the fixed-point discovery. The halting criterion is simply that the functional connectivity produced by the perturbed structural configuration should match the unstable functional connectivity computed from the original structural configuration. $\mathbf{K}(\tilde{\mathbf{W}}^{\mathbf{n}}; \widehat{\mathbf{p}})$ is the Jacobian of row r of the functional connectivity at the fixed point with respect to row r of the structural connectivity evaluated at some iteration. The general form is

$$K_{ij}(\mathbf{W}; \widehat{\mathbf{p}}) = \frac{\partial J_{ri}(\mathbf{w}_{\mathbf{r}}; \widehat{\mathbf{p}})}{\partial w_{rj}} \tag{A.7a}$$

$$= [f''(\mathbf{w}_{\mathbf{r}}{}^T \widehat{\mathbf{p}} + h_r)\widehat{p}_j w_{ri} + \delta_{ij} f'(\mathbf{w}_{\mathbf{r}}{}^T \widehat{\mathbf{p}} + h_r)](1 - \widehat{p}_r) \tag{A.7b}$$

$$-\delta_{ri} f'(\mathbf{w}_{\mathbf{r}}{}^T \widehat{\mathbf{p}} + h_r)\widehat{p}_j. \tag{A.7c}$$

In order to satisfy the fixed-point invariance condition in equation 2.36b a, a constrained gradient descent must be performed with

$$\tilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{n+1}} - \tilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{n}} \perp \widehat{\mathbf{p}}. \tag{A.8}$$

This is accomplished by constructing $\mathbf{B} \in \mathbb{R}^{N \times (N-1)}$ where

$$R(\mathbf{B}) = R^{\perp}(\widehat{\mathbf{p}}). \tag{A.9}$$

The gradient is then projected into a space orthogonal to $\widehat{\mathbf{p}}$ by premultiplying by $\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$. The modified iteration scheme is.

$$\tilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{n+1}} = \tilde{\mathbf{w}}_{\mathbf{r}}^{\mathbf{n}} - \gamma \mathbf{B}[\mathbf{B}^T\mathbf{B}]^{-1}\mathbf{B}^T\mathbf{K}^T(\tilde{\mathbf{W}}^{\mathbf{n}}; \widehat{\mathbf{p}})[\mathbf{J}(\tilde{\mathbf{W}}^{\mathbf{n}}; \widehat{\mathbf{p}}) - \tilde{\mathbf{J}}(\widehat{\mathbf{p}}; \mathbf{W})]^T \mathbf{e}_{\mathbf{r}}. \tag{A.10}$$

Then $\mathbf{\Gamma}_{\mathbf{W}}$ in Figure 2.31b, the modifications to the inbound structural connections to node r, is given by

$$\mathbf{\Gamma}_{\mathbf{W}} = \tilde{\mathbf{w}}_{\mathbf{r}} - \mathbf{w}_{\mathbf{r}}. \tag{A.11}$$

Since a gradient descent procedure is being used, conceptually at least, the smallest structural perturbation is being found to accomplish the minimum energy functional perturbation.

Suppose that only existing connections should be modified without adding new ones. Let $\{c_i\}$ be the set of m nodes such that $w_{rc_i} \neq 0$. Instead of equation A.9, choose $\mathbf{B}$ such that

$$R(\mathbf{B}) = R^{\perp}(\widehat{\mathbf{p}}) \cap span^{\perp}\{\mathbf{e}_{\mathbf{c_1}}, \ldots, \mathbf{e}_{\mathbf{c_m}}\}. \tag{A.12}$$

Note that in general, there is no guarantee that a change in structural connectivity can give rise to a particular perturbed functional connectivity (the halting criterion in the gradient descent cannot be met). This poses natural constraints on the way a network can become unstable.

