

Parameter Space Structure of Continuous-Time Recurrent Neural Networks

Randall D. Beer
Dept. of Electrical Engineering and Computer Science
Dept. of Biology
Case Western Reserve University
Cleveland, OH 44106

Submitted – 10/5/05
Revised – 3/27/06

Abstract

A fundamental challenge for any general theory of neural circuits is how to characterize the structure of the space of all possible circuits over a given model neuron. As a first step in this direction, this paper begins a systematic study of the global parameter space structure of continuous-time recurrent neural networks (CTRNNs), a class of neural models which, though simple, is dynamically universal. First, we explicitly compute the local bifurcation manifolds of CTRNNs. We then visualize the structure of these manifolds in net input space for small circuits. These visualizations reveal a set of extremal saddle-node bifurcation manifolds that divide CTRNN parameter space into regions of dynamics with different effective dimensionality. Next, we completely characterize the combinatorics and geometry of an asymptotically exact approximation to these regions for circuits of arbitrary size. Finally, we show how these regions can be used to calculate estimates of the probability of encountering different kinds of dynamics in CTRNN parameter space.

Please address all correspondence to:

Randall D. Beer
Dept. of Electrical Engineering and Computer Science
Case Western Reserve University
Cleveland, OH 44106-7071

Phone: (216) 368-2816
Fax: (216) 368-1534
Email: beer@eeecs.cwru.edu
URL: <http://vorlon.cwru.edu/~beer>

1. Introduction

Although tremendous progress has been made on the computational modeling of specific neural circuits, the need for a more general theory of such circuits is becoming widely recognized in neuroscience. This is especially true in work on invertebrate pattern generation, where detailed experimental analyses of several pattern generation circuits have brought questions of general principles to the fore (Selverston, 1980; Getting, 1989; Marder & Abbott, 1995; Marder & Calabrese, 1996). For example, recent computational and experimental studies of neurons and small circuits from the crustacean stomatogastric ganglion (STG) have become increasingly concerned with the overall parameter space structure of this circuit. In studies of both model and biological STG neurons, Goldman et al. (2001) found that activity patterns were quite robust to some sets of conductance variations, but extremely sensitive to other sets. Golowasch et al. (2002) found that averaging the conductance values of different models of an STG cell exhibiting a given firing characteristic could sometimes fail to produce an average model with the same characteristic due to nonlinearities in conductance space. Prinz et al. (2003) found that widely disparate parameter sets in a 3-cell pyloric network model could give rise to almost indistinguishable network activity. More recently, Prinz et al. (2004) have constructed maps of the parameter spaces of model STG neurons through sampling on a coarse grid of conductance values.

A general theory of neural circuits would benefit such studies in many ways. It would tell us what kinds of dynamical behavior lie in different regions of a circuit's parameter space and allow us to quantify the probability of encountering these different dynamics. It would help us to understand a circuit's response to sensory inputs or neuromodulators by distinguishing those directions of variation in parameter space to which its behavior is most sensitive from those to which it is robust. It would support the classification of novel circuits from their parameters alone, without requiring an exhaustive dynamical analysis of each new circuit. It would provide mathematical and computational tools for calculating experimentally testable predictions

regarding novel manipulations. It would provide a methodology for the design of neural circuits with desired behavior. Perhaps most importantly, it would supply a general context within which to understand the behavior of particular circuits, allowing us to situate the details of the actual within the space of the possible.

Characterizing the general parameter space structure of neural circuits is an extremely difficult problem and is likely to be impossible in full detail for most models of interest. Nevertheless, even a partial characterization of the parameter space structure of the simplest nontrivial model neural circuits would be extremely useful. Studies of simpler models can help us to build intuition about what such a general theory might look like. Such studies would also allow us to develop the conceptual framework necessary to formulate the theory and the mathematical and computational machinery required to derive its consequences. In addition, they would give us a better appreciation for the kinds of questions that we can reasonably expect to answer with such a theory, and the kinds of questions that are likely to be beyond its reach. Only armed with this experience can we hope to extend our understanding to the parameter space structure of more biophysically-realistic models.

In this paper, we begin the task of characterizing the parameter space structure of a class of model neural circuits which, though simple, is dynamically universal. Section 2 introduces the continuous-time recurrent neural network (CTRNN) model we will study and reviews some of its key properties. In Section 3, we explicitly compute the local bifurcation manifolds of CTRNNs and visualize their detailed structure for small N . An approximation to the overall structure of the outer envelope of these manifolds for arbitrary N is then derived analytically in Section 4. Sections 5 and 6 use the theory to calculate estimates of the probability of encountering different kinds of dynamics in CTRNN parameter space. The paper concludes with a discussion of the present status of the theory and directions for future work. An electronic supplement in the form of a *Mathematica* (Wolfram, 2003) notebook provides tools for the reader to reproduce the main calculations in the paper and to carry out their own explorations (Beer, 2005).

2. Continuous-Time Recurrent Neural Networks

Continuous-time recurrent neural networks are among the simplest possible nonlinear continuous-time neural models. CTRNNs are defined by the vector differential equation

$$\tau \dot{\mathbf{y}} = -\mathbf{y} + \mathbf{W}\sigma(\mathbf{y} + \boldsymbol{\theta}) + \mathbf{I} \quad (2.1)$$

where τ , $\dot{\mathbf{y}}$, \mathbf{y} , $\boldsymbol{\theta}$, and \mathbf{I} are length N vectors, $\mathbf{W} = \{w_{ij}\}$ is an $N \times N$ matrix, and all vector operations (including the application of the output function $\sigma(x) = 1 / (1 + e^{-x})$) are performed element-wise.

The standard neurobiological interpretation of this model is that y_i represents the mean membrane potential of the i^{th} neuron, $\sigma(\cdot)$ represents its mean firing rate, τ_i represents its membrane time constant, θ_i represents its threshold/bias, I_i represents an external input, the weights $w_{ij, j \neq i}$ represent synaptic connections from neuron j to neuron i , and the self-interaction w_{ii} represents a simple active conductance. This model can also be interpreted as representing nonspiking neurons (Dunn, et al., 2004). In this case, $\sigma(\cdot)$ represents saturating nonlinearities in synaptic input. Note that the distinction between I and θ is merely semantic; with respect to the output dynamics of (2.1), only the net input $I + \theta$ matters, since Eqn. (2.1) can be rewritten in the form $\tau \dot{\mathbf{x}} = -\mathbf{x} + \sigma(\mathbf{W}\mathbf{x} + \mathbf{I} + \boldsymbol{\theta})$ using the substitution $\mathbf{y} \mapsto \mathbf{W}\mathbf{x} + \mathbf{I}$. Without loss of generality, we will often assume that $\mathbf{I} = \mathbf{0}$, so the net input to a CTRNN is given simply by $\boldsymbol{\theta}$. Thus, an N -neuron CTRNN has N time constants, N net inputs and N^2 weights, giving $\mathcal{C}_{\text{CTRNN}}(N)$, the space of all possible CTRNNs on N neurons, an $(N^2 + 2N)$ -dimensional parameter space.

Compared to more biologically-realistic neural models, the dynamics of an individual CTRNN neuron is quite trivial. However, small networks of CTRNNs can reproduce qualitatively the full range of nerve cell phenomenology, including spiking, plateau potentials, bursting, etc. More importantly, CTRNNs are known to be universal approximators of smooth dynamics (Funahashi & Nakamura, 1993; Kimura & Nakano, 1998; Chow & Li, 2000). Thus, at least in principle, the use of CTRNNs implies no restriction whatsoever on biological realism.

CTRNNs can be thought of as a basis dynamics from which any other dynamics can be reproduced to any desired degree of accuracy.

CTRNNs are a special case of the general class of additive neural network models $\tau \dot{\mathbf{y}} = -\mathbf{y} + \mathbf{W}\xi(\mathbf{y} + \boldsymbol{\theta}) + \mathbf{I}$ (Grossberg, 1988). Additive neural networks have been extensively studied in both their continuous-time and discrete-time versions (Cowan & Ermentrout, 1978; Cohen & Grossberg, 1983; Hopfield, 1984; Hirsch, 1989; Borisuk & Kirillov, 1992; Blum & Wang, 1992; Zhaojue, Schieve, & Das, 1993; Beer, 1995; Hoppensteadt & Izhikevich, 1997; Tiňo, Horn & Giles, 2001; Haschke & Steil, 2005). Although our analysis will focus on (2.1), the results we obtain would be qualitatively identical for any additive model with a (smooth, monotone, bounded) sigmoidal activation function $\xi(x)$.

A particularly convenient class of activation functions can be parameterized as

$$\sigma_{\alpha,\beta,\mu}(x) = \frac{\alpha}{1 + e^{-\mu x}} + \beta$$

where $\alpha, \mu \in \mathbb{R}_+$, $\beta \in \mathbb{R}$ (Tiňo, Horne & Giles, 2001). This class contains several common activation functions, including both the one used here ($\sigma_{1,0,1}$) and the hyperbolic tangent function ($\sigma_{2,-1,2}$). All additive neural network models using activation functions in this class share an important property: their dynamics are topologically conjugate (Haschke, 2004). Specifically, the quantitative results we obtain for (2.1) can be directly translated to any other additive model $\tau' \dot{\mathbf{y}} = -\mathbf{y}' + \mathbf{W}' \sigma_{\alpha,\beta,\mu}(\mathbf{y}' + \boldsymbol{\theta}') + \mathbf{I}'$ via the change of variables

$$\begin{aligned} \mathbf{y}' &= \mu^{-1} \mathbf{y} \\ \tau' &= \tau \\ \mathbf{W}' &= (\alpha \mu)^{-1} \mathbf{W} \\ \boldsymbol{\theta}' &= \mu^{-1} \boldsymbol{\theta} \\ \mathbf{I}' &= \mu^{-1} \mathbf{I} - (\alpha \mu)^{-1} \mathbf{W} \cdot \boldsymbol{\beta} \end{aligned}$$

where $\boldsymbol{\beta}$ is the length N vector (β, \dots, β) .

The steady-state input/output (SSIO) curve of a single CTRNN neuron will play an important role in this paper (Figure 1). By transforming (2.1) to the output space defined by $o \equiv \sigma(y + \theta)$ and setting the time derivative to 0, we find that the SSIO curve of a neuron with self-weight w is

given by $I + \theta = \sigma^{-1}(o) - wo$. A single additive model neuron can exhibit either unstable or bistable dynamics, depending on the strength of its self-weight and its net input (Cowan & Ermentrout, 1978). In a single CTRNN neuron, only unstable dynamics are possible when $w < 4$ (Figure 1A). When $w > 4$, bistable dynamics occurs when $I_L(w) \leq I + \theta \leq I_R(w)$ (Figure 1B), where the left and right edges of the fold are given by (Beer, 1995)

$$I_L(w), I_R(w) = \pm 2 \ln \left(\frac{\sqrt{w} + \sqrt{w-4}}{2} \right) - \frac{w \pm \sqrt{w(w-4)}}{2}$$

and it is convenient to define the width of the fold as $I_w(w) \equiv I_R(w) - I_L(w)$. Analogous expressions can be derived for arbitrary sigmoidal activation functions.

[Insert Figure 1 Here]

As we study the parameter space of CTRNNs, we will repeatedly encounter *center-crossing circuits* defined by the condition $\theta_i^* = -\sum_{j=1}^N w_{ij}/2$ (Beer, 1995). When this condition is satisfied, the null manifolds of each neuron intersect at their centers of symmetry, or, equivalently, the SSIO of each neuron is centered over the range of synaptic inputs that it receives from the other neurons. Center-crossing circuits are important for a variety of reasons. First, the richest possible dynamics can be found in the neighborhood of such circuits. Second, the bifurcations of the central equilibrium point of a center-crossing circuit can often be fully characterized analytically. Finally, for any given weight matrix, the corresponding center-crossing circuit serves as a symmetry point in the net input parameter space for that circuit.

3. Visualizing Local Bifurcation Manifolds

Before we attempt to characterize the general structure of $\mathcal{C}_{\text{CTRNN}}(N)$, it would be helpful to directly visualize this structure, at least for small N . As is typical in dynamical systems theory, we will abstract over the details of individual trajectories and study instead the equivalence classes induced by topological conjugacy of entire flows (Kuznetsov, 1998). Under this equivalence relation, the parameter space of a dynamical system is divided into regions of

topologically-equivalent dynamics by bifurcation manifolds. If we wish to understand $\mathcal{C}_{\text{CTRNN}}(N)$, then characterizing the structure of these manifolds is a good place to begin.

