2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Multistability, Bifurcations, and Biological Neural Networks: A Synaptic Drive Firing Model for Cerebral Cortex Transition in the Induction of General Anesthesia

Wassim M. Haddad, Qing Hui, and James M. Bailey

Abstract— **This paper focuses on multistability theory for discontinuous dynamical systems having a set of multiple isolated equilibria and/or a continuum of equilibria. Multistability is the property whereby the solutions of a dynamical system can alternate between two or more mutually exclusive Lyapunov stable and convergent equilibrium states under asymptotically slowly changing inputs or system parameters. In this paper, we extend the definition and theory of multistability to discontinuous autonomous dynamical systems. In particular, nontangency Lyapunov-based tests for multistability of discontinuous systems with Filippov solutions are established. The results are then applied to excitatory and inhibitory biological neuronal networks to explain the underlying mechanism of action for anesthesia and consciousness from a multistable dynamical system perspective, thereby providing a theoretical foundation for general anesthesia using the network properties of the brain.**

I. Introduction

Advances in neuroscience have been closely linked to mathematical modeling beginning with the integrate-and-fire model of Lapicque [1] and proceeding through the modeling of the action potential by Hodgkin and Huxley [2] to the current era of mathematical neuroscience; see [3] and the numerous references therein. Neuroscience has always had models to interpret experimental results from a high-level complex systems perspective; however, expressing these models with dynamic equations rather than words fosters precision, completeness, and self-consistency. Nonlinear dynamical system theory, in particular, can provide a framework for a rigorous description of the behavior of large-scale networks of neurons. A particularly interesting application of nonlinear dynamical systems theory to the neurosciences is to study phenomena of the central nervous system that exhibit nearly discontinuous transitions between macroscopic states. One such example exhibiting this phenomenon is the induction of general anesthesia [4–7].

The rational, safe, and effective utilization of any drug in the practice of medicine is grounded in an understanding of the pharmacodynamics of the drug, loosely defined as what the drug does to the body [8]. A very important measure of the pharmacodynamics of any drug is the drug concentration parameter EC_{50}, which reflects the drug dose at which the therapeutic effect is achieved in 50% of the cases. This concept is certainly applicable for the administration of general inhalational anesthetics, where the potency of the drug is defined by the minimum alveolar concentration (MAC) of the drug needed to prevent a response to noxious stimuli in 50% of administrations [9].

The MAC concept is intrinsically embedded in a probabilistic framework [8]. It is the concentration at which the probability of a response to a noxious stimulus is 0.5. Typically the MAC of a particular anesthetic is determined by administering various doses of the agent to a population of patients and determining the dose at which there is a 0.5 chance of responding to a noxious stimulus. (Technically, we identify the concentration in the alveoli, the fundamental functional gas exchange units of the lung, at which the chance of response is 0.5.) It has been possible, however, to conduct studies of single subjects, varying the anesthetic concentration and determining responsiveness. When this has been done, it has been noted that the transition from responsiveness to non-responsiveness in the individual patient is very sharp, almost an all-or-none transition [10]. This simply confirms the observations of generations of clinicians. And this raises the question of how to account for such a transition in terms of the known molecular properties of the anesthetic agent.

Although general anesthesia has been used in the clinical practice of medicine for over 150 years, the mechanism of action is still not fully understood [11] and is still under considerable investigation [4–7]. Theories range from a nonspecific perturbation of the lipid bilayer membrane of neurons, the cells responsible for the "information" function of the central nervous system, to the interaction of the anesthetic agent with specific protein receptors [11]. Early theories postulated that anesthesia is produced by disturbance of the physical properties of cell membranes. The work of Meyer and Overton [12], [13] demonstrated that for some anesthetics there was a correlation between anesthetic potency and solubility in fat-like solvents. This led to a theory that anesthesia resulted from a nonspecific perturbation of the lipid bilayer membrane of neurons [7], [14]. Subsequent research then found that membrane proteins performed functions of excitability and this led to a focus on anesthetic binding and perturbation of hydrophobic regions of membrane proteins [15]. Further research also revealed that some anesthetic gases follow the Meyer-Overton correlation but do not produce anesthesia and some Meyer-Overton gases are excitatory and can cause seizures [16]. These results led to the more common modern focus on the interaction of the anesthetic agent with specific protein receptors [11].

In particular, there has been extensive investigation of the influence of anesthetic agents on the binding of neurotransmitters to their postsynaptic receptors [6], [7]. A plethora of receptors have been investigated, including receptors for glycine, serotonin type 2 and 3, N-methyl-d-aspartate (NMDA), α-2 adrenoreceptors, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), histamine, acetycholine, and γ-aminobutyric acid (GABA). One attractive aspect of this focus on postsynaptic receptors is it

This work was supported in part by DTRA under Grant HDTRA1-10-1-0090, the US Army MRMC under Grant 08108002, and AFOSR under Grant FA9550-09-1-0249.

Q. Hui is with the Department of Mechanical Engineering, Texas Tech University, Lubbock, TX 79409-1021, USA (qing.hui@ttu.edu).

W. M. Haddad is with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0150, USA (wm.haddad@aerospace.gatech.edu).

J. M. Bailey is with the Northeast Georgia Medical Center, Gainesville, GA 30503, USA (james.bailey@nghs.com).

facilitates mathematical analysis on the basis of the effect of receptor binding on the postsynaptic potential. This is in marked contrast to the Meyer-Overton hypothesis, which failed to explicitly detail how a nonspecific perturbation of the lipid membrane would result in the anesthetic state.

In parallel with the investigation of the molecular interactions of general anesthetic agents, there has also been active investigation of the anatomic pathways involved in the transition from consciousness to anesthesia [4]. There is compelling evidence that the immobility created by some anesthetics is mediated at the level of the spinal cord. In contrast, functional imaging and electroencephalograph analysis has suggested that the site of suppression of consciousness is the thalamus, and thalamocortical tracts may play a critical role in the suppression of consciousness [7].

Despite these advances in our understanding of the molecular interactions of anesthetic agents and of specific anatomic loci for the action of anesthetic agents, there has been less development of a mathematical framework to understand this fascinating and clinically important phenomenon. It is certainly possible that if the mechanism of general anesthesia is the binding of the anesthetic agent to a specific receptor protein, then the nearly all-or-none transition from the awake state to the anesthetized state could be explained by a highly cooperative binding of the anesthetic to the receptor. In fact, it has been common to mathematically model the probability of responsiveness to drug concentration using the Hill equation, a simplified equation originally derived in 1909 to describe the cooperative binding of oxygen to the hemoglobin molecule [8]. However, to date, no single unifying receptor mediating general anesthesia has been identified.

Rather, the most likely explanation for the mechanisms of action of anesthetics lies in the network properties of the brain. It is well established that there are two general types of neurons in the central nervous system—excitatory and inhibitory—interconnected in a complex network. The action potential of a spiking neuron is propagated along the axon to synapses where chemical neurotransmitters are released that generate a postsynaptic potential on the dendrites of connected neurons. Excitatory neurons generate a depolarizing postsynaptic potential on the dendrite of the connected neuron and if the depolarization is of sufficient magnitude, then a spike will be induced in the connected neuron. In contrast, inhibitory neurons generate a hyperpolarizing postsynaptic potential; an effect that acts to maintain a quiescent state.

The human central nervous system involves a complex large-scale interconnected neural network involving feedforward and feedback (or recurrent) networks, with the brain serving as the central element of this network system. The brain is interconnected to receptors that transmit sensory information to the brain, and in turn the brain delivers action commands to effectors. The neural network of the brain consists of approximately 10^{11} neurons (nerve cells) with each having 10^4 to 10^5 connections interconnected through subnetworks or nuclei. The nuclei in turn consist of clusters of neurons each of which performs a specific and defined function.

The most basic characteristic of the neurons that comprise the central nervous system is the electrochemical potential gradient across the cell membrane. All cells of the human body maintain an electrochemical potential gradient between the inside of the cell and the surrounding milieu. Neurons have the capacity of excitability. If stimulated beyond a threshold the neuron will "fire" and produce a large voltage spike (the action potential) before returning to the resting potential [3], [17]. The neurons of the brain are connected in a complex network in which the firing of one neuron can be the stimulus for the firing of another neuron. A major focus of theoretical neuroscience has been describing neuronal behavior in terms of this electrochemical potential, both at the single neuron level but more ambitiously, at the level of multi-neuron networks. In this type of analysis the specific properties of the single neuron that are most relevant are how the spike of a one neuron alters the electrochemical potential of another neuron, and how this change in the potential results in a neuronal spike. The physical connection between neurons occurs in the synapse, a small gap between the axon, the extension of the cell body of the transmitting neuron, and the dendrite, the extension of the receiving neuron. The signal is transmitted by the release of a neurotransmitter from the axon into the synapse. This neurotransmitter diffuses across the synapse, binds to a postsynaptic receptor membrane protein on the dendrite, and alters the electrochemical potential of the receiving neuron.