A.4 Neuronal Simulation Paradigm. The steps in the Gillespie stochastic simulation algorithm are described below for a network of N nodes treated as individual neurons:

Step 1: Pick some initial condition for the nodal states, $x_i(0) \in \{0, 1\}$.
Step 2: Evaluate synaptic inputs using equation 2.23.
Step 3: Calculate transition rates, r_i, at each node.

$$r_i = \alpha x_i(t) + f\left(s_i(t)\right)\left(1 - x_i(t)\right). \tag{A.13}$$

Since $x_i(t)$ is a Boolean variable, r_i is the transition rate out of the node's current state (α if active or $f\left(s_i(t)\right)$ if inactive).
Step 4: Determine the network transition rate:

$$r_{net} = \sum_{i=1}^{N} r_i \tag{A.14}$$

Step 5: Switch the state of a single node i with a probability of $\frac{r_i}{r_{net}}$ and update the state vector. A node is more likely to be the one that switches states at a given time if it has the highest transition rate in the network.
Step 6: Draw the step time dt from an exponential distribution. Note that if the network transition rate is high, the frequency of state changes is high and the step time is therefore smaller:

$$u \sim U(0, 1), \tag{A.15a}$$

$$dt = \frac{-log(u)}{r_{net}}. \tag{A.15b}$$

Step 7: Increment the time counter by dt and iterate from Step 2 for the desired duration.

A.5 Extension to Populations. Suppose that each of the N nodes now represents a population of neurons, where there is all-to-all connectivity with homogeneous synaptic strength between neurons in a population and

all neurons in a population receive the same external input. Then $p_i(t)$ in equation 2.21, instead of signifying the active probability of a single neuron, represents the active proportion of a population. The remainder of the derivation for the mean behavior is identical, including the perturbation results, but the elements in the structural and functional connectivity matrices now apply to interactions between populations.

In order for realizations of the stochastic process to be accurate, however, the unit activity of neurons within populations must still be simulated correctly. Suppose that node i in the network has a population of n_i neurons; then an augmented state vector, $\mathbf{x_{pop}} \in \mathbb{R}^{\sum_{i=1}^{N} n_i}$, must be used during simulations that tracks the activity of all neurons in all populations individually and not simply the mean activity across a population. The external input is likewise augmented to $\mathbf{h_{pop}} \in \mathbb{R}^{\sum_{i=1}^{N} n_i}$, consisting of homogeneous block vectors $\mathbf{h_i} \in \mathbb{R}^{n_i}$ where

$$\mathbf{h_i} = h_i \mathbf{1^{n_i}}. \tag{A.16}$$

Similarly, the structural connectivity must be augmented so that $\mathbf{W_{pop}} \in \mathbb{R}^{(\sum_{i=1}^{N} n_i)\times(\sum_{i=1}^{N} n_i)}$ consists of homogeneous block matrices $\mathbf{W_{pop_{ij}}} \in \mathbb{R}^{n_i \times n_j}$, whose entries are scaled by population size as shown:

$$\mathbf{W_{pop_{ij}}} = \frac{w_{ij}}{n_j} \mathbf{1^{n_i}} \left(\mathbf{1^{n_j}}\right)^T. \tag{A.17}$$

In the population context, w_{ij} is therefore the synaptic drive to a neuron in node i when the entire population in node j is active.

The population model can be simulated using the steps in section A.4 but with $\mathbf{x_{pop}}$, $\mathbf{h_{pop}}$, and $\mathbf{W_{pop}}$ instead of $\mathbf{x}$, $\mathbf{h}$, and $\mathbf{W}$. In addition, the transition rates in equation A.13 must be computed on a neuron-by-neuron basis and the network transition rate in equation A.14 is summed across all neurons.

Acknowledgments

S.V.S. was supported by the Burroughs Wellcome Fund CASI Award 1007274 and the National Science Foundation CAREER Award 1055560.

References

Annegers, J. F., & Coan, S. P. (2000). The risks of epilepsy after traumatic brain injury. *Seizure*, *9*, 453–457.

Beattie, E. C., Stellwagen, D., Morishita, W., Bresnahan, J. C., Ha, B. K., Zastrow, M. V., . . . Malenka, R. C. (2002). Control of synaptic strength by glial TNFα. *Science*, *295*(5563), 2282–2285.