Bifurcations can be either local or global. Local bifurcations involve changes in the neighborhood of a limit set and can be explicitly defined by algebraic conditions on the vectorfield and its derivatives in that neighborhood. For example, the change of stability of an equilibrium point as parameters are varied is a local bifurcation. In contrast, global bifurcations involve changes that are not localized to any particular limit set and can usually only be studied numerically. For example, saddle connections are global bifurcations in which the unstable manifold of one equilibrium point coincides with the stable manifold of another. We will only consider local bifurcations in this paper.

The two most common local bifurcations are the saddle-node and Hopf bifurcations (Kuznetsov, 1998). In a saddle-node bifurcation, the real eigenvalues of an equilibrium point change sign, signaling a change in its stability. In a Hopf bifurcation, the real parts of a complex conjugate pair of eigenvalues of an equilibrium point change sign, signaling a change in its stability and the production of a limit cycle. These bifurcations are defined by the conditions

$$\det(\mathbf{J}) = 0 \quad (\text{Saddle-Node Bifurcation})$$

$$\det(2\mathbf{J} \odot \mathbf{1}) = 0 \quad (\text{Hopf Bifurcation})$$

where $\mathbf{J} = \{\partial f_i / \partial y_j\}$ is the Jacobian matrix of partial derivatives of the vectorfield \mathbf{f} with respect to \mathbf{y} ($\mathbf{f}(\mathbf{y}) \equiv (-\mathbf{y} + \mathbf{W} \cdot \sigma(\mathbf{y} + \boldsymbol{\Theta})) / \tau$ for CTRNNs) evaluated at an equilibrium point $\bar{\mathbf{y}}$, $\mathbf{1}$ denotes the identity matrix, and \odot is the bialternate matrix product (Guckenheimer, Myers & Sturmfels, 1997). Given two $N \times N$ matrices $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$, $\mathbf{A} \odot \mathbf{B}$ is the $\frac{1}{2}N(N-1) \times \frac{1}{2}N(N-1)$ matrix whose rows are labeled by the multi-index (p, q) (where $p = 2, 3, \dots, N$ and $q = 1, 2, \dots, p-1$), whose columns are labeled by the multi-index (r, s) (where $r = 2, 3, \dots, N$ and $s = 1, 2, \dots, r-1$), and whose elements are given by (Kuznetsov, 1998)

$$(\mathbf{A} \odot \mathbf{B})_{(p,q)(r,s)} = \frac{1}{2} \left\{ \begin{vmatrix} a_{pr} & a_{ps} \\ b_{qr} & b_{qs} \end{vmatrix} + \begin{vmatrix} b_{pr} & b_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} \right\}$$

An implementation of \odot is included in the electronic supplement.

The significance of the bialternate matrix product lies in the fact that if an $N \times N$ matrix \mathbf{M} has eigenvalues $\lambda_1, \dots, \lambda_N$, then $2\mathbf{M} \odot \mathbf{1}$ has eigenvalues $\lambda_i + \lambda_j$, with $1 \leq j < i \leq N$. Thus, the Hopf condition $\det(2\mathbf{J} \odot \mathbf{1}) = 0$ is satisfied whenever a complex conjugate pair of eigenvalues have zero real parts. Note that an equilibrium point having a pair of real eigenvalues with equal magnitude but opposite sign will also satisfy this condition (such a point is called a neutral saddle). Since such points are not actually Hopf bifurcations, the portions of a solution manifold of $\det(2\mathbf{J} \odot \mathbf{1}) = 0$ for which the eigenvalues are not pure imaginary must be removed in a post-processing step.

Haschke (2004; Haschke & Steil, 2005) developed a method for computing local bifurcation manifolds of discrete-time recurrent neural networks in net input space that can easily be adapted to CTRNNs. If we define $\psi \equiv \sigma'(\bar{\mathbf{y}} + \theta)$ to be the vector of derivatives of $\sigma(\cdot)$ at an equilibrium point $\bar{\mathbf{y}}$, with $0 < \psi_i \leq 1/4$, then the CTRNN Jacobian at that point can be written as

$$\mathbf{J} = \text{diag}(\boldsymbol{\tau}^{-1}) \cdot (\text{diag}(\psi) \cdot \mathbf{W} - \mathbf{1})$$

where $\boldsymbol{\tau}^{-1}$ denotes the vector of reciprocals of the time constants $\boldsymbol{\tau}$ and $\text{diag}(\mathbf{v})$ denotes the diagonal matrix containing the vector \mathbf{v} . This allows the saddle-node and Hopf bifurcation conditions to be expressed directly in terms of ψ .

For example, for a 2-neuron CTRNN, the saddle-node bifurcation manifolds are given by the solutions (ψ_1, ψ_2) to

$$\det(\mathbf{J}) = \det \begin{pmatrix} \frac{w_{11}\psi_1 - 1}{\tau_1} & \frac{w_{21}\psi_1}{\tau_1} \\ \frac{w_{12}\psi_2}{\tau_2} & \frac{w_{22}\psi_2 - 1}{\tau_2} \end{pmatrix} = 1 - w_{11}\psi_1 - w_{22}\psi_2 + \psi_1\psi_2 \det \mathbf{W} = 0$$

where the time constants have been eliminated by multiplying through by $\tau_1\tau_2$. The Hopf bifurcation manifolds are given by the solutions to

$$\det(2\mathbf{J} \odot \mathbf{1}) = \det \left(2 \begin{pmatrix} \frac{w_{11}\psi_1 - 1}{\tau_1} & \frac{w_{21}\psi_1}{\tau_1} \\ \frac{w_{12}\psi_2}{\tau_2} & \frac{w_{22}\psi_2 - 1}{\tau_2} \end{pmatrix} \odot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = \frac{w_{11}\psi_1 - 1}{\tau_1} + \frac{w_{22}\psi_2 - 1}{\tau_2} = 0$$

We can visualize these local bifurcation manifolds either by numerically approximating the solution manifolds defined implicitly by each bifurcation condition or by explicitly solving for one ψ_i in terms of the others. The results for two sample 2-neuron circuits are shown in Figures 2A and 2B. Note that these manifolds are functions of \mathbf{W} (for saddle-node bifurcations) or \mathbf{W} and $\boldsymbol{\tau}$ (for Hopf bifurcations).

[Insert Figure 2 Here]

The bifurcation manifolds in $\boldsymbol{\psi}$ must then be transformed to net input space. Using the fact that $\bar{\mathbf{y}} = \sigma'^{-1}(\boldsymbol{\psi}) - \boldsymbol{\theta}$ (where $\sigma'^{-1}(\cdot)$ denotes the inverse of the derivative of $\sigma(\cdot)$) and then substituting this for $\bar{\mathbf{y}}$ in the equilibrium point condition $\mathbf{f}(\bar{\mathbf{y}}) = 0$, we obtain

$$\theta_i = \sigma'^{-1}(\psi_i) - \mathbf{W} \cdot \sigma(\sigma'^{-1}(\psi_i)) \quad (3.1)$$

where

$$\sigma'^{-1}(\psi_i) = \ln \frac{1 \pm \sqrt{1 - 4\psi_i}}{2\psi_i}$$

Note that $\sigma'^{-1}(\cdot)$ is 2-valued for a sigmoidal function. Since each component of $\boldsymbol{\theta}$ can come from either branch, each bifurcation manifold in $\boldsymbol{\psi}$ space can therefore generate up to 2^N bifurcation manifolds in net input space. Figures 2C and 2D show the $\boldsymbol{\theta}$ -space bifurcation manifolds corresponding to the $\boldsymbol{\psi}$ -space manifolds shown in Figures 2A and 2B, respectively.

Using the same approach, we can also calculate and display the local bifurcation manifolds of 3-neuron circuits. Figures 3A and 3B provide an external view of the local bifurcation manifolds for two different 3-neuron circuits and the slices in Figures 3C and 3D reveal some of the rich internal structure. In principle, this method can be applied to circuits of any size. However, the exponential scaling of the number of bifurcation manifolds in net input space and the difficulty of visualizing manifolds in dimensions greater than three make it practical for small

circuits only. *Mathematica* code for the visualization of the local bifurcation manifolds of CTRNNs in 2 and 3 dimensions can be found in the electronic supplement.

[Insert Figure 3 Here]

4. The Global Structure of Local Bifurcation Manifolds

How can we characterize the structure of CTRNN parameter space? As Figures 2 and 3 demonstrate, the local bifurcation structure of CTRNNs can be quite complex even in small circuits. Of course, including numerically-computed global bifurcation manifolds would serve only to further complicate these plots (Borisuk & Kirillov, 1992; Hoppensteadt & Izhikevich, 1997). If we have any hope of achieving an understanding of $\mathcal{C}_{\text{CTRNN}}(N)$ that can be scaled to large N , then we must focus on the overall structure of these manifolds rather than their fine details.

The plots in Figures 2 and 3 illustrate two key features of the overall structure of CTRNN parameter space. First, there is always a compact central region in net input space (whose location and extent depend on the connection weights) with the richest dynamics and the highest density of bifurcations. As it turns out, the center of each of these regions corresponds to the center-crossing network for that weight matrix. In these central regions, all N neurons are *dynamically active*: the range of net input each neuron receives from the other neurons overlaps the most sensitive region of its activation function $\sigma(\cdot)$. Dynamically active neurons can respond to changes in the outputs of other neurons and are therefore capable of participating in nontrivial dynamics.

A second key feature of CTRNN parameter space apparent in Figures 2 and 3 is that the local bifurcation manifolds flatten out as we move away from the central region, forming quasirectangular regions with an apparent combinatorial structure. This structure is produced by different subsets of neurons becoming *saturated*: the range of net input received is such that $\sigma(\cdot) \approx 0$ or $\sigma(\cdot) \approx 1$. Saturated neurons effectively drop out of the dynamics and become constant inputs to other neurons because their outputs are insensitive to changes in the outputs of

those other neurons (Figure 4). For example, in Figure 3A, we see a central “cube” surrounded by six “poles”, which are in turn interconnected by twelve “slabs”. In the central cube, all three neurons are dynamically active. In each pole, only two neurons are dynamically active; one of the neurons is saturated either on or off. Thus, in these regions, the dynamics of the 3-neuron circuit effectively becomes 2-dimensional. Indeed, the structure of the local bifurcation manifolds visible in the pole cross-sections in Figure 3A is very similar to those of the 2-neuron circuit in Figure 2C (Haschke, 2004). In the slabs, only one neuron is dynamically active, while in the eight remaining corner regions of this plot, all three neurons are saturated.

[Insert Figure 4 Here]

We can use this observation to partition the net input space of a CTRNN into regions of dynamics with different effective dimension depending on the number of neurons that are dynamically active. A CTRNN with an *effective dimension* of M has limit sets whose extent, distribution and responses to perturbations span an M -dimensional subspace of the N -dimensional output space of an N -neuron circuit, with $0 \leq M \leq N$ (Figure 4). In this section, we will completely characterize the combinatorics and geometry of an approximation to these regions for arbitrary N . Due to some differences in the details, we will first consider the case when all $w_{ii} > 4$ and then the case when some $w_{ii} < 4$.

4.1 All $w_{ii} > 4$

When all $w_{ii} > 4$, the SSIO curves of all neurons are folded (Figure 1B), and the edges of these folds play an important role in structuring CTRNN parameter space (Figure 4). Specifically, when the right edge of a neuron’s synaptic input range falls below its left fold, that neuron will be saturated off regardless of the states of the other neurons in the circuit (left rectangle in Figure 1B). Likewise, when the left edge of a neuron’s synaptic input range falls above its right fold, that neuron will be saturated on (right rectangle in Figure 1B). At the boundaries of these saturated regions (when a fold is tangent to the corresponding edge of the synaptic input range), saddle-node bifurcations occur. These saddle-node bifurcations are

extremal in the sense that, as long as a neuron remains saturated in this way, it cannot participate in any further bifurcations. Thus, all bifurcations (both local and global) that a given subset of neurons can undergo must be contained within the extremal bifurcation manifolds that delineate the dynamically active ranges of those neurons. When all $w_{ii} > 4$, the extremal bifurcation manifolds are the boundaries that divide CTRNN parameter space into regions of dynamics with different effective dimension.

Although the extremal bifurcation manifolds can be calculated analytically using the method described in the previous section, the expressions become extremely unwieldy for larger N . Thus, it is far more convenient for arbitrary N to work with approximations to these manifolds. Our approximation will be based on the fact that, as we move away from the center-crossing point in net input space, these extremal bifurcation manifolds become asymptotically flat due to the saturation of $\sigma(\cdot)$. By extrapolating inward toward the center-crossing point, we can approximate the extremal bifurcation manifolds by the hyperplanes that they asymptotically approach. Figures 5A-D show the region approximations that correspond to Figures 2C, 2D, 3A and 3B, respectively. Note that, while these approximations are asymptotically exact and generally quite good, there can be nontrivial differences when nonlinear effects come into significant play in the neighborhood of the center-crossing point (Figures 5B and 5D).