There is considerable evidence that general anesthetics alter postsynaptic potentials [18], [19]. For example, it is possible that the anesthetic bifurcation to unconsciousness or the nearly all-or-none characteristic induction of anesthesia is a type of phase transition of the neural network. This possibility was first considered by Steyn-Ross *et al*. (see [20] and the references therein). Their focus was on the mean voltage of the soma, or cell body, of neurons. Specifically, the authors in [20] show that the biological change of state to anesthetic unconsciousness is analogous to a thermodynamic phase change involving a liquid to solid phase transition. For certain ranges of anesthetic concentrations, their first-order model predicts the existence of multiple steady states for brain activity leading to a transition from normal levels of cerebral cortical activity to a quiescent, low-firing state.

In this paper, we present an alternative approach to the possibility of neuronal network phase transition in terms of neuronal firing rates, using the concept of multistability for dynamical systems. Multistability is the property whereby the solutions of a dynamical system can alternate between two or more mutually exclusive Lyapunov stable and convergent states under asymptotically slowly changing inputs or system parameters. In particular, multistable systems give rise to the existence of multiple (isolated and/or a continuum of) stable equilibria involving a quasistatic-like behavior between these multiple semistable steady states [21]. Semistability is the property whereby the solutions to a dynamical system converge to Lyapunov stable equilibrium points determined by the system initial conditions [22], [23]. Multistability is ubiquitous in biological systems ranging from biochemical networks to ecosystems to gene regulation and cell replication [24]. Since molecular studies suggest that one possible mechanism of action of anesthetics is the inhibition of synaptic transmission in cortical neurons [18], [19], this suggests that general anesthesia is a phenomenon in which different equilibria can be attained with changing anesthetic agent concentrations. Hence, multistability theory can potentially provide a theoretical foundation for describing general anesthesia. Finally, we note that the proofs of the results of this paper can be found in [24].

II. Biological Neural Networks

The fundamental building block of the central nervous system, the *neuron*, can be divided into three functionally distinct parts, namely, the *dendrites*, *soma* (or cell body), and *axon*. The dendrites play the role of input devices that collect signals from other neurons and transmit them to the soma; whereas the soma generates a signal that is transmitted to other neurons by the axon. The axons of other neurons connect to the dendrites and soma surfaces by means of connectors called *synapses*. The behavior of the neuron is best described in terms of the electrochemical potential gradient across the cell membrane. If the voltage gradient

across the membrane increases to a critical threshold value, then there is a subsequent abrupt step-like increase in the potential gradient, the action potential. This action potential is transmitted from the soma along the axon to a dendrite of a receiving neuron. The action potential elicits the release of neurotransmitter molecules that diffuse to the dendrite of a "receiving" neuron. This alters the voltage gradient across the receiving neuron.

The electrochemical potential for a neuron can be described by a nonlinear four-state system [3]. Coupling these system equations for each neuron in a large neural population is computationally prohibitive. To simplify the mathematical modeling, it has been common to use phenomenological firing rate models for studying neural coding, memory, and network dynamics [3]. Firing rate models involve the averaged behavior of the spiking rates of groups of neurons rather than tracking the spike rate of each individual neuron cell. In such population models, the activity of a neuron, that is, the rate at which the neuron generates an action potential ("fires") is modeled as a function of the voltage (across the membrane). The "firing" of a neuron evokes voltage changes, postsynaptic potentials on receiving neurons; that is, neurons electrically connected to the firing neurons via axon-dendrite connections. In general, neurons are either excitatory or inhibitory depending on whether the postsynaptic potential increases or decreases the potential of the receiving neuron. In particular, excitatory neurotransmitters *depolarize* postsynaptic membranes by increasing membrane potentials and can collectively generate an action potential. Inhibitory neurotransmitters *hyperpolarize* the postsynaptic membrane by decreasing membrane potentials, thereby nullifying the actions of excitatory neurotransmitters and in certain cases prevent the generation of action potentials.

Biological neural network models predict a voltage in the receiving or postsynaptic neuron given by

$$V(t) = \sum_{i=1}^{n_{\mathrm{E}}} \sum_{j} \alpha_i^{\mathrm{E}}(t - t_j) + \sum_{i'=1}^{n_{\mathrm{I}}} \sum_{j'} \alpha_{i'}^{\mathrm{I}}(t - t_{j'}), \qquad (1)$$

where $i \in \{1, \ldots, n_{\mathrm{E}}\}$ and $i' \in \{1, \ldots, n_{\mathrm{I}}\}$ enumerate the action potential or firings of the excitatory and inhibitory transmitting (presynaptic) neurons at firing times t_j and $t_{j'}$, respectively, and $\alpha_i^{\mathrm{E}}(\cdot)$ and $\alpha_{i'}^{\mathrm{I}}(\cdot)$ are functions describing the evolution of the excitatory and inhibitory postsynaptic potentials, respectively.

Using a (possibly discontinuous) function $f_i(\cdot)$ to represent the firing rate of the ith neuron and assuming the firing rate is a function of the voltage $v_i^{\mathrm{E}}(\cdot)$ (resp., $v_i^{\mathrm{I}}(\cdot)$) across the membrane of the ith neuron given by $f_i(v_i^{\mathrm{E}}(A^{\mathrm{XY}}))$ (resp., $f_i(v_i^{\mathrm{I}}(A^{\mathrm{XY}}))$), it follows that

$$\begin{aligned} v_i^{\mathrm{E}}(t) &= \sum_{j=1, j\neq i}^{n_{\mathrm{E}}} A_{ij}^{\mathrm{EE}} \int_{-\infty}^{t} \alpha_j^{\mathrm{E}}(t-\tau) f_i(v_j^{\mathrm{E}}(\tau)) \mathrm{d}\tau \\ &\quad + \sum_{j'=1}^{n_{\mathrm{I}}} A_{ij'}^{\mathrm{EI}} \int_{-\infty}^{t} \alpha_{j'}^{\mathrm{I}}(t-\tau) f_i(v_{j'}^{\mathrm{I}}(\tau)) \mathrm{d}\tau \\ &\quad + I_i^{\mathrm{E}}(t), \quad i = 1, \ldots, n_{\mathrm{E}}, \qquad (2) \\ v_i^{\mathrm{I}}(t) &= \sum_{j=1}^{n_{\mathrm{E}}} A_{ij}^{\mathrm{IE}} \int_{-\infty}^{t} \alpha_j^{\mathrm{E}}(t-\tau) f_i(v_j^{\mathrm{E}}(\tau)) \mathrm{d}\tau \\ &\quad + \sum_{j'=1, j'\neq i}^{n_{\mathrm{I}}} A_{ij'}^{\mathrm{II}} \int_{-\infty}^{t} \alpha_{j'}^{\mathrm{I}}(t-\tau) f_i(v_{j'}^{\mathrm{I}}(\tau)) \mathrm{d}\tau \\ &\quad + I_i^{\mathrm{I}}(t), \quad i = 1, \ldots, n_{\mathrm{I}}, \qquad (3) \end{aligned}$$

where the *neuronal connectivity matrix* A^{XY} is such that $A_{ij}^{\mathrm{XY}} \neq 0$, $\mathrm{X}, \mathrm{Y} \in \{\mathrm{E}, \mathrm{I}\}$, if the jth neuron is connected (i.e., contributes a postsynaptic potential) to the ith neuron and $A_{ij}^{\mathrm{XY}} = 0$ otherwise, and where $I_i^{\mathrm{E}}(\cdot)$ and $I_i^{\mathrm{I}}(\cdot)$ are continuous synaptic current functions. Note that $A_{ii}^{\mathrm{EE}} = A_{ii}^{\mathrm{II}} = 0$ by definition.

Next, defining the *synaptic drive* of each (excitatory or inhibitory) neuron by

$$S_i^{(\mathrm{E,I})}(t) \triangleq \int_{-\infty}^{t} \alpha_i^{(\mathrm{E,I})}(t-\tau) f_i(v_i^{(\mathrm{E,I})}(\tau)) \mathrm{d}\tau, \qquad (4)$$

and assuming $\alpha_i^{(\mathrm{E,I})}(t) = B^{(\mathrm{E,I})} e^{-\frac{t}{\lambda_i^{(\mathrm{E,I})}}}$, where $B^{(\mathrm{E,I})} = B^{\mathrm{E}}$ if the ith neuron is excitatory and $B^{(\mathrm{E,I})} = B^{\mathrm{I}}$ if the ith neuron is inhibitory, and similarly for $S_i^{(\mathrm{E,I})}$, $v_i^{(\mathrm{E,I})}$, $\alpha_i^{(\mathrm{E,I})}$, and $\lambda_i^{(\mathrm{E,I})}$, it follows from (4) and the expresion for $\alpha_i^{(\mathrm{E,I})}(t)$ that

$$\frac{\mathrm{d}S_i^{(\mathrm{E,I})}(t)}{\mathrm{d}t} = -\frac{1}{\lambda_i^{(\mathrm{E,I})}} S_i^{(\mathrm{E,I})}(t) + B^{(\mathrm{E,I})} f_i(v_i^{(\mathrm{E,I})}(t)). \qquad (5)$$