Benayoun, M., Cowan, J. D., van Drongelen, W., & Wallace, E. (2010). Avalanches in a stochastic model of spiking neurons. *PLoS Computational Biology*, *6*(7), 1–13.

Bradford, H. F. (1995). Glutamate, GABA and epilepsy. *Progress in Neurobiology*, *47*, 477–511.

Burns, S. P., Sritharan, D., Jouny, C., Bergey, G., Crone, N., Anderson, W. S., & Sarma, S. V. (2012). A network analysis of the dynamics of seizure. In *34th Annual International Conference of the IEEE EMBS*. Piscataway, NJ: IEEE.

de Boer, H. M., Mula, M., & Sander, J. W. (2008). The global burden and stigma of epilepsy. *Epilepsy and Behavior*, *12*, 540–546.

DeFelipe, J. (1999). Chandelier cells and epilepsy. *Brain*, *122*, 1807–1822.

Ehrens, D., Sritharan, D., & Sarma, S. V. (2014). Instability detector of a fragile neural network: Application to seizure detection in epilepsy. In *Proceedings of the 36th Annual IEEE Engineering in Medicine and Biology Society Conference*. Chicago, IL.

Fröhlich, F., Sejnowski, T. J., & Bazhenov, M. (2010). Network bistability mediates spontaneous transitions between normal and pathological brain states. *Journal of Neuroscience*, *30*(32), 10734–10743.

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS Computational Biology*, *5*(10), 1–26.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, *81*(25), 2340–2361.

Goutsias, J., & Jenkinson, G. (2013). Markovian dynamics on complex reaction networks. *Physics Reports*, *529*(2), 199–264.

Houweling, A. R., Bazhenov, M., Timofeev, I., Steriade, M., & Sejnowski, T. J. (2005). Homeostatic synaptic plasticity can explain post-traumatic epileptogenesis in chronically isolated neocortex. *Cerebral Cortex*, *15*(6), 834–845.

Jeffrey, N. L. (2003). The biology of epilepsy genes. *Ann. Rev. Neurosci.*, *26*, 599–625.

Jin, X., Prince, D. A., & Huguenard, J. R. (2006). Enhanced excitatory synaptic connectivity in layer V pyramidal neurons of chronically injured epileptogenic neocortex in rats. *Journal of Neuroscience*, *26*, 4891–4900.

Kramer, M. A., & Cash, S. S. (2012). Epilepsy as a disorder of cortical network organization. *Neuroscientist*, *18*(4), 360–372.

Kramer, M. A., Eden, U. T., Kolaczyk, E. D., Zepeda, R., Eskandar, E. N., & Cash, S. S. (2010). Coalescence and fragmentation of cortical networks during focal seizures. *Journal of Neuroscience*, *30*(30), 10076–10085.

Kramer, M. A., Kolaczyk, E. D., & Kirsch, H. E. (2008). Emergent network topology at seizure onset in humans. *Epilepsy research*, *79*(2), 173–186.

Kramer, M. A., Szeri, A. J., Sleigh, J. W., & Kirsch, H. E. (2007). Mechanisms of seizure propagation in a cortical model. *Journal of Computational Neuroscience*, *22*(1), 63–80.

Kramer, M. A., Truccolo, W., Eden, U. T., Lepage, K. Q., Hochberg, L. R., Eskandar, E. N., … & Cash, S. S. (2012). Human seizures self-terminate across spatial scales via a critical transition. *Proceedings of the National Academy of Sciences*, *109*(51), 21116–21121.

Lancman, M. E., & Morris, H. H. (1996). Epilepsy after central nervous system infection: clinical characteristics and outcome after epilepsy surgery. *Epilepsy Research*, *25*(3), 285–290.