In the remainder of this section, we will formally characterize the structure of the regions bounded by these hyperplanes. We will proceed in three steps. First, we describe the region of M -dimensional dynamics in an N -neuron circuit (denoted by the symbol \mathcal{R}) as a union of disjoint polytopes (denoted by the symbol \mathcal{Q}) corresponding to different subsets of $N - M$ neurons in saturation. Second, we decompose the structure of each \mathcal{Q} into a union of overlapping rectangular hypersolids (denoted by the symbol \mathcal{B}). Third, we write the boundaries of each \mathcal{B} as a set of N inequalities involving functions of the connection weights.

[Insert Figure 5 Here]

Let $\mathcal{R}_M^N(\mathbf{W})$ be the asymptotic hyperplane approximation to the region of M -dimensional dynamics in the net input space of an N -neuron circuit with weight matrix \mathbf{W} , $0 \leq M \leq N$.

Because the saddle-node bifurcation condition is independent of the time constants, \mathcal{R}_M^N does not depend on τ . Thus, \mathcal{R}_M^N is an N -dimensional volume characterized by N^2 weight parameters. For example, Figure 6 shows the structure of \mathcal{R}_M^3 for $M = 0, \dots, 3$ for the same circuit shown in Figures 3A and 5C.

[Insert Figure 6 Here]

As can be clearly seen in Figure 6, $\mathcal{R}_M^N(\mathbf{W})$ is composed of disjoint polytopes corresponding to different subsets of $N - M$ neurons in saturation. Let ${}^N\mathcal{Q}_{i_1\dots i_k}^{i_{k+1}\dots i_{N-M}}(\mathbf{W})$ be the polytope with neurons $i_1\dots i_k$ saturated off and neurons $i_{k+1}\dots i_{N-M}$ saturated on for any $0 \leq k \leq N - M$. Let $\mathcal{Z}(\mathcal{S}) \equiv \bigtimes_{i \in \mathcal{S}} \{i, \overset{i}{\circ}\}$ be the set containing the indices in \mathcal{S} in all possible raised and lowered combinations, e.g., $\mathcal{Z}(\{1, 2\}) = \{\{1, \overset{1}{\circ}\}, \{2, \overset{2}{\circ}\}, \{1, 2\}, \{1, \overset{1}{\circ}, \overset{2}{\circ}\}\}$. Then

$$\mathcal{R}_M^N(\mathbf{W}) = \bigcup_{\substack{\mathcal{S} \in \mathcal{K}_{N-M}^N \\ \mathcal{J} \in \mathcal{Z}(\mathcal{S})}} {}^N\mathcal{Q}|\mathcal{J}(\mathbf{W}) \quad (4.1)$$

where \mathcal{K}_K^N denotes the K -subsets of $\{1, \dots, N\}$ and ${}^N\mathcal{Q}|\mathcal{J}$ denotes the concatenation of the indices \mathcal{J} onto ${}^N\mathcal{Q}$. For example, $\mathcal{R}_0^2 = {}^2\mathcal{Q}_{12} \cup {}^2\mathcal{Q}^{12} \cup {}^2\mathcal{Q}_1^2 \cup {}^2\mathcal{Q}_2^1$. Note that the index set \mathcal{J} is *unordered*: ${}^2\mathcal{Q}_{12} = {}^2\mathcal{Q}_{21}$. Since there are $\binom{N}{N-M} = \binom{N}{M}$ different ways of selecting $N - M$ neurons to saturate and each of these may be saturated either off or on, \mathcal{R}_M^N is composed of $\binom{N}{M} 2^{N-M}$ nonoverlapping \mathcal{Q} polytopes.

Each ${}^N\mathcal{Q}|\mathcal{J}$ polytope is in general nonconvex. In order to simplify the description of \mathcal{R}_M^N , we will further decompose these ${}^N\mathcal{Q}|\mathcal{J}$ s into unions of rectangular hypersolids ${}^N\mathcal{B}|\mathcal{L}(\mathbf{W})$, where \mathcal{L} is an *ordered* list of raised and lowered indices: ${}^3\mathcal{B}_2^1 \neq {}^3\mathcal{B}_1^2$. Then

$${}^N\mathcal{Q}|\mathcal{J}(\mathbf{W}) = \bigcup_{\mathcal{L} \in \mathcal{P}(\mathcal{J})} {}^N\mathcal{B}|\mathcal{L}(\mathbf{W}) \quad (4.2)$$

where $\mathcal{P}(\mathcal{J})$ is the set of permutations of the raised and lowered indices \mathcal{J} . For example, ${}^3\mathcal{Q}^{12} = {}^3\mathcal{B}^{12} \cup {}^3\mathcal{B}^{21}$ (Figure 7). Here ${}^3\mathcal{B}^{12}$ can be interpreted to mean that neuron 1 is on and neuron 2 is on given that neuron 1 is on, and ${}^3\mathcal{B}^{21}$ means that neuron 2 is on and neuron 1 is on

given that neuron 2 is on. Since \mathcal{L} contains $N - M$ elements, each ${}^N\mathcal{Q}|\mathcal{J}$ is composed of $(N - M)!$ possibly overlapping ${}^N\mathcal{B}|\mathcal{L}$ s. Thus, \mathcal{R}_M^N is composed of a total of $\binom{N}{M} 2^{N-M} (N - M)! = \frac{N!}{M!} 2^{N-M}$ rectangular hypersolids, which obviously grows very quickly with the codimension $N - M$.

[Insert Figure 7 Here]

In order to calculate the bounds of a ${}^N\mathcal{B}|\mathcal{L}$ rectangular hypersolid, we will need the *synaptic input spectrum* $\mathcal{I}_i^S(\mathbf{W})$ of neuron i : the set of possible synaptic inputs received by neuron i from the subcircuit consisting of the neurons in S . The min and max elements of this set define the range of synaptic inputs that neuron i can receive (Figure 1). \mathcal{I} can be defined as

$$\mathcal{I}_i^S(\mathbf{W}) \equiv \Sigma \mathcal{P}ow(\{w_{ij} \mid j \in S \text{ and } j \neq i\}) \quad (4.3)$$

where $\mathcal{P}ow(\mathcal{W})$ denotes the power set of the set of weights \mathcal{W} and $\Sigma \mathcal{P}ow(\mathcal{W})$ denotes the set of sums of the elements of the sets in $\mathcal{P}ow(\mathcal{W})$. For example, the synaptic input spectrum of neuron 2 in the $\{2, 3, 4\}$ subcircuit of a 4-neuron network is

$$\begin{aligned} \mathcal{I}_2^{\{2,3,4\}} &= \Sigma \mathcal{P}ow(\{w_{23}, w_{24}\}) \\ &= \Sigma \{\{\}, \{w_{23}\}, \{w_{24}\}, \{w_{23}, w_{24}\}\} \\ &= \{0, w_{23}, w_{24}, w_{23} + w_{24}\} \end{aligned}$$

We can now write each ${}^N\mathcal{B}|\mathcal{L}(\mathbf{W})$ as a set of N inequalities. In order to obtain the tightest possible bounds, we must treat the saturated and dynamically active neurons separately. Consider a neuron i in a circuit with a set of saturated neurons S and a set of dynamically active neurons A and let I_i be the net synaptic input that i receives from the neurons in S . Then neuron i will be saturated off when the left edge of its fold ($I_L(w_{ii}) - (\theta_i + I_i)$) falls above the right edge of the range of synaptic input it receives from the active neurons ($\max \mathcal{I}_i^A$) (left rectangle in Figure 1B), it will be saturated on when the right edge of its fold ($I_R(w_{ii}) - (\theta_i + I_i)$) falls below the left edge of the range of synaptic input it receives from the active neurons ($\min \mathcal{I}_i^A$) (right rectangle in Figure 1B), and it will be dynamically active otherwise.

More formally, each saturated neuron whose index appears in \mathcal{L} leads to an inequality of the form $l_{v_i} < \theta_{v_i} < u_{v_i}$. Write the ordered index list \mathcal{L} as the element-wise product of an elevation vector \mathbf{e} and an index vector \mathbf{v} : $\mathcal{L}_i = e_i v_i$, where $e_i = 1$ if index v_i is raised in \mathcal{L} and 0 if it is lowered. Then the inequality bounds for a saturated neuron can be written as

$$l_{v_i} = \begin{cases} -\infty & \text{if } e_i = 0 \\ I_R(w_{v_i v_i}) - \min \mathcal{I}_{v_i}^{\mathcal{N} \setminus \{v_1, \dots, v_{i-1}\}} - \sum_{j=1}^{i-1} e_j w_{v_i v_j} & \text{if } e_i = 1 \end{cases} \quad (4.4)$$

$$u_{v_i} = \begin{cases} I_L(w_{v_i v_i}) - \max \mathcal{I}_{v_i}^{\mathcal{N} \setminus \{v_1, \dots, v_{i-1}\}} - \sum_{j=1}^{i-1} e_j w_{v_i v_j} & \text{if } e_i = 0 \\ \infty & \text{if } e_i = 1 \end{cases}$$

where $\mathcal{N} = \{1, \dots, N\}$ and the set difference $\mathcal{S}_1 \setminus \mathcal{S}_2$ is the set consisting of all elements of \mathcal{S}_1 that are not in \mathcal{S}_2 .

For example, for ${}^4\mathcal{B}^4{}_1^3$ we have $\mathcal{L} = \{4, 1, 3\}$, with $\mathbf{e} = (1, 0, 1)$ and $\mathbf{v} = (4, 1, 3)$. Consider neuron 1 in this circuit, which occurs at an index of $i = 2$ in \mathcal{L} . The notation ${}^4\mathcal{B}^4{}_1^3$ tells us that neuron 1 is saturated off ($e_2 = 0$), which requires that the left edge of its fold falls to the right of the maximum synaptic input $\max \mathcal{I}_1^{\{1, 2, 3, 4\} \setminus \{4\}} = \max \mathcal{I}_1^{\{1, 2, 3\}}$ it receives (see Figure 1B). The left edge of its fold is given by $I_L(w_{11})$ offset by the net input it receives from the other saturated neurons relative to its own bias: $I_L(w_{11}) - \theta_1 - \sum_{j=1}^1 e_j w_{v_i v_j} = I_L(w_{11}) - \theta_1 - e_1 w_{14} = I_L(w_{11}) - \theta_1 - w_{14}$. Thus, the neuron 1 boundary of ${}^4\mathcal{B}^4{}_1^3$ is $I_L(w_{11}) - \theta_1 - w_{14} > \max \mathcal{I}_1^{\{1, 2, 3\}}$ or $\theta_1 < I_L(w_{11}) - \max \mathcal{I}_1^{\{1, 2, 3\}} - w_{14}$.

On the other hand, each dynamically active neuron i in $\mathcal{A} = \mathcal{N} \setminus \mathcal{L}$ leads to an inequality of the form $l_i < \theta_i < u_i$, with

$$l_i = I_L(w_{ii}) - \max \mathcal{I}_i^{\mathcal{A}} - \sum_{j=1}^{N-M} e_j w_{i v_j} \quad (4.5)$$

$$u_i = I_R(w_{ii}) - \min \mathcal{I}_i^{\mathcal{A}} - \sum_{j=1}^{N-M} e_j w_{i v_j}$$

For example, for ${}^4\mathcal{B}^4{}_1{}^3$ we have $\mathcal{A} = \{1, 2, 3, 4\} \setminus \{4, 1, 3\} = \{2\}$. For neuron 2 to be dynamically active, its fold must overlap the range of synaptic input it receives from the other dynamically active neurons. The minimum and maximum synaptic input are given by $\min \mathcal{I}_2^{\{2\}} = 0$ and $\max \mathcal{I}_2^{\{2\}} = 0$, respectively. The left (resp. right) edges of the fold are given by $I_L(w_{22})$ (resp. $I_R(w_{22})$) offset by the net input it receives from the saturated neurons relative to its own bias: $I_L(w_{22}) - \theta_2 - \sum_{j=1}^3 e_j w_{2v_j} = I_L(w_{22}) - \theta_2 - (e_1 w_{24} + e_2 w_{21} + e_3 w_{23}) = I_L(w_{22}) - \theta_2 - w_{23} - w_{24}$ (resp. $I_R(w_{22}) - \theta_2 - w_{23} - w_{24}$). Thus, the neuron 2 bound is $I_L(w_{22}) - w_{23} - w_{24} < \theta_2 < I_R(w_{22}) - w_{23} - w_{24}$.