Now, using the expressions for the excitatory and inhibitory voltage given by (2) and (3), respectively, it follows that

$$\begin{aligned} \frac{\mathrm{d}S_i^{\mathrm{E}}(t)}{\mathrm{d}t} &= -\frac{1}{\lambda_i^{\mathrm{E}}} S_i^{\mathrm{E}}(t) + B^{\mathrm{E}} f_i \left(\sum_{j=1, j\neq i}^{n_{\mathrm{E}}} A_{ij}^{\mathrm{EE}} S_j^{\mathrm{E}}(t) \right. \\ &\quad \left. + \sum_{j'=1}^{n_{\mathrm{I}}} A_{ij'}^{\mathrm{EI}} S_{j'}^{\mathrm{I}}(t) + I_i^{\mathrm{E}}(t) \right), \ i = 1, \ldots, n_{\mathrm{E}}, \qquad (6) \\ \frac{\mathrm{d}S_i^{\mathrm{I}}(t)}{\mathrm{d}t} &= -\frac{1}{\lambda_i^{\mathrm{I}}} S_i^{\mathrm{I}}(t) + B^{\mathrm{I}} f_i \left(\sum_{j'=1, j'\neq i}^{n_{\mathrm{I}}} A_{ij'}^{\mathrm{II}} S_{j'}^{\mathrm{I}}(t) \right. \\ &\quad \left. + \sum_{j=1}^{n_{\mathrm{E}}} A_{ij}^{\mathrm{IE}} S_j^{\mathrm{E}}(t) + I_i^{\mathrm{I}}(t) \right), \ i = 1, \ldots, n_{\mathrm{I}}. \qquad (7) \end{aligned}$$

The above analysis reveals that a form for capturing the neuroelectronic behavior of biological excitatory or inhibitory neuronal networks can be written as

$$\begin{aligned} \frac{\mathrm{d}S_i(t)}{\mathrm{d}t} &= -\tau_i S_i(t) + f_i \left(\sum_{j=1}^{n} A_{ij} S_j(t) + I_i(t) \right), \\ S_i(0) &= S_{i0}, \quad t \geq 0, \quad i = 1, \ldots, n, \qquad (8) \end{aligned}$$

where $S_i(t) \in \mathcal{D} \subseteq \mathbb{R}$ is the ith synaptic drive, $I_i(t) \in \mathbb{R}$ denotes the synaptic current of the ith neuron, A_{ij} is a constant representing the coupling strength of the jth neuron on the ith neuron, $\tau_i \triangleq 1/\lambda_i$ is a time constant, and $f_i(\cdot)$ is a nonlinear activation function describing the relationship between the synaptic current and the firing rate of the neuron. In this paper, we assume that $f_i(\cdot)$ can be a discontinuous function such as a hard limiter or a continuous function such as a half-wave rectification function or a sigmoidal function. Specifically, for a typical neuron

$$f_i(x) = [x]_+, \qquad (9)$$

where $i \in \{1, \ldots, n\}$ and $[x]_+ = x$ if $x \geq 0$, and $[x]_+ = 0$ otherwise. Alternatively, we can approximate $f_i(x), i \in \{1, \ldots, n\}$, by the sigmoidal function

$$f_i(x) = \frac{x e^{\gamma x}}{1 + e^{\gamma x}}, \quad \gamma >> 0. \qquad (10)$$

III. Multistability Theory

Multistability is the property whereby the solutions of a dynamical system can alternate between two or more mutually exclusive semistable states under asymptotically slowly changing inputs or system parameters. In particular, the state of a multistable system converges to Lyapunov stable equilibria that belong to an equilibrium set that has a multivalued hybrid topological structure consisting of isolated points and closed sets homeomorphic to intervals on the real line. In this and the next section, we develop a *general*, nontangency- and weak stability-based framework for addressing multistability of nonlinear dynamical systems. This work is inspired by the study of nontangency-based Lyapunov tests for convergence and stability addressed in [22] as well as weak stability notions for dynamical systems addressed in [25].

To develop the notion of multistability, consider the autonomous differential equation given by

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0, \quad \text{a.a.} \quad t \geq 0, \tag{11}$$

where $f : \mathbb{R}^q \to \mathbb{R}^q$ is Lebesgue measurable and locally essentially bounded [26], that is, f is bounded on a bounded neighborhood of every point, excluding sets of measure zero, and let $\mathcal{E}_{\mathrm{e}} \triangleq \{x \in \mathbb{R}^q : f(x) = 0\}$ denote the set of equilibria for (11).

Definition 3.1 ([26]): An absolutely continuous function $x : [0, \tau] \to \mathbb{R}^q$ is said to be a *Filippov solution* of (11) on the interval $[0, \tau]$ with initial condition $x(0) = x_0$, if $x(t)$ satisfies $\frac{\mathrm{d}}{\mathrm{d}t}x(t) \in \mathcal{K}[f](x(t))$ for almost every $t \in [0, \tau]$, where the *Filippov set-valued map* $\mathcal{K}[f] : \mathbb{R}^q \to \mathcal{B}(\mathbb{R}^q)$ is defined by $\mathcal{K}[f](x) \triangleq \bigcap_{\delta>0} \bigcap_{\mu(\mathcal{S})=0} \overline{\mathrm{co}}\,\{f(\mathcal{B}_\delta(x)\backslash\mathcal{S})\}$ for $x \in \mathbb{R}^q$, where $\mathcal{B}_\delta(x)$ denotes the *open ball centered* at x with *radius* δ, $\mathcal{B}(\mathbb{R}^q)$ denotes the collection of all subsets of $\mathbb{R}^q$, $\mu(\cdot)$ denotes the Lebesgue measure in $\mathbb{R}^q$, and "$\overline{\mathrm{co}}$" denotes the convex closure.

Note that $\mathcal{K}[f] : \mathbb{R}^q \to \mathcal{B}(\mathbb{R}^q)$ is a map that assigns sets to points. Dynamical systems of the form given by $\frac{\mathrm{d}}{\mathrm{d}t}x(t) \in \mathcal{K}[f](x(t))$ are called *differential inclusions* [27] and for each state $x \in \mathbb{R}^q$, they specify a *set* of possible evolutions rather than a single one. Note that an equilibrium point of (11) is a point $x_{\mathrm{e}} \in \mathbb{R}^q$ such that $0 \in \mathcal{K}[f](x_{\mathrm{e}})$.

Definition 3.2: Consider the nonlinear dynamical system (11). We say that the dynamical system (11) is *multistable* if i) there exists more than one equilibrium point of (11) in $\mathbb{R}^q$; ii) all solutions to (11) converge to one of these equilibrium points; and iii) almost all solutions to (11) converge to Lyapunov stable equilibria; that is, the set of initial conditions driving the solutions of (11) to unstable equilibria has Lebesgue measure zero.

It is important to note that our definition of multistability is different from the definition given in [28]. Specifically, pertaining to condition iii), the definition of multistability given in [28] requires that almost all solutions to (11) converge to asymptotically stable equilibria. This key difference allows for the dynamical system (11) to possess a continuum of equilibria, rather than merely isolated equilibria. As we see later, if f_i is of the form given by (9), then (8) has a continuum of equilibria under certain conditions, and hence, (11) is semistable in the sense of iii) [27]. Hence, in this case, it is more appropriate to use Definition 3.2 to characterize multistability.

Almost all of the existing results on multistability theory rely on linearization techniques based on the Hartman-Grobman theorem involving the fact that the linearized system has the same topological property as the original system around a hyperbolic fixed point. When the system fixed point is not hyperbolic, however, these techniques fail to predict multistability. In this case, checking multistability becomes a daunting task. Rather than checking the transversality condition for hyperbolicity, in this paper we present a new approach for guaranteeing multistability using equilibria-independent, semidefinite Lyapunov function methods. In particular, using the geometric structure of the vector field f for a given dynamical system, we develop nontangency-based Lyapunov tests for verifying conditions ii) and iii) in Definition 3.2 involving convergence and Lyapunov stability almost everywhere.

IV. Direction Cones, Nontangency, Convergence, and Nonsmooth Multistability

To show condition ii) in Definition 3.2 holds for dynamical systems of the form given by (11), we adopt the notion of nontangency [22], [27] to develop nontangency-based Lyapunov tests for convergence. Specifically, the authors in [22] develop a general framework for nontangency-based Lyapunov tests for the convergence of dynamical systems described by ordinary differential equations with continuous vector fields. In [27], the authors extend some of the results of [22] to nonsmooth dynamical systems, that is, systems described by ordinary differential equations with the discontinuous right-hand sides. Since the vector field f characterizing biological neural networks can involve either continuous or discontinuous vector fields, we use the more general definition for nontangency presented in [27]. Before stating our results, we introduce some notation and definitions.