Santaniello, S., Burns, S. P., Golby, A. J., Singer, J. M., Anderson, W. S., & Sarma, S. V. (2011). Quickest detection of drug-resistant seizures: An optimal control approach. *Epilepsy and Behavior*, *22*, S49–S60.

Santaniello, S., Sherman, D. L., Mirski, M. A., Thakor, N. V., & Sarma, S. V. (2011). A Bayesian framework for analyzing iEEG data from a rat model of epilepsy. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1435–1438). Piscataway, NJ: IEEE.

Santaniello, S., Sherman, D. L., Thakor, N. V., Eskandar, E. N., & Sarma, S. V. (2012). Optimal control-based Bayesian detection of clinical and behavioral state transitions. *IEEE Trans. Neural. Syst. Rehabil. Eng.*, *20*(5), 708–719.

Sritharan, D. (2013). *Network fragility: Application to neuronal networks in epilepsy*. Master's thesis, Johns Hopkins University.

Stellwagen, D., & Malenka, R. C. (2006). Synaptic scaling mediated by glial TNF-α. *Nature*, *440*(7087), 1054–1059.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos*. Cambridge: Perseus Books.

Traub, R. D., Whittington, M. A., Buhl, E. H., LeBeau, F. E. N., Bibbig, A., Boyd, S., . . . Baldeweg, T. (2001). A possible role for gap junctions in generation of very fast EEG oscillations preceding the onset of, and perhaps initiating, seizures. *Epilepsia*, *42*(2), 153–170.

Truccolo, W., Donoghue, J. A., Hochberg, L. R., Eskandar, E. N., Madsen, J. R., Anderson, W. S., . . . Cash, S. S. (2011). Single-neuron dynamics in human focal epilepsy. *Nature Neuroscience*, *14*(5), 635–643.

Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, *5*(2), 97–107.

Uhlhaas, P. J., & Singer, W. (2003). Neural synchrony in brain disorders: Relevance for cognitive dysfunctions and pathophysiology. *Neuron*, *52*(26), 599–625.

Volman, V., Bazhenov, M., & Sejnowski, T. J. (2011). Pattern of trauma determines the threshold for epileptic activity in a model of cortical deafferentation. *Proceedings of the National Academy of Sciences*, *108*(37), 15402–15407.

Volman, V., Sejnowski, T. J., & Bazhenov, M. (2011). Topological basis of epileptogenesis in a model of severe cortical trauma. *Journal of Neurophysiology*, *106*(4), 1933–1942.

Wallace, E., Benayoun, M., van Drongelen, W., & Cowan, J. D. (2011). Emergent oscillations in networks of stochastic spiking neurons. *PLoS ONE*, *6*(5), 1–16.

Whalen, A. J., Brennan, S. N., Sauer, T. D., & Schiff, S. J. (2012). Observability of neuronal network motifs. In *Proceedings of the 46th Annual Conference on Information Sciences and Systems* (pp. 1–5). Piscataway, NJ: IEEE.

Zhao, L., Beverlin, B., Netoff, T., & Nykamp, D. Q. (2011). Synchronization from second order network connectivity statistics. *Frontiers in Computational Neuroscience*, *5*(28).

Received January 10, 2014; accepted April 24, 2014.

This article has been cited by:

1. Daniel Ehrens, Duluxan Sritharan, Sridevi V. Sarma. 2015. Closed-loop control of a fragile network: application to seizure-like dynamics of an epilepsy model. *Frontiers in Neuroscience* **9**. . [CrossRef]
2. Robert B. Yaffe, Philip Borger, Pierre Megevand, David M. Groppe, Mark A. Kramer, Catherine J. Chu, Sabato Santaniello, Christian Meisel, Ashesh D. Mehta, Sridevi V. Sarma. 2015. Physiology of functional and effective networks in epilepsy. *Clinical Neurophysiology* **126**, 227-236. [CrossRef]
3. Daniel Ehrens, Duluxan Sritharan, Sridevi V. SarmaInstability detector of a fragile neural network: Application to seizure detection in epilepsy 6569-6572. [CrossRef]