Completing this example, the rectangular hypersolid ${}^4\mathcal{B}^4{}_1{}^3$ would be defined by the four inequalities

$$\begin{aligned}\theta_1 &< I_L(w_{11}) - \max \mathcal{I}_1^{\{1,2,3\}} - w_{14} \\ I_L(w_{22}) - w_{23} - w_{24} &< \theta_2 < I_R(w_{22}) - w_{23} - w_{24} \\ I_R(w_{33}) - \min \mathcal{I}_3^{\{2,3\}} - w_{34} &< \theta_3 \\ I_R(w_{44}) - \min \mathcal{I}_4^{\{1,2,3,4\}} &< \theta_4\end{aligned}$$

Despite the notational complexity of this section, the basic idea is straightforward. When all $w_{ii} > 4$, the extremal saddle-node bifurcation manifolds divide the net input parameter space of an N -neuron CTRNN into regions of dynamics whose effective dimensionality range from 0 to N . Furthermore, the combinatorial structure, location and extent of an asymptotic approximation to each of these regions can be calculated explicitly using equations (4.1)-(4.5). *Mathematica* code for the computation (and display when $N = 2$ or 3) of $\mathcal{R}_M^N(\mathbf{W})$ can be found in the electronic supplement.

4.2 Some $w_{ii} < 4$

In contrast to the $w_{ii} > 4$ case, the SSIO of a neuron whose self-weight is less than 4 is unfolded (Figure 1A). Thus, such a neuron will not undergo the extremal saddle-node bifurcations that play such a crucial role in the parameter space structure characterized above. Can this analysis be extended to circuits containing such neurons?

To gain some insight into this question, Figure 8 compares the net input parameter space of a 2-neuron CTRNN with $w_{22} = 5$ (A) and $w_{22} = 3$ (B). Note that the left and right branches of saddle-node bifurcations disappear when w_{22} passes below 4, as expected. However, there are still differences in the effective dimensionality of the dynamics of this circuit as θ_2 is varied. For example, between the two saddle-node bifurcation curves, the 3 equilibrium point phase portrait changes from occupying the interior of the state space (and therefore being effectively 2-dimensional in the distribution of its equilibrium points and its response to perturbations) at the point C1 to occupying only the bottom edge (effectively 1-dimensional) at C2. Likewise, outside the saddle-node bifurcation curves, the single equilibrium point phase portrait changes from occupying the right edge (effectively 1-dimensional) at D1 to occupying the bottom right-hand corner (effectively 0-dimensional) at D2.

Thus, when $w_{22} < 4$, regions of dynamics with different effective dimensionality still exist, but there are no sharp boundaries between them because these regions are no longer delineated by saddle-node bifurcations. If we wish to extend our definitions of the $\mathcal{R}_M^N(\mathbf{W})$ boundaries from the previous section to the case when some self-weights are less than 4, then we need to identify some feature of a neuron's unfolded SSIO curve against which we can compare the range of synaptic inputs that it receives. Different choices will lead to somewhat different boundaries.

[Insert Figure 8 Here]

Perhaps the simplest way to accomplish this is to make use of the piecewise linear approximation (dashed lines in Figure 1A)

$$\tilde{\sigma}(y + \theta) = \begin{cases} 0 & y < -\theta - 2 \\ \frac{y + \theta}{4} + \frac{1}{2} & -\theta - 2 \leq y \leq -\theta + 2 \\ 1 & y > -\theta + 2 \end{cases}$$

and use the points where the linear pieces intersect as markers for boundary calculations (black points in Figure 1A). By setting the resulting 1-neuron equation to 0 and solving for the net

input, we obtain left and right boundaries of -2 and $2 - w$, respectively, leading to the extended definitions

$$\tilde{I}_L(w) = \begin{cases} -2 & w < 4 \\ I_L(w) & w \geq 4 \end{cases}$$

$$\tilde{I}_R(w) = \begin{cases} 2 - w & w < 4 \\ I_R(w) & w \geq 4 \end{cases}$$

It is important to reiterate that these extended definitions can no longer be interpreted as folds when $w < 4$. Rather, they correspond to boundaries between saturated and dynamical active behavior.

If we everywhere replace $I_L(w)$ (resp. $I_R(w)$) by $\tilde{I}_L(w)$ (resp. $\tilde{I}_R(w)$) in our previous analysis, we obtain the extended regions $\tilde{\mathcal{R}}_M^N(\mathbf{W})$, ${}^N\tilde{\mathcal{Q}}|\mathcal{J}(\mathbf{W})$ and ${}^N\tilde{\mathcal{B}}|\mathcal{L}(\mathbf{W})$ that are valid for all w_{ii} . In our 2-neuron example, these extended definitions give rise to the shaded regions shown in Figure 8B. Although the original region definitions will be used in the remainder of this paper, the extended definitions could be substituted at any point.

5. Calculating \mathcal{R}_M^N Probabilities

In many applications, knowing what can happen in principle is often far less useful than knowing what typically does happen in practice. Probabilistic calculations can be used to characterize the most likely behavior under various conditions. In this section, we study the probability that a random parameter sample will encounter a region of M -dimensional dynamics in an N -neuron circuit. Such calculations provide estimates of the dynamical complexity of randomly-chosen circuits. In addition, because the \mathcal{R}_M^N boundaries become increasingly complex in higher dimensions, a statistical description can provide a very useful summary of the overall scaling of the structure of $\mathcal{C}_{\text{CTRNN}}(N)$ with N . Probabilistic calculations are also important for the application of stochastic search techniques such as evolutionary algorithms to CTRNNs (Beer & Gallagher, 1992; Harvey et al., 1997; Nolfi & Floreano, 2000), where search behavior is dominated by the most common dynamics in parameter space. They can help to determine how

best to seed the initial population of an evolutionary search. They can help select weight and bias parameter ranges (which is often done in an ad hoc manner) so as to maximize the chances of encountering interesting dynamics. They can also help to explain the dynamics of the evolutionary search itself, and to assess the evolvability of different types of dynamics.

Assuming that weights are in the range $[w_{\min}, w_{\max}]$ and biases are in the range $[\theta_{\min}, \theta_{\max}]$, this probability is given by the fraction of parameter space volume occupied by the region of interest

$$P(\mathcal{R}_M^N) = \frac{\text{vol}(\mathcal{R}_M^N)}{(w_{\max} - w_{\min})^{N^2} (\theta_{\max} - \theta_{\min})^N} \quad (5.1)$$

where $\text{vol}(\mathcal{R})$ denotes the volume of the region \mathcal{R} .

Since \mathcal{R}_M^N is composed of nonoverlapping polytopes ${}^N\mathcal{Q}|\mathcal{J}$, we have

$$\text{vol}(\mathcal{R}_M^N) = \sum_{\substack{\mathcal{S} \in \mathcal{K}_{N-M}^N \\ \mathcal{J} \in \mathcal{Z}(\mathcal{S})}} \text{vol}({}^N\mathcal{Q}|\mathcal{J}) \quad (5.2)$$

It is more difficult to compute $\text{vol}({}^N\mathcal{Q}|\mathcal{J})$ because its constituent ${}^N\mathcal{B}|\mathcal{L}$ rectangular hypersolids can overlap. In general, the volume of these possibly overlapping sets is given by the sum of the volumes of the individual sets adjusted by the volumes of their various overlaps:

$$\text{vol}({}^N\mathcal{Q}|\mathcal{J}) = \sum_{\mathcal{L} \in \mathcal{P}(\mathcal{J})} \text{vol}({}^N\mathcal{B}|\mathcal{L}) - \sum_{i=2}^{|\mathcal{P}(\mathcal{J})|} (-1)^i \sum_{\mathcal{H} \in \mathcal{K}_i^{\mathcal{P}(\mathcal{J})}} \text{vol}\left(\bigcap_{\mathcal{L} \in \mathcal{H}} {}^N\mathcal{B}|\mathcal{L}\right) \quad (5.3)$$

where $\mathcal{P}(\mathcal{J})$ once again denotes the permutations of \mathcal{J} , $|\mathcal{P}(\mathcal{J})|$ denotes the cardinality of $\mathcal{P}(\mathcal{J})$, and $\mathcal{K}_K^{\mathcal{P}(\mathcal{J})}$ denotes the K -subsets of $\mathcal{P}(\mathcal{J})$. For example, $\text{vol}({}^3\mathcal{Q}^{12}) = \text{vol}({}^3\mathcal{B}^{12}) + \text{vol}({}^3\mathcal{B}^{21}) - \text{vol}({}^3\mathcal{B}^{12} \cap {}^3\mathcal{B}^{21})$ (see Figure 7). Since the ${}^N\mathcal{B}|\mathcal{L}$ are rectangular hypersolids, each intersection in Eqn. (5.3) will also be a rectangular hypersolid, the bounds of which can be found by taking the appropriate maxs and mins of the bounds of the constituent ${}^N\mathcal{B}|\mathcal{L}$ s.

Finally, the volume of each rectangular hypersolid ${}^N\mathcal{B}|\mathcal{L}$ is given by

$$\text{vol}({}^N\mathcal{B}|\mathcal{L}) = \int_{\mathcal{W}} \prod_{i=1}^N ([u_i]_{\theta_{\min}}^{\theta_{\max}} - [l_i]_{\theta_{\min}}^{\theta_{\max}})$$

where \mathcal{W} is the hypercube $[w_{\min}, w_{\max}]^{N^2}$, the expressions u_i and l_i for the bounds of the i^{th} dimension of ${}^N\mathcal{B}|\mathcal{L}$ are given in Eqns. (4.4) and (4.5), and the notation $[x]_{\min}^{\max}$ means to clip x to the bounds $[\min, \max]$. Since u_i and l_i depend only on the N weights coming into neuron i , denoted \mathcal{W}_i , this N^2 -dimensional integral can be factored into the product of N N -dimensional integrals as

$$\text{vol}({}^N\mathcal{B}|\mathcal{L}) = \prod_{i=1}^N \int_{\mathcal{W}_i} ([u_i]_{\theta_{\min}}^{\theta_{\max}} - [l_i]_{\theta_{\min}}^{\theta_{\max}}) \quad (5.4)$$

Thus, in order to calculate $P(\mathcal{R}_M^N)$, we must evaluate Eqns. (5.1)-(5.4). *Mathematica* code supporting the construction and evaluation of such expressions for sufficiently small $N - M$ is provided in the electronic supplement. Note that these expressions are not as efficient as they could be. By taking into account integral symmetries, it should be possible to derive equivalent expressions that involve the evaluation of considerably fewer integrals.

As a concrete illustration of the calculation of $P(\mathcal{R}_M^N)$, consider the region \mathcal{R}_N^N of N -dimensional dynamics in an N -neuron circuit. This is not only the simplest case, but also the most important, because all N neurons are dynamically active. In this case, \mathcal{R}_N^N consists of a single rectangular hypersolid and thus

$$\text{vol}(\mathcal{R}_N^N) = \text{vol}({}^N\mathcal{Q}) = \text{vol}({}^N\mathcal{B})$$

Since all N integrals are identical, (5.4) can be written as

$$\text{vol}(\mathcal{R}_N^N) = \left(\int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} [I_R(w) - \min \mathcal{I}^N]_{\theta_{\min}}^{\theta_{\max}} - [I_L(w) - \max \mathcal{I}^N]_{\theta_{\min}}^{\theta_{\max}} dw_1 \cdots dw_{N-1} dw \right)^N \quad (5.5)$$

where w_i are the incoming weights to an arbitrary neuron, w is that neuron's self-connection, and \mathcal{I}^N is an abbreviation for $\mathcal{I}_i^{\{1, \dots, N\}}$ for arbitrary i . Note that the lower limit of the outermost integral must be 4 because we are using the original region definitions (which are only defined

for $w \geq 4$) rather than the extended ones. Note also that we have assumed that $w_{\min} = -w_{\max}$ for simplicity and $w_{\max} \geq 4$ so that $\text{vol}(\mathcal{R}_N^N)$ is nonzero.

Although it is unclear in general how to evaluate these arbitrarily-iterated piecewise integrals in closed-form, it is possible to evaluate them for fixed w and θ bounds and fixed N (see Appendix A.1). In addition, depending on the range of θ_{\min} and θ_{\max} relative to the points θ_{\min}^* and θ_{\max}^* where clipping begins to occur, there are two cases of interest where evaluation of these integrals for general N is relatively straightforward: (1) when clipping dominates Eqn. (5.5) and (2) when no clipping occurs.