A set $\mathcal{E} \subseteq \mathbb{R}^q$ is *connected* if and only if every pair of open sets $\mathcal{U}_i \subseteq \mathbb{R}^q$, $i = 1, 2$, satisfying $\mathcal{E} \subseteq \mathcal{U}_1 \cup \mathcal{U}_2$ and $\mathcal{U}_i \cap \mathcal{E} \neq \varnothing$, $i = 1, 2$, has a nonempty intersection. A *connected component* of the set $\mathcal{E} \subseteq \mathbb{R}^q$ is a connected subset of $\mathcal{E}$ that is not properly contained in any connected subset of $\mathcal{E}$. Given a set $\mathcal{E} \subseteq \mathbb{R}^q$, let $\operatorname{coco}\mathcal{E}$ denote the convex cone generated by $\mathcal{E}$.

Definition 4.1: Given $x \in \mathbb{R}^q$, the *direction cone* $\mathcal{F}_x$ of the vector field f at x is the intersection of closed convex cones of the form $\overline{\bigcap_{\mu(\mathcal{S})=0} \operatorname{coco}\{f(\mathcal{U}\backslash\mathcal{S})\}}$, where $\mathcal{U} \subseteq \mathbb{R}^q$ is an open neighborhood of x and $\overline{\mathcal{Q}}$ denotes the closure of the set $\mathcal{Q}$. Let $\mathcal{E} \subseteq \mathbb{R}^q$. A vector $v \in \mathbb{R}^q$ is *tangent* to $\mathcal{E}$ at $z \in \mathcal{E}$ if there exist a sequence $\{z_i\}_{i=1}^{\infty}$ in $\mathcal{E}$ converging to z and a sequence $\{h_i\}_{i=1}^{\infty}$ of positive real numbers converging to zero such that $\lim_{i\to\infty} \frac{1}{h_i}(z_i - z) = v$. The *tangent cone* to $\mathcal{E}$ at z is the closed cone $T_z\mathcal{E}$ of all vectors tangent to $\mathcal{E}$ at z. Finally, the vector field f is *nontangent* to the set $\mathcal{E}$ at the point $z \in \mathcal{E}$ if $T_z\mathcal{E} \cap \mathcal{F}_z \subseteq \{0\}$.

Next, let $\omega(x)$ be the positive limit set of (11) at x and let $\dot{V}$ denote the *set-valued Lie derivative* [27] for Filippov solutions to (11) for a given lower semicontinuous function $V(\cdot)$. The next result generalizes the Krasovskii-LaSalle invariant set theorem to the case where $V(\cdot)$ is lower semicontinuous and $f(\cdot)$ is Lebesgue measurable and locally essentially bounded.

Proposition 4.1: Assume that $V : \mathbb{R}^q \to \mathbb{R}$ is a lower semicontinuous function such that $\dot{V}$ is defined on $\mathbb{R}^q$ and $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^q$. Let $x \in \mathbb{R}^q$ be such that a solution $\psi(t, x)$ of (11) is bounded. Then $\omega(x) \subseteq \mathcal{M}$, where $\mathcal{M} \triangleq \bigcup_{\gamma\in\mathbb{R}} \mathcal{M}_\gamma$ and $\mathcal{M}_\gamma$ denotes the largest weakly invariant set contained in $\mathcal{R}_\gamma \triangleq \bigcap_{c>\gamma} \overline{V^{-1}([\gamma, c])}$.

The following theorem gives sufficient conditions for convergence using the nontangency between the vector field f and invariant subsets of the level sets of a lower semicontinuous semidefinite Lyapunov function.

Theorem 4.1: Assume that $V : \mathbb{R}^q \to \mathbb{R}$ is a lower semicontinuous function such that $\dot{V}$ is defined on $\mathbb{R}^q$ and $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^q$. Let $x \in \mathbb{R}^q$ be such that the solution of (11) is bounded. If $\mathcal{M}$ defined in Proposition 4.1 is composed of isolated equilibria of (11) or f is nontangent to $\mathcal{M}$ at every point in $\mathcal{M}$, then $\lim_{t\to\infty} \psi(t, x)$ exists.

Corollary 4.1: Assume that $V : \mathbb{R}^q \to \mathbb{R}$ is a continuous function such that $\dot{V}$ is defined on $\mathbb{R}^q$ and $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^q$. Let $x \in \mathbb{R}^q$ be such that the solution of (11) is bounded and let $\mathcal{N}$ denote the largest weakly invariant set contained in $\dot{V}^{-1}(0)$. If $\mathcal{N}$ is composed of isolated equilibria of (11) or f is nontangent to $\mathcal{N}$ at every point in $\mathcal{N}$, then $\lim_{t\to\infty} \psi(t, x)$ exists.

Next, we present convergence results for (11) in the case where a subset of the equilibria of (11) are Lyapunov stable.

Theorem 4.2: Assume that there exists a lower semicontinuous function $V : \mathbb{R}^q \to \mathbb{R}$ such that $\dot{V}$ is defined on $\mathbb{R}^q$ and satisfies $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^q$. Let $x \in \mathbb{R}^q$ be such that the solution of (11) is bounded, let $\mathcal{S}_{\mathrm{e}} \subseteq \mathcal{E}_{\mathrm{e}}$ denote the set of equilibria of (11) that are Lyapunov stable, and let $\mathcal{C}_0$ denote the largest weakly invariant set contained in $\mathcal{M}$. In addition, for every $k = 0, 1, 2, \ldots$, let $\mathcal{H}_k \subseteq \mathcal{C}_k$ denote the set of points in $\mathcal{C}_k$ where f is not nontangent to $\mathcal{C}_k$ and let $\mathcal{C}_{k+1} \subseteq \mathcal{H}_k$ denote the largest weakly invariant set contained in $\mathcal{H}_k$. If $\mathcal{H}_k \subseteq \mathcal{S}_{\mathrm{e}}$ for some $k \in \{0, 1, 2, \ldots\}$, then $\lim_{t\to\infty} \psi(t, x)$ exists.

Corollary 4.2: Assume that $V : \mathbb{R}^q \to \mathbb{R}$ is a continuous function such that $\dot{V}$ is defined on $\mathbb{R}^q$ and $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^q$. Let $x \in \mathbb{R}^q$ be such that the solution of (11) is bounded and let $\mathcal{N}$ denote the largest weakly invariant set contained in $\dot{V}^{-1}(0)$. If every point in $\mathcal{N}$ is Lyapunov stable, then $\lim_{t\to\infty} \psi(t, x)$ exists.

The results of this section can be used to verify condition $ii)$ of Definition 3.2, that is, convergence of all solutions to (11). Now, we need only show that all solutions converging to unstable equilibria correspond to a Lebesgue zero-measure set of initial conditions to establish multistability. This is addressed in the next section.

V. Semistability Almost Everywhere

In this section, we use the results of the previous section to derive sufficient conditions for convergence and Lyapunov stability almost everywhere of (11). This result does not require any assumptions on the sign definiteness of the Lyapunov function. The assumption needed is a nontangency condition on the vector field to the closure of the zero-level set of the Lyapunov function derivative.

Theorem 5.1: Assume that $V : \mathbb{R}^q \to \mathbb{R}$ is a continuous function such that $\dot{V}$ is defined almost everywhere on $\mathbb{R}^q$ and $\dot{V} \leq 0$ wherever $\dot{V}$ is well defined. Let $x \in \mathbb{R}^q$ be such that the solution of (11) is bounded and let $\mathcal{N}$ denote the largest weakly invariant set contained in $\dot{V}^{-1}(0)$. If either every point in $\mathcal{N}$ is Lyapunov stable or f is nontangent to $\mathcal{N}$ at every point in $\mathcal{N}$, then almost all solutions of (11) converge to Lyapunov stable equilibria.

VI. Applications to Excitatory-Inhibitory Biological Networks

The form of biological neural network models given by (8) represents a wide range of firing rate population models appearing in neuroscience [3], [17]. In this section, we will consider an important class of these network systems involving *excitatory-inhibitory networks*. The firing rate is a nonnegative quantity representing the probability of the firing action potential by the neuron and can be interpreted as a measure of the neuron's activity. Since the firing rate of the excitatory-inhibitory network is nonnegative, all solutions of physical interest always take values in the nonnegative orthant of the state space for nonnegative initial conditions. For such systems, which evolve on possibly closed positively invariant subsets of $\mathbb{R}^q$, it is natural to consider the nonnegative orthant $\overline{\mathbb{R}}^q_+$ as their state space, and hence, these systems are nonnegative dynamical systems [23]. In this case, all of our stability and convergence results developed in Sections IV and V hold with respect to $\overline{\mathbb{R}}^q_+$ by replacing $\mathbb{R}^q$ with $\overline{\mathbb{R}}^q_+$.

The following result, which follows from Proposition 2.1 of [23], gives necessary and sufficient conditions for the activation function of the excitatory-inhibitory networks such that the firing rates $S_i(t)$ remain in the nonnegative orthant of the state space. For the statement of the next result recall that f is *nonnegative* if and only if $f(x) \geq\geq 0$, $x \in \overline{\mathbb{R}}^n_+$, where "$\geq\geq$" denotes a component-wise inequality.