The points θ_{\min}^* and θ_{\max}^* can be defined as follows. For $4 \leq w \leq w_{\max}$, we have $I_R(w) \in [I_R(w_{\max}), -2]$ and $I_L(w) \in [I_L(w_{\max}), -2]$. Since \mathcal{I}^N has the form $\{0, \pm w_{\max}, \dots, \pm w_{\max}(N-1)\}$, we can conclude that $I_R(w) - \min \mathcal{I}^N \in [I_R(w), w_{\max}(N-1)-2]$ and that $I_L(w) - \max \mathcal{I}^N \in [I_L(w_{\max}) - w_{\max}(N-1), -2]$, giving

$$\begin{aligned}\theta_{\min}^* &= I_L(w_{\max}) - w_{\max}(N-1) \\ \theta_{\max}^* &= w_{\max}(N-1) - 2\end{aligned}$$

The first case in which the iterated integrals can be evaluated in closed form is when $\theta_{\min} \gg \theta_{\min}^*$ and $\theta_{\max} \ll \theta_{\max}^*$, which will occur when N becomes sufficiently large relative to fixed θ_{\min} and θ_{\max} . In this case, the integrands are almost everywhere clipped to either θ_{\min} or θ_{\max} and the iterated integrals evaluate to

$$\text{vol}_{\infty}(\mathcal{R}_N^N) = \left((\theta_{\max} - \theta_{\min})(w_{\max} - 4)(2w_{\max})^{N-1} \right)^N$$

where the ∞ subscript reminds us that this expression is accurate only for sufficiently large N . The probability of a random parameter sample hitting \mathcal{R}_N^N therefore scales as

$$P_{\infty}(\mathcal{R}_N^N) = \frac{\text{vol}_{\infty}(\mathcal{R}_N^N)}{(2w_{\max})^{N^2} (\theta_{\max} - \theta_{\min})^N} = \left(\frac{w_{\max} - 4}{2w_{\max}} \right)^N$$

The second case in which the iterated integrals can be calculated in closed form is when $\theta_{\min} \leq \theta_{\min}^*$ and $\theta_{\max} \geq \theta_{\max}^*$, which will occur when N is small relative to fixed θ_{\min} and θ_{\max} . In

this case, the θ bounds are sufficiently large that no clipping occurs and the $[\cdot]_{\theta_{\min}}^{\theta_{\max}}$ can be dropped from the integrands. The integrals can then be evaluated (see Appendix A.2) to obtain

$$\text{vol}_0(\mathcal{R}_N^N) = \left(2^{N-2} (w_{\max})^{N-1} \left(w_{\max} \left(N(w_{\max} - 4) - w_{\max} + \sqrt{w_{\max}(w_{\max} - 4)} + \ln 256 + 4 \right) - 8(w_{\max} - 1) \ln \left(\sqrt{w_{\max} - 4} + \sqrt{w_{\max}} \right) + 2\sqrt{w_{\max}(w_{\max} - 4)} - 8\ln 2 \right) \right)^N$$

where the 0 subscript reminds us that this expression is accurate only for sufficiently small N . The probability in this case thus scales as

$$P_0(\mathcal{R}_N^N) = \frac{\text{vol}_0(\mathcal{R}_N^N)}{(2w_{\max})^{N^2} (\theta_{\max} - \theta_{\min})^N}$$

Figure 9A shows the scaling of the two approximations P_0 and P_∞ with N for the particular case $w_{\max} = 16$, $\theta_{\min} = -24$ and $\theta_{\max} = 24$ (note that here and throughout the remainder of the paper, such specific values are for illustrative purposes only). Superimposed on these curves are data points taken from 10^6 random samples of parameter space at each N . Note that $P_0(\mathcal{R}_N^N)$ provides the better fit for $N < 5$, whereas $P_\infty(\mathcal{R}_N^N)$ provides the better fit for $N > 7$. The data actually begins to deviate from P_0 by $N = 2$ (since $\theta_{\min} = -24 < \theta_{\min}^* = I_L(16) - 16(N-1)$ for $N > 1$), but the largest error occurs in the crossover region between these two curves, where the θ clipping begins to become significant. This becomes even more apparent for the narrower bounds $\theta_{\min} = -16$ and $\theta_{\max} = 16$ (Figure 9B). If higher accuracy is required in this crossover region, then the full iterated integrals for $\text{vol}(\mathcal{R}_N^N)$ must be evaluated (Appendix A.1). Such calculations can also be used to choose appropriate $[w_{\min}, w_{\max}]$ and $[\theta_{\min}, \theta_{\max}]$ parameter ranges so as to maximize $P(\mathcal{R}_M^N)$ for a CTRNN of a given size.

[Insert Figure 9 Here]

6. Calculating Phase Portrait Probabilities

As classification schemes, the effective dimensionality equivalence relation described above and the topological conjugacy equivalence relation normally studied in dynamical systems theory are distinct. A parameter space region containing dynamics with a given effective

dimension may include many nonequivalent phase portraits. Conversely, a given phase portrait may appear in multiple regions with different effective dimensionality. However, when the defining conditions are simple enough, we can estimate the densities of individual phase portraits using geometric reasoning similar to that employed above. In this section, we give two examples of such calculations.

6.1 Maximal Phase Portraits

The maximum number of equilibrium points that an N -neuron CTRNN can exhibit is 3^N (Beer, 1995). A sufficient (but not quite necessary) condition for the occurrence of this maximal phase portrait $\mathcal{P}_{3^N}^N$ is that the range of synaptic input each neuron receives from the other neurons should fall entirely within its SSIO fold, i.e. $I_L(w_{ii}) - \min \mathcal{I}_i^N \leq \theta_i \leq I_R(w_{ii}) - \max \mathcal{I}_i^N$ for all neurons i . Thus, the density of this phase portrait in the parameter space of an N -neuron circuit can be estimated as

$$P(\mathcal{P}_{3^N}^N) = \frac{\left(\int_4^{w_{\max}} \int_{w_{\min}}^{w_{\max}} \cdots \int_{w_{\min}}^{w_{\max}} \left[\left[I_R(w) - \max \mathcal{I}^N \right]_{\theta_{\min}}^{\theta_{\max}} - \left[I_L(w) - \min \mathcal{I}^N \right]_{\theta_{\min}}^{\theta_{\max}} \right]_0 dw_1 \cdots dw_{N-1} dw \right)^N}{(w_{\max} - w_{\min})^{N^2} (\theta_{\max} - \theta_{\min})^N}$$

where the outer clipping to 0 in the integrand enforces the condition $I_W(w_{ii}) > \text{range } \mathcal{I}_i^N \equiv \max \mathcal{I}^N - \min \mathcal{I}^N$, ensuring the existence of this phase portrait for a given set of weights. Not surprisingly, this arbitrarily-iterated piecewise integral is also very difficult to evaluate in general, but can often be evaluated for particular N and parameter ranges.

As an example, consider the density of phase portraits with 9 equilibrium points \mathcal{P}_9^2 in a 2-neuron CTRNN whose weights and biases are in the range [-16, 16]. In this case, the previous integral reduces to

$$P(\mathcal{P}_9^2) = \frac{\left(\int_4^{16} \int_{-16}^{16} \left[\left[I_R(w) - \max(0, w_1) \right]_{-16}^{16} - \left[I_L(w) - \min(0, w_1) \right]_{-16}^{16} \right]_0 dw_1 dw \right)^2}{32^4 32^2} \quad (6.1)$$

This expression can then be evaluated (see Appendix A.3) to obtain

$$P(\mathcal{P}_9^2) = \frac{(1152 - 576\sqrt{3}\ln(2 + \sqrt{3}) + 240\ln(2 + \sqrt{3})^2)^2}{1073741824} \approx 0.0060\%$$

For comparison, a random sample of 10^6 2-neuron circuits exhibited the 9 equilibrium point phase portrait with a probability of 0.0059%.

Clearly, the maximal phase portrait is very rare in the parameter space of 2-neuron CTRNNs, and only becomes more so in larger networks. However, we can use our knowledge of the SSIO geometry underlying this phase portrait to substantially increase its probability. For example, if we randomly choose the cross-weights, then randomly chose each self-weight such that $I_W(w_{ii}) > \text{range } \mathcal{I}_i^N$ (assuming $w_{ii} > 4$), and then randomly choose each bias from the range $[I_L(w_{ii}) - \min \mathcal{I}_i^N, I_R(w_{ii}) - \max \mathcal{I}_i^N]$, we are *guaranteed* to obtain the maximal phase portrait.

6.2 Central Oscillation Phase Portraits

Some phase portraits do not have even approximate defining conditions that can be expressed solely in terms of the geometry of SSIO curves. However, we can still sometimes calculate a good estimate of occurrence probability in such cases. Consider, for example, the density of central oscillatory dynamics in a 2-neuron CTRNN. By “central”, I mean a phase portrait \mathcal{P}_{ILC}^2 in which a single stable limit cycle surrounds a single unstable equilibrium point in the neighborhood of a center-crossing configuration. While there are other oscillatory phase portraits that can occur in 2-neuron CTRNNs (Borisuk & Kirillov, 1992; Beer, 1995; Hoppensteadt & Izhikevich, 1997; Ermentrout, 1998), their defining conditions are delicate and they are therefore likely to make only a small contribution to oscillation probability.

By linearizing about the center-crossing point (θ_1^*, θ_2^*) and solving for the Hopf bifurcation condition (see Appendix A.4), we can derive two pairs of expressions $H_1^\pm(\Delta\theta_2; \mathbf{W}, \tau)$ and $H_2^\pm(\Delta\theta_1; \mathbf{W}, \tau)$ such that $\theta_1^* + H_1^\pm(\theta_2 - \theta_2^*; \mathbf{W}, \tau)$ and $\theta_2^* + H_2^\pm(\theta_1 - \theta_1^*; \mathbf{W}, \tau)$ approximate the Hopf bifurcation curve in net input space. A comparison of these approximations (gray curves) to the actual Hopf curves (dashed curves) is shown in Figure 10. Note that the approximate Hopf curves can extend beyond the inner edge of the saddle-node bifurcation “tongues” even though

the central oscillations terminate there in saddle-node bifurcations on a loop. Note also that the character of these curves changes from elliptical to hyperbolic as one of the self-weights passes through 0. Indeed, when one self-weight becomes sufficiently negative, the oscillatory region splits into two distinct regions and central oscillations no longer exist.

[Insert Figure 10 Here]

There are several different ways we can use these curves to estimate $P(\mathcal{P}_{ILC}^2)$. Perhaps the simplest approach is to approximate the oscillatory region by the rectangle $[\theta_1^* + H_1^-(0; \mathbf{W}, \tau), \theta_1^* + H_1^+(0; \mathbf{W}, \tau)] \times [\theta_2^* + H_2^-(0; \mathbf{W}, \tau), \theta_2^* + H_2^+(0; \mathbf{W}, \tau)]$ obtained from the upper and lower bounds of the Hopf curve approximation at the center-crossing point (gray rectangles in Figure 10). If we clip this rectangle to the θ bounds and saddle-node bifurcation manifolds, clip its width and height to 0, integrate over the weights and time constants, and divide by the total volume of parameter space, we obtain

$$P(\mathcal{P}_{ILC}^2) = \frac{2 \int_{\mathcal{D}} \left[[\theta_1^* + H_1^+(0; \mathbf{W}, \tau)]_{a_1}^{b_1} - [\theta_1^* - H_1^-(0; \mathbf{W}, \tau)]_{a_1}^{b_1} \right]_0 \times \left[[\theta_2^* + H_2^+(0; \mathbf{W}, \tau)]_{a_2}^{b_2} - [\theta_2^* - H_2^-(0; \mathbf{W}, \tau)]_{a_2}^{b_2} \right]_0}{(w_{\max} - w_{\min})^N (\theta_{\max} - \theta_{\min})^N (\tau_{\max} - \tau_{\min})^N}$$

where

$$a_i = \begin{cases} \max(\theta_{\min}, I_R(w_{ii}) - \max(0, w_{ij, i \neq j})) & w_{ii} \geq 4 \\ \theta_{\min} & w_{ii} < 4 \end{cases}$$

$$b_i = \begin{cases} \min(\theta_{\max}, I_L(w_{ii}) - \min(0, w_{ij, i \neq j})) & w_{ii} \geq 4 \\ \theta_{\max} & w_{ii} < 4 \end{cases}$$

clip the Hopf curve to the saddle-node bifurcation manifolds and θ bounds, and

$$\mathcal{D} : \{(w_{11}, w_{12}, w_{21}, w_{22}, \tau_1, \tau_2) \mid w_{\min} \leq w_{ii} \leq w_{\max}, 0 \leq w_{12} \leq w_{\max}, w_{\min} \leq w_{21} \leq 0, \tau_{\min} \leq \tau_i \leq \tau_{\max}, \kappa\eta_1, \kappa\eta_2 \geq 0\}$$

gives the domain of integration (see Appendix A.4 for an explanation of the last two conditions). Since it is well known that oscillations can only occur in a 2-neuron CTRNN when the cross-

weights are oppositely signed (Ermentrout, 1995), we have assumed above that $w_{12} > 0$ and $w_{21} < 0$ and doubled the integral to account for the opposite possibility.

We will not attempt to evaluate this integral in closed form. Assuming $w_{ij}, \theta_i \in [-16, 16]$ and $\tau_i \in [0.5, 10]$, numerical integration using a quasirandom Monte Carlo method gives $P(\mathcal{P}_{1LC}^2) \approx 0.22\%$, which accords quite well with the empirical probability of 0.24% observed in a random sample of 10^6 2-neuron circuits. The empirical estimate was obtained by randomly generating 10 initial conditions in the range $y_i \in [-16, 16]$, integrating each with the forward Euler method for 2500 integration steps of size 0.1 to skip transients, and then integrating for an additional 500 integration steps. If the output of either neuron varied by more than 0.05 during this second integration for any initial condition, then the circuit was classified as oscillatory. Since the empirical value includes both central and noncentral oscillations, it would be expected to be slightly higher.

How does the probability of oscillation $P(\mathcal{O}^N)$ scale with N in CTRNNs? By “oscillation”, I mean any asymptotic behavior other than an equilibrium point, so that periodic, quasiperiodic and chaotic dynamics are included. Although this question is beyond the theory described in this paper, we can examine it empirically. A plot of $P(\mathcal{O}^N)$ is shown in Figure 11A (black curve), with the $P(\mathcal{P}_{1LC}^2)$ value calculated above corresponding to the $N = 2$ point in this plot. For comparison, the scaling of oscillation probability with N for random center-crossing circuits is also shown (gray curve). Note that both curves monotonically increase toward 100%, although oscillations are clearly much more likely in random center-crossing circuits than they are in completely random CTRNNs (Beer, 1995; Mathayomchan & Beer, 2002). Interestingly, samples of oscillatory circuits taken from this data set suggest that chaotic dynamics becomes increasingly common for large N , which is consistent with other work on chaos in additive neural networks (Sompolinsky & Crisanti, 1988).

[Insert Figure 11 Here]

In order to gain some insight into the underlying structure of $P(\mathcal{O}^N)$, the probability $P(\mathcal{O}_M^N)$ that exactly M neurons are oscillating in an N -neuron circuit (with the remaining $N - M$ neurons

in saturation) is plotted in Figure 11B. As N increases, note that (1) the most probable oscillating subcircuit size (denoted M^*) increases, (2) the distribution of oscillatory subcircuits broadens, and (3) the probability of M^* increases. There are several factors underlying these features. First, the distribution broadens and shifts to the right because the range of possible subcircuit sizes grows with N . At least within the range of this plot, the shift in peak location with N is roughly $M^* \approx 1.26\sqrt{N} - 0.074$. Second, the probability of M^* increases both because the number of possible subcircuits $\sum_M \binom{N}{M} \sim 2^N$ grows exponentially and because the parameter ranges over which a subcircuit of a given size can oscillate increases with $N - M$. As long as $P(\mathcal{O}^N) = \sum_M P(\mathcal{O}_M^N) < 1$, the probability of M^* can continue to increase. However, as $P(\mathcal{O}^N)$ approaches 1, one would expect the probability of M^* to decrease as a fixed area is distributed across an increasing range of subcircuit sizes. Finally, the quantitative details of $P(\mathcal{O}_M^N)$ obviously depend on the relative proportion of different oscillatory regions that fall within the range of allowable bias values.

7. Discussion

Despite the extreme difficulty involved, there is a growing recognition of the need for a more general theory of neural circuits, even a partial one. One path toward such a theory involves the systematic study of the structure of the space of all possible circuits over a given model neuron. This paper begins such a study for the relatively simple but still dynamically universal CTRNN model. I have explicitly computed the local bifurcation manifolds of arbitrary CTRNNs in net input space, and presented visualizations of their structure for small N . I have also shown how the outermost envelope of saddle-node bifurcations formed by the saturation of the sigmoid activation function divides net input space into regions of dynamics with different effective dimensionality, and I have derived analytical approximations to these regions. While these regions by no means exhaust the intricate structure of CTRNN parameter space, they do provide a coarse map on which more detailed behavior can be situated. I have also demonstrated how to calculate estimates of the probabilities of these different regions and of specific phase portraits.

Although saturation is the most obvious feature of $\sigma(\cdot)$, it is still remarkable how this simple property almost completely dominates the structure of $\mathcal{C}_{\text{CTRNN}}(N)$. As N increases, the probability of finding circuits with saturated subcircuits exponentially overwhelms the probability of finding circuits in which all neurons are dynamically active. However, either activity-dependent regulation (Turrigiano, Abbott & Marder, 1994; Williams & Noble, in press) or biased sampling (Beer, 1995; Mathayomchan & Beer, 2002) can counteract this trend. The domination of saturation also suggests the utility of applying techniques from the logical analysis of switching networks (Lewis & Glass, 1992; Edwards, 2000; Thomas & Kaufman, 2001) or modular decomposition (Chiel, Beer & Gallagher, 1999; Beer, Chiel & Gallagher, 1999) to CTRNNs.

While there is no question that the richest dynamics that an N -neuron CTRNN can exhibit occurs when all N neurons are dynamically active, it would be a mistake to dismiss circuits in which some neurons are saturated as somehow less interesting. Even in, say, a 23-neuron circuit, 17-dimensional dynamics can be quite rich. Furthermore, it can be much easier to evolve 17-dimensional dynamics in a 23-neuron circuit than in a 17-neuron circuit, since $P(\mathcal{R}_{17}^{23})$ can be considerably higher than $P(\mathcal{R}_{17}^{17})$. In addition, N -neuron CTRNNs containing saturated subcircuits may be switchable into different M -dimensional dynamical modes by external inputs.

Despite the progress reported in this paper, much work remains to be done. It would be very useful to have an exact closed-form expression for $P(\mathcal{R}_N^N)$ for general parameter ranges, as well as either approximate or exact closed-form expressions for $P(\mathcal{R}_M^N)$ and $P(\mathcal{P}_{3^N}^N)$. Although the piecewise function capabilities of *Mathematica* (version 5.1 and above) are quite powerful, further advances in the evaluation of the arbitrarily-iterated piecewise integrals that arise in these calculations will be necessary. It would also be interesting to derive at least an approximate closed form for the $P(\mathcal{P}_{1LC}^2)$ calculations in Section 6.2 for general parameter ranges, as well as to improve the accuracy of these calculations by using nonrectangular approximations to the oscillatory region. More generally, it may be possible to calculate probability estimates for at least some of the other 2-neuron phase portraits, and perhaps some phase portraits in larger

circuits as well. Empirical estimates of the probabilities of many of the 2-neuron phase portraits have already been determined for a specific set of parameter ranges (Izquierdo-Torres, 2004).

It would also be interesting to characterize the structure of the nonextremal local bifurcation manifolds that appear in Figures 2 and 3. Further examination of the submanifold of center-crossing circuits may be fruitful in this regard. Note that, while there is only one true center-crossing point in net input space for any fixed set of weights, each subcircuit region has its own center-crossing submanifold. For example, at the centers of symmetry of the poles in Figure 3 lie lines of 2-neuron center-crossing circuits with the third neuron in saturation, while the centers of symmetry of the slabs likewise contain planes of 1-neuron center-crossing circuits. Thus, the center-crossing (sub)circuits form a kind of “skeleton” for $\mathcal{C}_{\text{CTRNN}}(N)$ within the “skin” of the extremal saddle-node bifurcations. By studying small perturbations from these center-crossing submanifolds, it may be possible to gain some insight into the structure of the nonextremal local bifurcation manifolds that lie between the skeleton and the skin.

Even in its present form, the theory described in this paper has significant applications to evolutionary robotics, where CTRNNs are widely used as controllers (Beer & Gallagher, 1992; Harvey et al., 1997; Nolfi & Floreano, 2000). For example, it would be interesting to explore the utility of seeding evolutionary searches with circuits drawn uniformly from \mathcal{R}_N^N , which includes a much greater variety of rich dynamics than the highly symmetric center-crossing circuits that have previously been utilized as seeds (Beer, 1995; Mathayomchan & Beer, 2002). The theory can also be used to select $[w_{\min}, w_{\max}]$ and $[\theta_{\min}, \theta_{\max}]$ parameter ranges which maximize $P(\mathcal{R}_N^N)$ for a CTRNN of a given size. Studies of neutrality (Izquierdo-Torres, 2004) and evolutionary dynamics in CTRNN parameter spaces (Seys & Beer, 2004) should also benefit from the theoretical picture of $\mathcal{C}_{\text{CTRNN}}(N)$ presented here, and the probability calculations we have demonstrated should be directly applicable to empirical studies of the implications of saturation for the evolvability of CTRNNs (Williams & Noble, in press). The theory could also be used to further work on the impact of network architecture on CTRNN dynamics (Psujek, Ames & Beer, in press), since different circuit architectures simply correspond to subspaces of $\mathcal{C}_{\text{CTRNN}}(N)$ with

some connection weights fixed to 0. Finally, given the qualitative similarities between CTRNNs and some models of genetic regulatory networks (de Jong, 2002), the work described here may even have applications in this area.

Returning to the recent results on the parameter spaces of invertebrate pattern generator models with which this paper began (Goldman et al., 2001; Golowasch et al., 2002; Prinz, Billimoria, & Marder, 2003; Prinz, Bucher & Marder, 2004), what might our CTRNN results tell us about the maximum conductance parameter spaces of more biophysically-realistic neural models? Although the parameter space structure of such models is considerably more complex, some of the insights gained from studies of CTRNNs, and the mathematical and computational tools developed to support such studies, might carry over. For example, various kinds of saturation effects, which play such an important role in structuring $\mathcal{C}_{\text{CTRNN}}(N)$, also occur throughout more realistic models. Thus, one would predict that saturation would similarly dominate the parameter spaces of conductance-based models. Furthermore, one would expect that the proper functioning of such models would impose constraints on the interactions between conductances analogous to those defining the \mathcal{R}_N^N region in CTRNNs. More generally, the study of CTRNN parameter space can serve as an exemplar for parameter space studies of conductance-based models, suggesting fruitful questions and directions and making tentative predictions. Indeed, the discovery of sensitivity and robustness to different combinations of conductance variations, failure of averaging, and multiple-instantiability in conductance-based models of invertebrate pattern generators was partly anticipated by earlier studies with CTRNNs (Chiel, Beer, & Gallagher, 1999; Beer, Chiel & Gallagher, 1999).

Ultimately, the work described in this paper needs to be extended to these more biophysically-realistic neural circuits. Perhaps the best place to begin such a complexification would be with 2-dimensional model neurons, which can be configured to qualitatively reproduce a wide range of nerve cell phenomenology (Rowat & Selverston, 1997; Rinzel & Ermentrout, 1998). Indeed, one interesting possibility would be to study circuits of 2-dimensional model neurons formed from pairs of CTRNN units, since the theory developed here could be directly

applied. From there, more complex model neurons can be explored, such as the 3-dimensional generic bursting model recently examined by Ghigliazza and Holmes (2004). Only by studying the simplest possible model circuits and then incrementally complicating them as our understanding progresses will we ever be able to approach a more general theory of neural circuits that can do justice to the parameter space complexity that studies of biological circuits are revealing.

Acknowledgments. I would like to thank Jeff Ames, Michael Branicky, Alan Calvitti, Hillel Chiel, Bard Ermentrout, Eldan Goldenberg, Robert Haschke, and Eduardo Izquierdo-Torres for their feedback on an earlier draft of this paper. This research was supported in part by NSF grant EIA-0130773.