Proposition 6.1: Consider the excitatory-inhibitory network given by (8). The firing rate vector $S(t) \triangleq [S_1(t), \ldots, S_n(t)]^{\mathrm{T}} \in \mathbb{R}^n$ remains in the nonnegative orthant of the state space $\overline{\mathbb{R}}^n_+$ for all $t \geq 0$ if and only if for every $S_i \geq 0$ and $I_i \geq 0$, $i = 1, \ldots, n$, the function $\tilde{f} = [f_1(\sum_{j=1}^n A_{1j}S_j + I_1), \ldots, f_n(\sum_{j=1}^n A_{nj}S_j + I_n)]^{\mathrm{T}} : \mathbb{R}^n \to \mathbb{R}^n$ is nonnegative.

The vector-matrix form of (8) can be written as

$$\dot{S}(t) = -LS(t) + \tilde{f}(AS(t) + I(t)), \quad S(0) = S_0, \qquad (12)$$

where $t \geq 0, L \triangleq \mathrm{diag}[\tau_1, \ldots, \tau_n] \in \mathbb{R}^{n\times n}$ is a time constant matrix, $A \triangleq [A_{ij}] \in \mathbb{R}^{n\times n}$ is a matrix representing the strength of the synaptic interconnections, and $I(t) \triangleq [I_1(t), \ldots, I_n(t)]^{\mathrm{T}} \in \mathbb{R}^n$, $t \geq 0$, is a vector of synaptic currents.

First, we consider the case where $f_i(\cdot)$ is given by (9) and $I(t) \equiv 0$. By Proposition 6.1, for $S_0 \in \overline{\mathbb{R}}^n_+$, $S(t) \in \overline{\mathbb{R}}^n_+$ for all $t \geq 0$ if and only if A is a *nonnegative matrix*, that is, $A_{ij} \geq 0$, $i, j = 1, \ldots, n$. Hence, let A be nonnegative. Then, (12) collapses to the linear model given by

$$\dot{S}(t) = (A - L)S(t), \quad S(0) = S_0 \in \overline{\mathbb{R}}^n_+, \quad t \geq 0. \qquad (13)$$

The linear system (13) has a continuum, and hence, multiple equilibria if and only if the rank of $A - L$ is less than n. To analyze the multistability of (13), we assume that the rank of $A - L$ is less than n. In this case, (13) has a continuum of equilibria and $\lim_{t\to\infty} S(t)$ exists if and only if $A - L$ is semistable [23]. Hence, multistability of (13) reduces to checking semistability of (13).

Next, we consider the case where $f_i(\cdot)$ is given by (10) and $I(t) \equiv 0$. In this case, we assume that $A_{ij} \geq 0$, $i, j = 1, \ldots, n$. Consider the nonnegative function $U(x) = p^{\mathrm{T}}x$, where $x \in \overline{\mathbb{R}}^n_+$ and $p \in \overline{\mathbb{R}}^n_+$. It follows from the Perron-Frobenius theorem [23] that there exist positive vectors $p, r \in \mathbb{R}^n_+$ such that $(A - L)^{\mathrm{T}}p + r = 0$. Hence,

$$\begin{aligned} \dot{U}(S) &= -p^{\mathrm{T}}LS + p^{\mathrm{T}}f(AS) \leq -p^{\mathrm{T}}LS + p^{\mathrm{T}}AS \\ &= p^{\mathrm{T}}(A - L)S = -r^{\mathrm{T}}S \leq 0, \quad S \in \overline{\mathbb{R}}^n_+. \end{aligned} \qquad (14)$$

Thus, all the solutions of (8) are bounded.

Next, assume that the set $\mathcal{E}_{\mathrm{e}}$ has a nonzero element, that is, $\tilde{f}(Ax) - Lx = 0$ has a nonzero solution for $x \in \overline{\mathbb{R}}^n_+$. The following result is immediate.

Proposition 6.2: If $\mathcal{E}_{\mathrm{e}}$ consists of multiple isolated equilibria, then rank $(A - L) = n$.

If rank $(A-L) < n$, then $\mathcal{E}_{\rm e}$ has a continuum of equilibria. Now, consider the following two cases.

Case 1. Either $\mathcal{E}_{\rm e}$ consists of multiple isolated equilibria or $\mathcal{E}_{\rm e}$ consists of both multiple isolated equilibria and a continuum of equilibria, and rank $(L-A) = n$.

In this case, consider the function $V(x) = \frac{1}{2}x^{\rm T}Kx$, $x \in \overline{\mathbb{R}}_+^n$, where K is symmetric but not sign definite, and note that

$$\dot{V}(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} K_{ij}x_j\dot{x}_i = -\sum_{i=1}^{n}\sum_{j=1}^{n}\tau_i K_{ij}x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_{ij}x_j(\sum_{k=1}^{n}A_{ik}x_k)e^{\gamma(\sum_{k=1}^{n}A_{ik}x_k)}}{1+e^{\gamma(\sum_{k=1}^{n}A_{ik}x_k)}}. \quad (15)$$

Now, since $\sum_{k=1}^{n} A_{ik}x_k \geq 0$, it follows that

$$\frac{1}{2} \leq \frac{e^{\gamma(\sum_{k=1}^{n}A_{ik}x_k)}}{1+e^{\gamma(\sum_{k=1}^{n}A_{ik}x_k)}} < 1. \quad (16)$$

Hence,

$$\begin{aligned}\dot{V}(x) &\leq -\sum_{i=1}^{n}\sum_{j=1}^{n}\tau_i K_{ij}x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left[\frac{3}{4}+\frac{1}{4}{\rm sign}(K_{ij})\right]K_{ij}A_{ik}x_k x_j \\ &= -x^{\rm T}KLx + x^{\rm T}\overline{K}Ax = -x^{\rm T}(KL-\overline{K}A)x, \quad (17)\end{aligned}$$

where the entries of $\overline{K}$ are given by $\overline{K}_{ij} = \left[\frac{3}{4}+\frac{1}{4}{\rm sign}(K_{ij})\right]K_{ij}$. If K is chosen such that $(KL-\overline{K}A) \geq 0$, then $\dot{V}(x) \leq 0$ for all $x \in \overline{\mathbb{R}}_+^n$. Thus, $\dot{V}^{-1}(0) \subseteq \mathcal{N}(KL-\overline{K}A)$, where $\mathcal{N}(\cdot)$ denotes null space, which implies that the largest weakly invariant set of $\dot{V}^{-1}(0)$ is contained in $\mathcal{N}(KL-\overline{K}A)\cap\mathcal{E}_{\rm e}$.

If the system

$$\dot{x}(t) = (\overline{K}A - KL)x(t), \quad x(0) = x_0, \quad t \geq 0, \quad (18)$$

is Lyapunov stable, then it follows from Corollary 4.2 that all the solutions converge to one of the Lyapunov stable equilibria in $\mathcal{N}(KL-\overline{K}A)\cap\mathcal{E}_{\rm e}$ for (12). Hence, it follows from Theorem 5.1 that (12) is multistable.

Alternatively, if the vector field f of (12) is nontangent to $\mathcal{N}(KL-\overline{K}A)\cap\mathcal{E}_{\rm e}$ at every point in $\mathcal{N}(KL-\overline{K}A)\cap\mathcal{E}_{\rm e}$, then it follows from Corollary 4.1 that all the solutions of (12) converge to one of the equilibria in $\mathcal{N}(KL-\overline{K}A)\cap\mathcal{E}_{\rm e}$. Hence, it follows from Theorem 5.1 that (12) is multistable.

Case 2. rank $(L-A) < n$.

In this case, the null space of $L-A$ is a subset of $\mathcal{E}_{\rm e}$ since $-Lx+\tilde{f}(Ax) \leq\leq (A-L)x$ for all $x \in \overline{\mathbb{R}}_+^n$. Consider the function $V(x) = \frac{1}{2}x^{\rm T}x$, $x \in \overline{\mathbb{R}}_+^n$, and note that

$$\begin{aligned}\dot{V}(x) &= -\sum_{i=1}^{n}\tau_i x_i^2 + \sum_{i=1}^{n}\frac{x_i(\sum_{j=1}^{n}A_{ij}x_j)e^{\gamma(\sum_{j=1}^{n}A_{ij}x_j)}}{1+e^{\gamma(\sum_{j=1}^{n}A_{ij}x_j)}} \\ &\leq -\sum_{i=1}^{n}\tau_i x_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}x_i x_j = -x^{\rm T}(L-A)x. \quad (19)\end{aligned}$$

If all the τ_i's are sufficiently large, then we can ensure that $L-A \geq 0$. In this case, $\dot{V}(x) \leq 0$ for all $x \in \overline{\mathbb{R}}_+^n$. Thus, $\dot{V}^{-1}(0) \subseteq \mathcal{N}(L-A) \subseteq \mathcal{E}_{\rm e}$ and the largest weakly invariant set of $\dot{V}^{-1}(0)$ is contained in $\mathcal{N}(L-A)$.

If (13) is Lyapunov stable, then it follows from Corollary 4.2 that all the solutions of (12) converge to one of the Lyapunov stable equilibria in $\mathcal{N}(L-A)$. Now, it follows from Theorem 5.1 that (12) is multistable. Alternatively, if the vector field f of (12) is nontangent to $\mathcal{N}(L-A)$ at every point in $\mathcal{N}(L-A)$, then it follows from Corollary 4.1 that all the solutions of (12) converge to one of the equilibria in $\mathcal{N}(L-A)$. Hence, it follows from Theorem 5.1 that (12) is multistable.

VII. A Two-Class Mean Excitatory and Inhibitory Synaptic Drive Model

The excitatory and inhibitory neural network model given by (6) and (7) possesses multiple equilibria. For certain values of the model parameters it can be shown that as the inhibitory time constants $\lambda_i^{\rm I}$ get larger, the equilibrium states can flip their stabilities. Since molecular studies suggest that one possible mechanism of action of anesthetics is the prolongation of the time constants of inhibitory neurons [18], [19], this suggests that general anesthesia is a phenomenon in which different equilibria can be attained with changing anesthetic agent concentrations. In this section, we develop a simplified model involving mean excitatory and inhibitory drives to explore this multistability phenomenon.

Consider the excitatory and inhibitory synaptic drive model given by (6) and (7) with $f_i(\cdot) = f(\cdot)$, $I_i^{\rm E}(t) = I^{\rm E}$, $I_i^{\rm I}(t) = I^{\rm I}$, $B^{\rm E} = B^{\rm I} = 1$, $\lambda_i^{\rm E} = \lambda^{\rm E}$, and $\lambda_i^{\rm I} = \lambda^{\rm I}$. In this case, (6) and (7) become

$$\begin{aligned}\frac{{\rm d}S_i^{\rm E}(t)}{{\rm d}t} &= f\left(\sum_{j=1}^{n_{\rm E}}A_{ij}^{\rm EE}S_j^{\rm E}(t) + \sum_{k=1}^{n_{\rm I}}A_{ik}^{\rm EI}S_k^{\rm I}(t) + I^{\rm E}\right) - \frac{1}{\lambda^{\rm E}}S_i^{\rm E}(t), \quad i = 1,\ldots,n_{\rm E}, \quad (20)\\ \frac{{\rm d}S_i^{\rm I}(t)}{{\rm d}t} &= f\left(\sum_{j=1}^{n_{\rm E}}A_{ij}^{\rm IE}S_j^{\rm E}(t) + \sum_{k=1}^{n_{\rm I}}A_{ik}^{\rm II}S_k^{\rm I}(t) + I^{\rm I}\right) - \frac{1}{\lambda^{\rm I}}S_i^{\rm I}(t), \quad i = 1,\ldots,n_{\rm I}, \quad (21)\end{aligned}$$

where $f(\cdot)$ is given by (10) and $A_{ii}^{\rm EE} = A_{ii}^{\rm II} = 0$.

Next, let $A_{ij}^{\rm EE} = \overline{A}^{\rm EE} + \Delta_{ij}^{\rm EE}, A_{ij}^{\rm EI} = \overline{A}^{\rm EI} + \Delta_{ij}^{\rm EI}, A_{ij}^{\rm IE} = \overline{A}^{\rm IE} + \Delta_{ij}^{\rm IE}$, and $A_{ij}^{\rm II} = \overline{A}^{\rm II} + \Delta_{ij}^{\rm II}$, where $\overline{(\cdot)}$ denotes mean and $\Delta_{ij}^{\rm XY}, {\rm X,Y} \in \{{\rm E,I}\}$, are deviations from the mean. Furthermore, note that since, by definition, $\overline{A}^{\rm EE} = \left(\frac{1}{n_{\rm E}^2}\right)\sum_{i=1}^{n_{\rm E}}\sum_{j=1}^{n_{\rm E}}A_{ij}^{\rm EE}$, $\overline{A}^{\rm EI} = \left(\frac{1}{n_{\rm E}n_{\rm I}}\right)\sum_{i=1}^{n_{\rm E}}\sum_{j=1}^{n_{\rm I}}A_{ij}^{\rm EI}$, $\overline{A}^{\rm IE} = \left(\frac{1}{n_{\rm E}n_{\rm I}}\right)\sum_{i=1}^{n_{\rm I}}\sum_{j=1}^{n_{\rm E}}A_{ij}^{\rm IE}$, and $\overline{A}^{\rm II} = \left(\frac{1}{n_{\rm I}^2}\right)\sum_{i=1}^{n_{\rm I}}\sum_{j=1}^{n_{\rm I}}A_{ij}^{\rm II}$, it follows that

$$\begin{aligned}\sum_{i=1}^{n_{\rm E}}\sum_{j=1}^{n_{\rm E}}\Delta_{ij}^{\rm EE} &= \sum_{i=1}^{n_{\rm E}}\sum_{j=1}^{n_{\rm I}}\Delta_{ij}^{\rm EI} = \sum_{i=1}^{n_{\rm I}}\sum_{j=1}^{n_{\rm E}}\Delta_{ij}^{\rm IE} \\ &= \sum_{i=1}^{n_{\rm I}}\sum_{j=1}^{n_{\rm I}}\Delta_{ij}^{\rm II} = 0. \quad (22)\end{aligned}$$

Now, using the average and perturbed expressions for

$A_{ij}^{\mathrm{XY}}, \mathrm{X}, \mathrm{Y} \in \{\mathrm{E}, \mathrm{I}\}$, (20) and (21) can be rewritten as

$$\begin{aligned}\frac{\mathrm{d}S_i^{\mathrm{E}}(t)}{\mathrm{d}t} &= f\left(n_{\mathrm{E}}\overline{A}^{\mathrm{EE}}\overline{S}^{\mathrm{E}}(t) + \sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{EE}}S_j^{\mathrm{E}}(t)\right.\\ &\quad \left. + n_{\mathrm{I}}\overline{A}^{\mathrm{EI}}\overline{S}^{\mathrm{I}}(t) + \sum_{k=1}^{n_{\mathrm{I}}}\Delta_{ik}^{\mathrm{EI}}S_k^{\mathrm{I}}(t) + I^{\mathrm{E}}\right)\\ &\quad -\frac{1}{\lambda^{\mathrm{E}}}S_i^{\mathrm{E}}(t), \quad i = 1, \ldots, n_{\mathrm{E}}, \qquad (23)\\ \frac{\mathrm{d}S_i^{\mathrm{I}}(t)}{\mathrm{d}t} &= f\left(n_{\mathrm{E}}\overline{A}^{\mathrm{IE}}\overline{S}^{\mathrm{E}}(t) + \sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{IE}}S_j^{\mathrm{E}}(t)\right.\\ &\quad \left. + n_{\mathrm{I}}\overline{A}^{\mathrm{II}}\overline{S}^{\mathrm{I}}(t) + \sum_{k=1}^{n_{\mathrm{I}}}\Delta_{ik}^{\mathrm{II}}S_k^{\mathrm{I}}(t) + I^{\mathrm{I}}\right)\\ &\quad -\frac{1}{\lambda^{\mathrm{I}}}S_i^{\mathrm{I}}(t), \quad i = 1, \ldots, n_{\mathrm{I}}, \qquad (24)\end{aligned}$$

where $\overline{S}^{\mathrm{E}}(t) \triangleq \frac{1}{n_{\mathrm{E}}}\sum_{j=1}^{n_{\mathrm{E}}}S_j^{\mathrm{E}}(t)$ and $\overline{S}^{\mathrm{I}}(t) \triangleq \frac{1}{n_{\mathrm{I}}}\sum_{j=1}^{n_{\mathrm{I}}}S_j^{\mathrm{I}}(t)$.

Next, assume that $i)$ $\Delta_{ij}^{\mathrm{XY}}, \mathrm{X}, \mathrm{Y} \in \{\mathrm{E}, \mathrm{I}\}$, $i = 1, \ldots, n_{\mathrm{X}}$ and $j = 1, \ldots, n_{\mathrm{Y}}$, in (23) and (24) are small relative to the remaining terms in $f(\cdot)$, and $ii)$ $\sum_{i=1}^{n_{\mathrm{X}}}\Delta_{ij}^{\mathrm{XY}} = 0, \mathrm{X}, \mathrm{Y} \in \{\mathrm{E}, \mathrm{I}\}$, for each $j \in \{1, \ldots, n_{\mathrm{Y}}\}$, which asserts that the total influence of the jth neuron on all other neurons is identical for each $j \in \{1, \ldots, n_{\mathrm{Y}}\}$. It follows from assumption $i)$ that the first-order expansions of (23) and (24) are given by