References

- Beer, R.D. (2005). Electronic supplement to “Parameter space structure of continuous-time recurrent neural networks”. This *Mathematica* notebook is available for download at <http://vorlon.cwru.edu/~beer/CTRNNSupplement.nb>.
- Beer, R.D. (1995). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior* **3**:469-509.
- Beer, R.D., Chiel, H.J. and Gallagher, J.C. (1999). Evolution and analysis of model CPGs for walking II. General principles and individual variability. *J. Computational Neuroscience* **7**:119-147.
- Beer, R.D. and Gallagher, J.G. (1992). Evolving dynamical neural networks for adaptive behavior. *Adaptive Behavior* **1**:91-122.
- Borisuk, R.M. and Kirillov, A.B. (1992). Bifurcation analysis of a neural network model. *Biological Cybernetics* **66**:319-325.
- Blum, E.K. and Wang, X. (1992). Stability of fixed points and periodic orbits and bifurcations in analog neural networks. *Neural Networks* **5**:577-587.
- Chiel, H.J., Beer, R.D. and Gallagher, J.C. (1999). Evolution and analysis of model CPGs for walking I. Dynamical modules. *J. Computational Neuroscience* **7**:99-118.
- Chow, T.W.S. and Li, X.-D. (2000). Modeling of continuous time dynamical systems with input by recurrent neural networks. *IEEE Trans. On Circuits and Systems – I: Fundamental Theory and Applications* **47**:575-578.
- Cohen, M.A. and Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans. Systems, Man and Cybernetics* **13**:813-825.
- Cowan, J.D. and Ermentrout, G.B. (1978). Some aspects of the eigenbehavior of neural nets. In S.A. Levin (Ed.), *Studies in Mathematical Biology 1: Cellular Behavior and the Development of Pattern* (pp. 67-117). The Mathematical Association of America.
- de Jong, H. (2002). Modeling and simulation of genetic regulatory networks: A literature review. *J. Computational Biology* **9**:67-103.

- Dunn, N.A., Lockery, S.R., Pierce-Shimomura, J.T. and Conery, J.S. (2004). A neural network model of chemotaxis predicts functions of synaptic connections in the nematode *Caenorhabditis elegans*. *J. Computational Neuroscience* **17**:137-147.
- Edwards, R. (2000). Analysis of continuous-time switching networks. *Physica D* **146**:165-199.
- Ermentrout, B. (1995). Phase-plane analysis of neural activity. In M.A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks* (pp. 732-738). MIT Press.
- Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics* **61**:353-430.
- Funahashi, K.I. and Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks* **6**:801-806.
- Getting, P.A. (1989). Emerging principles governing the operation of neural networks. *Annual Review of Neuroscience* **12**:185-204.
- Ghigliazza, R.M. and Holmes, P. (2004). Minimal models of bursting neurons: How multiple currents, conductances and timescales affect bifurcation diagrams. *SIAM J. Applied Dynamical Systems* **3**:636-670.
- Goldman, M.S., Golowasch, J., Marder, E. and Abbott, L.F. (2001). Global structure, robustness and modulation of neuronal models. *J. Neuroscience* **21**:5229-5238.
- Golowasch, J., Goldman, M.S., Abbott, L.F. and Marder, E. (2002). Failure of averaging in the construction of a conductance-based neural model. *J. Neurophysiol.* **87**:1129-1131.
- Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks* **1**:17-61.
- Guckenheimer, J., Myers, M. and Sturmfels, B. (1997). Computing Hopf bifurcations I. *SIAM J. Numerical Analysis* **84**:1-21.
- Harvey, I., Husbands, P. Cliff, D., Thompson, A. and Jacobi, N. (1997). Evolutionary robotics: The Sussex approach. *Robotics and Autonomous Systems* **20**:205-224.
- Haschke, R. (2004). *Bifurcations in Discrete-Time Neural Networks: Controlling Complex Network Behavior with Inputs*. Ph.D. Thesis, Faculty of Technology, University of Bielefeld.

- Haschke, R. and Steil, J.J. (2005). Input space bifurcation manifolds of recurrent neural networks. *Neurocomputing* **64C**:25-38.
- Hirsch, M. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks* **2**:331-349.
- Hopfield, J.J. (1984). Neurons with graded response properties have collective computational properties like those of two-state neurons. *Proc. National Academy of Sciences* **81**:3088-3092.
- Hoppensteadt, F.C. and Izhikevich, E.M. (1997). *Weakly Connected Neural Networks*. Springer.
- Izquierdo-Torres, E. (2004). *Evolving Dynamical Systems: Nearly Neutral Regions in Continuous Fitness Landscapes*. M.S. Thesis, Dept. of Informatics, University of Sussex.
- Kimura, M. and Nakano, R. (1998). Learning dynamical systems by recurrent neural networks from orbits. *Neural Networks* **11**:1589-1599.
- Kuznetsov, Y.A. (1998). *Elements of Applied Bifurcation Theory* (Second edition). Springer.
- Lewis, J.E. and Glass, L. (2001). Nonlinear dynamics and symbolic dynamics of neural networks. *Neural Computation* **4**: 621-642.
- Marder, E. and Abbott, L.F. (1995). Theory in motion. *Current Opinion in Neurobiology* **5**:832-840.
- Marder, E. and Calabrese, R.L. (1996). Principles of rhythmic motor pattern generation. *Physiological Reviews* **76**:687-717.
- Mathayomchan, B. and Beer, R.D. (2002). Center-crossing recurrent neural networks for the evolution of rhythmic behavior. *Neural Computation* **14**:2043-2051.
- Nolfi, S. and Floreano, D. (2000). *Evolutionary Robotics*. MIT Press.
- Pasemann, F. (2002). Complex dynamics and the structure of small neural networks. *Network: Computation in Neural Systems* **13**:195-216.
- Prinz, A.A., Billimoria, C.P. and Marder, E. (2003). Alternative to hand-tuning conductance-based models: Construction and analysis of databases of model neurons. *J. Neurophysiol.* **90**:3998-4015.

- Prinz, A.A., Bucher, D. and Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nature Neuroscience* **7**:1345-1352.
- Psupek, S., Ames, J. and Beer, R.D. (2006). Connection and coordination: The interplay between architecture and dynamics in evolved model pattern generators. *Neural Computation* **18**:729-747.
- Rinzel, J. and Ermentrout, B. (1998). Analysis of neural excitability and oscillations. In C. Koch and I. Segev (Eds.), *Methods in Neuronal Modeling* (Second edition) (pp. 251-291). MIT Press.
- Rowat, P.F. and Selverston, A.I. (1997). Oscillatory mechanisms in pairs of neurons with fast inhibitory synapses. *J. Computational Neuroscience* **4**:103-127.
- Selverston, A.I. (1980). Are central pattern generators understandable? *Behavioral and Brain Sciences* **3**:535-571.
- Seys, C. W. and Beer, R.D. (2004). Evolving walking: The anatomy of an evolutionary search. In S. Schaal, A. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam and J.-A. Meyer (Eds.), *From Animals to Animats 8: Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior* (pp. 357-363). MIT Press.
- Sompolinsky, H. and Crisanti, A. (1988). Chaos in random neural networks. *Physical Review Letters* **61**(3):259-262.
- Tiňo, P. Horne, B.G., Giles, C.L. (2001). Attractive periodic sets in discrete-time recurrent neural networks (with emphasis on fixed-point stability and bifurcations in two-neuron networks). *Neural Computation* **13**:1379-1414.
- Thomas, R. and Kaufman, M. (2001). Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos* **11**:180-195.
- Turrigiano, G., Abbott, L.F. and Marder, E. (1994). Acitivity-dependent changes in the intrinsic properties of cultured neurons. *Science* **264**:974-976.

Williams, H. and Noble, J. (in press). Homeostatic plasticity improves signal propagation in continuous-time recurrent neural networks. To appear in *Biosystems*.

Wolfram, S. (2003). *The Mathematica Book* (Fifth edition). Wolfram Media.

Zhaojue, Z., Schieve, W.C. and Das, P.K. (1993). Two neuron dynamics and adiabatic elimination. *Physica D* **67**:224-236.

Appendix

A.1 Calculating $\text{vol}(\mathcal{R}_N^N)$ in closed form for fixed N

In this section, we show how to calculate the value of $\text{vol}(\mathcal{R}_N^N)$ from Eqn. (5.5) for fixed w and θ ranges and fixed N . We can write $\text{vol}(\mathcal{R}_N^N) = (R_N - L_N)^N$, where

$$R_N \equiv \int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} [I_R(w) - \min \mathcal{I}^N]_{\theta_{\min}}^{\theta_{\max}} dw_1 \cdots dw_{N-1} dw$$

$$L_N \equiv \int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} [I_L(w) - \max \mathcal{I}^N]_{\theta_{\min}}^{\theta_{\max}} dw_1 \cdots dw_{N-1} dw$$

By construction, \mathcal{I}^N has 2^{N-1} elements, with $\binom{N-1}{k}$ k -wise sums. Since the expression $\min \mathcal{I}^N$ will evaluate to each element of \mathcal{I}^N precisely when all terms in that element are negative, R_N can be written as

$$R_N = \binom{N-1}{0} \int_4^{w_{\max}} \int_0^{w_{\max}} \cdots \int_0^{w_{\max}} [I_R(w)]_{\theta_{\min}}^{\theta_{\max}} dw_1 \cdots dw_{N-1} dw$$

$$+ \binom{N-1}{1} \int_4^{w_{\max}} \int_0^{w_{\max}} \cdots \int_0^{w_{\max}} \int_{-w_{\max}}^0 [I_R(w) - w_1]_{\theta_{\min}}^{\theta_{\max}} dw_1 \cdots dw_{N-1} dw$$

$$+ \binom{N-1}{2} \int_4^{w_{\max}} \int_0^{w_{\max}} \cdots \int_0^{w_{\max}} \int_{-w_{\max}}^0 \int_{-w_{\max}}^0 [I_R(w) - w_1 - w_2]_{\theta_{\min}}^{\theta_{\max}} dw_1 \cdots dw_{N-1} dw$$

$$+ \cdots$$

$$+ \binom{N-1}{N-1} \int_4^{w_{\max}} \int_{-w_{\max}}^0 \cdots \int_{-w_{\max}}^0 \left[I_R(w) - \sum_{i=1}^{N-1} w_i \right]_{\theta_{\min}}^{\theta_{\max}} dw_1 \cdots dw_{N-1} dw$$

Evaluating the integrals involving positive weights (which do not appear in the corresponding integrands), we obtain

$$\begin{aligned}
R_N = & \binom{N-1}{0} (w_{\max})^{N-1} \int_0^{w_{\max}} [I_R(w)]_{\theta_{\min}}^{\theta_{\max}} dw \\
& + \binom{N-1}{1} (w_{\max})^{N-2} \int_4^{w_{\max}} \int_{-w_{\max}}^0 [I_R(w) - w_1]_{\theta_{\min}}^{\theta_{\max}} dw_1 dw \\
& + \binom{N-1}{2} (w_{\max})^{N-3} \int_4^{w_{\max}} \int_{-w_{\max}}^0 \int_{-w_{\max}}^0 [I_R(w) - w_1 - w_2]_{\theta_{\min}}^{\theta_{\max}} dw_1 dw_2 dw \\
& + \dots \\
& + \binom{N-1}{N-1} (w_{\max})^0 \int_4^{w_{\max}} \int_{-w_{\max}}^0 \dots \int_{-w_{\max}}^0 \left[I_R(w) - \sum_{i=1}^{N-1} w_i \right]_{\theta_{\min}}^{\theta_{\max}} dw_1 \dots dw_{N-1} dw
\end{aligned}$$

which sums to

$$R_N = \sum_{k=0}^{N-1} \binom{N-1}{k} (w_{\max})^{N-k-1} S_R^k \quad \text{with } S_R^k \equiv \int_4^{w_{\max}} \int_{-w_{\max}}^0 \dots \int_{-w_{\max}}^0 \left[I_R(w) - \sum_{i=1}^k w_i \right]_{\theta_{\min}}^{\theta_{\max}} dw_1 \dots dw_k dw$$

A similar derivation can be applied to L_N to obtain

$$L_N = \sum_{k=0}^{N-1} \binom{N-1}{k} (w_{\max})^{N-k-1} S_L^k \quad \text{with } S_L^k \equiv \int_4^{w_{\max}} \int_0^{w_{\max}} \dots \int_0^{w_{\max}} \left[I_L(w) - \sum_{i=1}^k w_i \right]_{\theta_{\min}}^{\theta_{\max}} dw_1 \dots dw_k dw$$

Unfortunately, it is not clear how to evaluate S_R^k and S_L^k in closed form for general k . There are two obstacles to be overcome. First, we must evaluate integrals of the form $\int_{\mathcal{D}^k} \sum_{i=1}^k x_i$ for general k , with $\mathcal{D}^k : \{(x_1, \dots, x_k) \mid -w_{\max} \leq x_i \leq 0, \sum_{i=1}^k x_i < I_R(w) - \theta_{\max}\}$ for S_R^k and $\mathcal{D}^k : \{(x_1, \dots, x_k) \mid 0 \leq x_i \leq w_{\max}, \sum_{i=1}^k x_i > I_L(w) - \theta_{\min}\}$ for S_L^k . Second, we must evaluate integrals of the form $\gamma_R^k \equiv \int_4^{w_{\max}} (I_R(w))^k dw$ and $\gamma_L^k \equiv \int_4^{w_{\max}} (I_L(w))^k dw$ for general k . Fortunately, using cylindrical decomposition, *Mathematica* can symbolically tabulate the values of such integrals for fixed k and fixed w and θ ranges. Thus, for modest N , it is possible to explicitly calculate $\text{vol}(\mathcal{R}_N^N)$ in closed form.