$$\begin{aligned}\frac{\mathrm{d}S_i^{\mathrm{E}}(t)}{\mathrm{d}t} &= f\left(n_{\mathrm{E}}\overline{A}^{\mathrm{EE}}\overline{S}^{\mathrm{E}}(t) + n_{\mathrm{I}}\overline{A}^{\mathrm{EI}}\overline{S}^{\mathrm{I}}(t) + I^{\mathrm{E}}\right)\\ &\quad + f'\left(n_{\mathrm{E}}\overline{A}^{\mathrm{EE}}\overline{S}^{\mathrm{E}}(t) + n_{\mathrm{I}}\overline{A}^{\mathrm{EI}}\overline{S}^{\mathrm{I}}(t) + I^{\mathrm{E}}\right)\\ &\quad \left[\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{EE}}S_j^{\mathrm{E}}(t) + \sum_{k=1}^{n_{\mathrm{I}}}\Delta_{ik}^{\mathrm{EI}}S_k^{\mathrm{I}}(t)\right]\\ &\quad -\frac{1}{\lambda^{\mathrm{E}}}S_i^{\mathrm{E}}(t), \quad i = 1, \ldots, n_{\mathrm{E}}, \qquad (25)\\ \frac{\mathrm{d}S_i^{\mathrm{I}}(t)}{\mathrm{d}t} &= f\left(n_{\mathrm{E}}\overline{A}^{\mathrm{IE}}\overline{S}^{\mathrm{E}}(t) + n_{\mathrm{I}}\overline{A}^{\mathrm{II}}\overline{S}^{\mathrm{I}}(t) + I^{\mathrm{I}}\right)\\ &\quad + f'\left(n_{\mathrm{E}}\overline{A}^{\mathrm{IE}}\overline{S}^{\mathrm{E}}(t) + n_{\mathrm{I}}\overline{A}^{\mathrm{II}}\overline{S}^{\mathrm{I}}(t) + I^{\mathrm{I}}\right)\\ &\quad \left[\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{IE}}S_j^{\mathrm{E}}(t) + \sum_{k=1}^{n_{\mathrm{I}}}\Delta_{ik}^{\mathrm{II}}S_k^{\mathrm{I}}(t)\right]\\ &\quad -\frac{1}{\lambda^{\mathrm{I}}}S_i^{\mathrm{I}}(t), \quad i = 1, \ldots, n_{\mathrm{I}}. \qquad (26)\end{aligned}$$

Now, letting $\overline{S}^{\mathrm{E}}$ and $\overline{S}^{\mathrm{I}}$ denote the mean excitatory synaptic drive and mean inhibitory synaptic drive, respectively, and defining $S_i^{\mathrm{E}}(t) \triangleq \overline{S}^{\mathrm{E}}(t) + \delta_i^{\mathrm{E}}(t)$ and $S_i^{\mathrm{I}}(t) \triangleq \overline{S}^{\mathrm{I}}(t) + \delta_i^{\mathrm{I}}(t)$, where $\delta_i^{\mathrm{E}}(t)$ and $\delta_i^{\mathrm{I}}(t)$ are deviations from the mean, it follows that

$$\begin{aligned}\sum_{i=1}^{n_{\mathrm{E}}}\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{EE}}S_j^{\mathrm{E}}(t) &= \sum_{i=1}^{n_{\mathrm{E}}}\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{EE}}\left(\overline{S}^{\mathrm{E}}(t) + \delta_i^{\mathrm{E}}(t)\right)\\ &= \sum_{i=1}^{n_{\mathrm{E}}}\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{EE}}\delta_i^{\mathrm{E}}(t), \qquad (27)\end{aligned}$$

since $\sum_{i=1}^{n_{\mathrm{E}}}\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{EE}}\overline{S}^{\mathrm{E}}(t) = 0$ by (22).

Next, it follows from assumption $ii)$ that

$$\sum_{i=1}^{n_{\mathrm{E}}}\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{EE}}\delta_j^{\mathrm{E}}(t) = 0, \quad \sum_{i=1}^{n_{\mathrm{E}}}\sum_{j=1}^{n_{\mathrm{I}}}\Delta_{ij}^{\mathrm{EI}}\delta_j^{\mathrm{I}}(t) = 0, \qquad (28)$$

$$\sum_{i=1}^{n_{\mathrm{I}}}\sum_{j=1}^{n_{\mathrm{E}}}\Delta_{ij}^{\mathrm{IE}}\delta_j^{\mathrm{E}}(t) = 0, \quad \sum_{i=1}^{n_{\mathrm{I}}}\sum_{j=1}^{n_{\mathrm{I}}}\Delta_{ij}^{\mathrm{II}}\delta_j^{\mathrm{I}}(t) = 0. \qquad (29)$$

Now, summing (25) and (26) over $i = 1, \ldots, n_{\mathrm{E}}$ and $i = 1, \ldots, n_{\mathrm{I}}$, and dividing by n_{E} and n_{I}, respectively, it follows that the average excitatory synaptic drive and the average inhibitory synaptic drive are given by

$$\begin{aligned}\frac{\mathrm{d}\overline{S}^{\mathrm{E}}(t)}{\mathrm{d}t} &= f\left(n_{\mathrm{E}}\overline{A}^{\mathrm{EE}}\overline{S}^{\mathrm{E}}(t) + n_{\mathrm{I}}\overline{A}^{\mathrm{EI}}\overline{S}^{\mathrm{I}}(t) + I^{\mathrm{E}}\right)\\ &\quad -\frac{1}{\lambda^{\mathrm{E}}}\overline{S}^{\mathrm{E}}(t), \quad t \geq 0, \qquad (30)\\ \frac{\mathrm{d}\overline{S}^{\mathrm{I}}(t)}{\mathrm{d}t} &= f\left(n_{\mathrm{E}}\overline{A}^{\mathrm{IE}}\overline{S}^{\mathrm{E}}(t) + n_{\mathrm{I}}\overline{A}^{\mathrm{II}}\overline{S}^{\mathrm{I}}(t) + I^{\mathrm{I}}\right)\\ &\quad -\frac{1}{\lambda^{\mathrm{I}}}\overline{S}^{\mathrm{I}}(t). \qquad (31)\end{aligned}$$

Next, note that (30) and (31) can be written in the form of (12) with

$$\begin{aligned}A &= \begin{bmatrix} n_{\mathrm{E}}\overline{A}^{\mathrm{EE}} & n_{\mathrm{I}}\overline{A}^{\mathrm{EI}} \\ n_{\mathrm{E}}\overline{A}^{\mathrm{IE}} & n_{\mathrm{I}}\overline{A}^{\mathrm{II}} \end{bmatrix}, \quad L = \begin{bmatrix} \frac{1}{\lambda^{\mathrm{E}}} & 0 \\ 0 & \frac{1}{\lambda^{\mathrm{I}}} \end{bmatrix},\\ A - L &= \begin{bmatrix} n_{\mathrm{E}}\overline{A}^{\mathrm{EE}} - \frac{1}{\lambda^{\mathrm{E}}} & n_{\mathrm{I}}\overline{A}^{\mathrm{EI}} \\ n_{\mathrm{E}}\overline{A}^{\mathrm{IE}} & n_{\mathrm{I}}\overline{A}^{\mathrm{II}} - \frac{1}{\lambda^{\mathrm{I}}} \end{bmatrix}. \qquad (32)\end{aligned}$$

If

$$\left(n_{\mathrm{E}}\overline{A}^{\mathrm{EE}} - \frac{1}{\lambda^{\mathrm{E}}}\right)\left(n_{\mathrm{I}}\overline{A}^{\mathrm{II}} - \frac{1}{\lambda^{\mathrm{I}}}\right) - n_{\mathrm{E}}\overline{A}^{\mathrm{IE}}n_{\mathrm{I}}\overline{A}^{\mathrm{EI}} = 0, \qquad (33)$$

then it follows that $\mathrm{rank}\,(A - L) < 2$. Hence, it follows from the analysis of Section VI that the dynamical system (30) and (31) exhibits multistability for λ^{I} and λ^{E} satisfying (33) and $n_{\mathrm{E}}\overline{A}^{\mathrm{EE}} + n_{\mathrm{I}}\overline{A}^{\mathrm{II}} < \frac{1}{\lambda^{\mathrm{E}}} + \frac{1}{\lambda^{\mathrm{I}}}$. Note that if λ^{I} and λ^{E} satisfy (33) and the previous inequality, then the eigenvalues of $A - L$ are given by 0 and $n_{\mathrm{E}}\overline{A}^{\mathrm{EE}} + n_{\mathrm{I}}\overline{A}^{\mathrm{II}} - \frac{1}{\lambda^{\mathrm{E}}} + \frac{1}{\lambda^{\mathrm{I}}} < 0$. Hence, $A - L$ is semistable [23].

To investigate (30) and (31) numerically, let $f(\cdot)$ be the given by (10) with $\gamma = 100$, $n_{\mathrm{E}}\overline{A}^{\mathrm{EE}} = 0.2$, $n_{\mathrm{I}}\overline{A}^{\mathrm{EI}} = 1$, $n_{\mathrm{E}}\overline{A}^{\mathrm{IE}} = 1$, $n_{\mathrm{I}}\overline{A}^{\mathrm{II}} = 0$, $\lambda^{\mathrm{E}} = 1$, $I^{\mathrm{E}} = 0$, and $I^{\mathrm{I}} = 0$, and let λ^{I} vary. In this case,

$$A = \begin{bmatrix} 0.2 & 1 \\ 1 & 0 \end{bmatrix}, L = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\lambda^{\mathrm{I}}} \end{bmatrix}, A - L = \begin{bmatrix} -0.8 & 1 \\ 1 & -\frac{1}{\lambda^{\mathrm{I}}} \end{bmatrix}.$$

Clearly, $\mathrm{rank}\,(A - L) < 2$ for $\lambda^{\mathrm{I}} = 0.8$. Hence, it follows from the analysis of Section VI that the dynamical system (30) and (31) exhibits multistability for $\lambda^{\mathrm{I}} = 0.8$. In this case, $A - L$ is semistable. For our simulation, we take $x_1 = \overline{S}^{\mathrm{E}}$ and $x_2 = \overline{S}^{\mathrm{I}}$, and use the initial condition $x(0) = [0.1, 0.5]^{\mathrm{T}}$. Figures 1 and 2 show the time response for the average excitatory and inhibitory synaptic drives, and the phase portrait for $\lambda^{\mathrm{I}} = 0.8$. Note that there is a zero-eigenvalue transcritical bifurcation at $\lambda^{\mathrm{I}} = 0.8$.

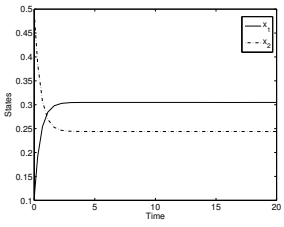

Fig. 1. State trajectories of (30) and (31) for $\lambda^{\mathrm{I}} = 0.8$.

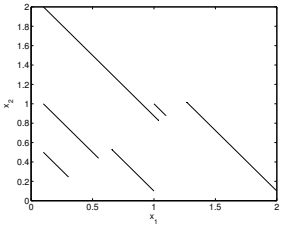

Fig. 2. Phase portrait of (30) and (31) for $\lambda^{\mathrm{I}} = 0.8$.

VIII. Conclusion

With advances in biochemistry, molecular biology, and neurochemistry there has been impressive progress in the understanding of the function of single neurons. Using the example of the mechanism of action of general anesthesia, the past decade has seen a remarkable explosion of our understanding of how anesthetic agents affect the properties of neurons. However, despite this advance, we still do not understand how molecular mechanisms translate into the induction of general anesthesia at the macroscopic level. In particular, there has been little focus on how the molecular properties of anesthetic agents lead to the observed macroscopic property that defines the anesthetic state, that is, lack of responsiveness to noxious stimuli. This clinical property leads to consideration of anesthesia as a binary (on-or-off) variable, and the relationship between the concentration of an anesthetic agent in the central nervous system and the anesthetic state is described in terms of the probability of responsiveness as a function of anesthetic concentration [8]. In clinical studies, the typical observation is that at low concentrations of anesthetic agent the probability of responsiveness (to noxious stimuli) is high, possibly unity. Then as the anesthetic concentration increases there is a sharp transition to a probability of responsiveness that is low and possibly zero.

In this paper, we used a synaptic drive firing rate model to model the central nervous system as a discontinuous autonomous dynamical system and showed that the transition to the anesthetic state exhibits multistability; that is, the system exhibits multiple attracting equilibria under asymptotically slowly changing parameters. The goal of this paper has been to specifically develop multistability theory as a framework for understanding central nervous system behavior characterized by abrupt transitions between mutually exclusive states. Such phenomena are not limited to general anesthesia and can be seen in biochemical systems, ecosystems, gene regulation and cell replication, as well as numerous medical conditions (e.g., seizures, schizophrenia, hallucinations, etc.) and are obviously of great clinical importance but have been lacking rigorous theoretical frameworks. The primary impact of such frameworks will be to allow for the development of models that go beyond words to dynamic equations, leading to mathematical models with greater precision and self-consistency. Mathematical formulations enforce self-consistency and while "self-consistency is not necessarily truth, self-inconsistency is certainly falsehood."

References

[1] L. Lapicque, "Recherches quantitatives sur l' excitation electiique des nerfs traitee comme une polarization," *J. Physiol. Gen.*, vol. 9, pp. 620–635, 1907.

[2] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, pp. 500–544, 1952.

[3] B. Ermentrout and D. H. Terman, *Mathematical Foundations of Neuroscience*. New York: Springer-Verlag, 2010.

[4] G. A. Mashour, "Consciousness unbound: Toward a paradigm of general anesthesia," *Anesthesiology*, vol. 100, pp. 428–433, 2004.

[5] A. Y. Zecharia and N. P. Franks, "General anesthesia and ascending arousal pathways," *Anesthesiology*, vol. 111, pp. 695–696, 2009.

[6] J. A. Campagna, K. W. Miller, and S. A. Forman, "Mechanisms of actions of inhaled anesthetics," *N. Engl. J. Med.*, vol. 348, pp. 2110–2124, 2003.

[7] E. R. John and L. S. Prichep, "The anesthetic cascade: A theory of how anesthesia suppresses consciousness," *Anesthesiology*, vol. 102, pp. 447–471, 2005.

[8] J. M. Bailey and W. M. Haddad, "Drug dosing control in clinical pharmacology: Paradigms, benefits, and challenges," *Contr. Syst. Mag.*, vol. 25, pp. 35–51, 2005.

[9] G. Merkel and E. I. Eger, "A comparative study of halothane and halopropane anesthesia including method for determining equipotency," *Anesthesiology*, vol. 24, pp. 346–357, 1963.

[10] J. Vuyk, T. Lim, F. H. M. Engbers, A. G. L. Burm, A. A. Vietter, and J. G. Bovill, "Pharmacodynamics of alfentanil as a supplement to propofol of nitrous oxide for lower abdominal surgery in female patients," *Anesthesiology*, vol. 78, pp. 1936–1945, 1993.

[11] S. Hameroff, "The entwined mysteries of anesthesia and consciousness. Is there a common underlying mechanism?" *Anesthesiology*, vol. 105, pp. 400–412, 2006.

[12] E. Overton, *Studien über die Narkose zugleich ein Beitrag zur allgemeinen Pharmakologie*. Jena: Verlag con Gustav Fischer, 1901.

[13] H. Meyer, "Welche eigenschaft der ansthetica bedingt ihre narkotische wirkung?" *Naunyn-Schmiedebergs Arch. Exp. Path. Pharmakol.*, vol. 42, pp. 109–118, 1889.

[14] I. Ueda, "Molecular mechanisms of anesthesia," *Anesth. Analg.*, vol. 63, pp. 929–945, 2003.

[15] N. P. Franks and W. R. Lieb, "Molecular and cellular mechanisms of general anesthesia," *Nature*, vol. 367, pp. 607–614, 1994.

[16] L. J. Voss, J. W. Sleigh, J. P. M. Barnard, and H. E. Kirsch, "The howling cortex: Seizurs and general anesthetic drugs," *Anesth. Analg.*, vol. 107, pp. 1689–1703, 2008.

[17] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press, 2005.

[18] A. Kiyamura, W. Marszalec, J. Z. Yeh, and T. Narahishi, "Effects of halothane and propofol on excitatory and inhibitory synaptic transmission in rat cortical neurons," *J. Pharmacol.*, vol. 304, pp. 162–171, 2002.

[19] A. Hutt and A. Longtin, "Effects of the anesthetic agent propofol on neural populations," *Cogn. Neurodyn.*, vol. 4, pp. 37–59, 2009.

[20] M. L. Steyn-Ross, D. A. Steyn-Ross, and J. W. Sleigh, "Modelling general anesthesia as a first-order phase transition in the cortex," *Prog. Biophysics Mol. Biol.*, vol. 85, pp. 369–385, 2004.

[21] D. Angeli, J. Ferrell, and E. D. Sontag, "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems," *Proc. Natl. Acad. Sci.*, vol. 101, pp. 1822–1827, 2004.

[22] S. P. Bhat and D. S. Bernstein, "Nontangency-based Lyapunov tests for convergence and stability in systems having a continuum of equilibra," *SIAM J. Control Optim.*, vol. 42, pp. 1745–1775, 2003.

[23] W. M. Haddad, V. Chellaboina, and Q. Hui, *Nonnegative and Compartmental Dynamical Systems*. Princeton, NJ: Princeton Univ. Press, 2010.

[24] Q. Hui, W. M. Haddad, and J. M. Bailey, "Multistability, bifurcations, and biological neural networks: A synaptic drive firing model for cerebral cortex transition in the induction of general anesthesia," *Nonlinear Analysis: Hybrid Systems*, submitted.

[25] U. Vaidya and P. G. Mehta, "Lyapunov measure for almost everywhere stability," *IEEE Trans. Autom. Control*, vol. 53, pp. 307–323, 2008.

[26] A. F. Filippov, *Differential Equations with Discontinuous Right-Hand Sides*. Dordrecht, The Netherlands: Kluwer, 1988.

[27] Q. Hui, W. M. Haddad, and S. P. Bhat, "Semistability, finite-time stability, differential inclusions, and discontinuous dynamical systems having a continuum of equilibria," *IEEE Trans. Autom. Control*, vol. 54, pp. 2465–2470, 2009.

[28] D. Angeli, "Multistability in systems with counter-clockwise input-output dynamics," *IEEE Trans. Autom. Control*, vol. 52, pp. 596–609, 2007.