For example, consider the point of greatest discrepancy between the approximate curves and the empirical data in Figure 9B, which occurs at $N=4$. For $w_{\max}=16$, $\theta_{\min}=-16$ and $\theta_{\max}=16$, we obtain

$$\begin{aligned}
\text{vol}(\mathcal{R}_4^4) &= \left(4096(S_R^0 - S_L^0) + 768(S_R^1 - S_L^1) + 48(S_R^2 - S_L^2) + (S_R^3 - S_L^3) \right)^4 \\
&= \frac{\left(-7798784\gamma_R^0 - 557056\gamma_R^1 + 10752\gamma_R^2 + 128\gamma_R^3 - 2\gamma_R^4 - 7798784\gamma_L^0 + 557056\gamma_L^1 + 19968\gamma_L^2 + 256\gamma_L^3 + \gamma_L^4 \right)^4}{331776}
\end{aligned}$$

which numerically evaluates to a probability of $P(\mathcal{R}_4^4) \approx 0.375\%$. This theoretical value (plus sign in Figure 9B) matches the empirical value of 0.376% quite closely. Details of these calculations, as well as the tabulations of S_R^k , S_L^k , γ_R^k and γ_L^k necessary to support them, can be found in the electronic supplement.

A.2 Calculating the approximation $\text{vol}_0(\mathcal{R}_N^N)$ in closed form

In Section 5, a closed-form expression for $\text{vol}_0(\mathcal{R}_N^N)$ was given. Here, we show the derivation of this expression. Since the integrand clipping can be dropped by assumption, the integral (5.5) becomes

$$\text{vol}_0(\mathcal{R}_N^N) = \left(\int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} (I_R(w) - \min \mathcal{I}^N) - (I_L(w) - \max \mathcal{I}^N) dw_1 \cdots dw_{N-1} dw \right)^N$$

which is equivalent to

$$\left(\int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} I_W(w) + \text{range } \mathcal{I}^N dw_1 \cdots dw_{N-1} dw \right)^N$$

where $\text{range } \mathcal{I}^N = \max \mathcal{I}^N - \min \mathcal{I}^N = |w_1| + \cdots + |w_{N-1}|$. Thus, we can rewrite this integral as

$$\left(\int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} I_W(w) dw_1 \cdots dw_{N-1} dw + \int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} \sum_{i=1}^{N-1} |w_i| dw_1 \cdots dw_{N-1} dw \right)^N$$

The first term can be reduced to a 1-dimensional integral of $I_W(w)$ that can be evaluated explicitly to obtain

$$\begin{aligned}
& \int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} I_W(w) dw_1 \cdots dw_{N-1} dw \\
&= (2w_{\max})^{N-1} \int_4^{w_{\max}} I_W(w) dw \\
&= (2w_{\max})^{N-1} \frac{w_{\max} \left(\sqrt{w_{\max}(w_{\max}-4)} + \ln 256 \right) - 8(w_{\max}-1) \ln \left(\sqrt{w_{\max}-4} + \sqrt{w_{\max}} \right) + 2\sqrt{w_{\max}(w_{\max}-4)} - \ln 256}{2}
\end{aligned}$$

In the second term, we can eliminate the absolute value operations by splitting the integral into 2^{N-1} parts, all of which have the same value. Then we distribute the integrals over the sum and evaluate to obtain

$$\begin{aligned}
& \int_4^{w_{\max}} \int_{-w_{\max}}^{w_{\max}} \cdots \int_{-w_{\max}}^{w_{\max}} \sum_{i=1}^{N-1} |w_i| dw_1 \cdots dw_{N-1} dw = 2^{N-1} \int_4^{w_{\max}} \int_0^{w_{\max}} \cdots \int_0^{w_{\max}} \sum_{i=1}^{N-1} w_i dw_1 \cdots dw_{N-1} dw \\
&= 2^{N-1} \sum_{i=1}^{N-1} \int_4^{w_{\max}} \int_0^{w_{\max}} \cdots \int_0^{w_{\max}} w_i dw_1 \cdots dw_{N-1} dw \\
&= 2^{N-1} (N-1) (w_{\max}-4) (w_{\max})^{N-2} \frac{(w_{\max})^2}{2}
\end{aligned}$$

Raising the sum of the previous two expressions to the N th power and simplifying gives

$$\begin{aligned}
\text{vol}_0(\mathcal{R}_N^N) &= \left(2^{N-2} (w_{\max})^{N-1} \left(w_{\max} (N(w_{\max}-4) - w_{\max} + \sqrt{w_{\max}(w_{\max}-4)} + \ln 256 + 4) - \right. \right. \\
&\quad \left. \left. 8(w_{\max}-1) \ln \left(\sqrt{w_{\max}-4} + \sqrt{w_{\max}} \right) + 2\sqrt{w_{\max}(w_{\max}-4)} - 8 \ln 2 \right) \right)^N
\end{aligned}$$

A.3 Calculating $P(\mathcal{P}_9^2)$ for 2-neuron circuits

In Section 6.1, a value for $P(\mathcal{P}_9^2)$ was given. Here we show the derivation of this value.

Given $4 \leq w \leq 16$ and $-16 \leq w_1, \theta_1, \theta_2 \leq 16$, the integrand of (6.1) can be simplified to obtain

$$\left[[I_R(w) - \max(0, w_1)]_{-16}^{16} - [I_L(w) - \min(0, w_1)]_{-16}^{16} \right]_0 = \begin{cases} I_W(w) - w_1 & 0 \leq w_1 \leq I_W(w) \\ I_W(w) + w_1 & -I_W(w) \leq w_1 \leq 0 \end{cases}$$

We can then split the inner integral of (6.1) and evaluate to obtain

$$\begin{aligned}
P(\mathcal{P}_9^2) &= \frac{\left(\int_4^{16} \left(\int_{-I_w(w)}^0 I_w(w) + w_1 dw_1 + \int_0^{I_w(w)} I_w(w) - w_1 dw_1 \right) dw \right)^2}{32^4 32^2} \\
&= \frac{\left(\int_4^{16} I_w(w)^2 dw \right)^2}{32^6} \\
&= \frac{\left(1152 - 576\sqrt{3} \ln(2 + \sqrt{3}) + 240 \ln(2 + \sqrt{3})^2 \right)^2}{1073741824}
\end{aligned}$$

A.4 Calculating an approximation to the Hopf curve for 2-neuron circuits

In Section 6.2, approximations $H_1^\pm(\Delta\theta_2; \mathbf{W}, \boldsymbol{\tau})$ and $H_2^\pm(\Delta\theta_1; \mathbf{W}, \boldsymbol{\tau})$ to the Hopf bifurcation curve in 2-neuron circuits were utilized. Here we give the derivation of these expressions. Within a small perturbation $\Delta\boldsymbol{\theta}$ of a center-crossing circuit, we can replace $\sigma(x)$ with its linearization $\hat{\sigma}(x) = x/4 + 1/2$ in order to obtain the linearized dynamics $\boldsymbol{\tau}\dot{\mathbf{y}} = -\mathbf{y} + \mathbf{W} \cdot \hat{\boldsymbol{\sigma}}(\mathbf{y} + \boldsymbol{\theta}^* + \Delta\boldsymbol{\theta})$. By solving for the simultaneous zeroes of these equations, we find that the central equilibrium point occurs at $\bar{\mathbf{y}} = -\boldsymbol{\theta}^* + \Delta\bar{\mathbf{y}}$ where, for a 2-neuron circuit, we have

$$\begin{aligned}
\Delta\bar{y}_1 &= \frac{(4w_{11} - w_{11}w_{22} + w_{12}w_{21})\Delta\theta_1 + 4w_{12}\Delta\theta_2}{w_{11}w_{22} - 4w_{11} - 4w_{22} - w_{12}w_{21} + 16} \\
\Delta\bar{y}_2 &= \frac{(4w_{22} - w_{11}w_{22} + w_{12}w_{21})\Delta\theta_2 + 4w_{21}\Delta\theta_1}{w_{11}w_{22} - 4w_{11} - 4w_{22} - w_{12}w_{21} + 16}
\end{aligned}$$

If $\hat{\mathbf{J}}$ is the Jacobian of the linearized system evaluated at $\bar{\mathbf{y}}$, then the Hopf bifurcation condition is given by

$$\det(2\hat{\mathbf{J}} \odot \mathbf{1}) = \frac{w_{11}\hat{\sigma}'(\bar{y}_1 + \theta_1^* + \Delta\theta_1) - 1}{\tau_1} + \frac{w_{22}\hat{\sigma}'(\bar{y}_2 + \theta_2^* + \Delta\theta_2) - 1}{\tau_2} = 0$$

where $\hat{\sigma}'(x) = 1/4 - x^2/16$ is the quadratic approximation to $\sigma'(x)$.

Substituting $\bar{y}_i = -\theta_i^* + \Delta\bar{y}_i$, solving for $\Delta\theta_1$ and $\Delta\theta_2$ and simplifying, we obtain

$$H_1^\pm(\Delta\theta_2; \mathbf{W}, \boldsymbol{\tau}) = \frac{\alpha\Delta\theta_2 \pm \sqrt{\beta(\chi(\Delta\theta_2)^2 + \kappa\eta_1)}}{2\eta_1}, \quad H_2^\pm(\Delta\theta_1; \mathbf{W}, \boldsymbol{\tau}) = \frac{\alpha\Delta\theta_1 \pm \sqrt{\beta(\chi(\Delta\theta_1)^2 + \kappa\eta_2)}}{2\eta_2}$$

where

$$\begin{aligned}
\alpha &= 2w_{22}w_{21}\tau_1(w_{11}-4) + 2w_{11}w_{12}\tau_2(w_{22}-4) \\
\beta &= (w_{12}w_{21} + 4(w_{22}-4) - w_{11}(w_{22}-4))^2 \\
\chi &= -4w_{11}w_{22}\tau_1\tau_2 \\
\kappa &= (\tau_2(w_{11}-4) + \tau_1(w_{22}-4)) \\
\eta_1 &= w_{11}\tau_2(w_{22}-4)^2 + w_{22}\tau_1(w_{21})^2 \\
\eta_2 &= w_{22}\tau_1(w_{11}-4)^2 + w_{11}\tau_2(w_{12})^2
\end{aligned}$$

Note that these H expressions are only real-valued when $\beta(\chi(\Delta\theta_2)^2 + \kappa\eta_1) > 0$, $\beta(\chi(\Delta\theta_1)^2 + \kappa\eta_2) \geq 0$ and that they have singularities at $\eta_1 = 0$ and $\eta_2 = 0$. The last two conditions in the domain of integration \mathcal{D} given in the main text arise from the requirement that $\kappa\eta_1, \kappa\eta_2 \geq 0$ for the H functions to be real-valued when $\Delta\theta_1 = \Delta\theta_2 = 0$, since β is strictly positive.

Figure 1

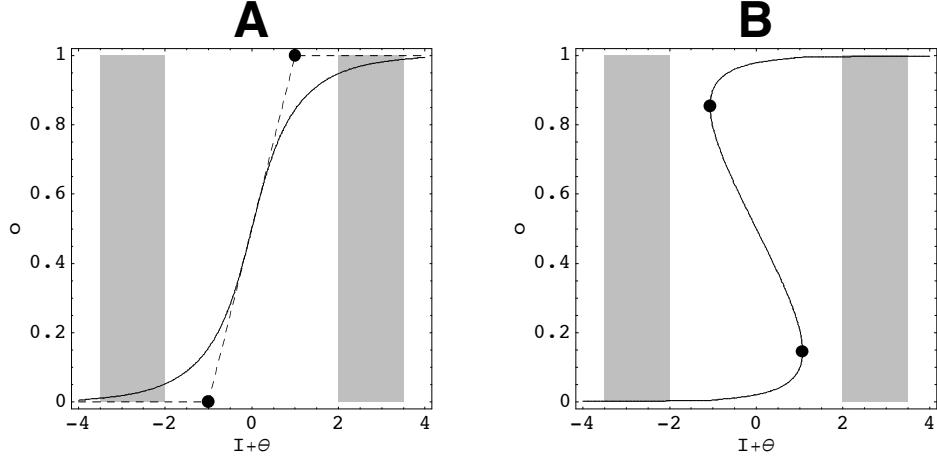


Figure 1: Representative steady-state input/output (SSIO) diagrams of a single CTRNN for (A) $w = 2$ and (B) $w = 8$. The solid line shows the output space location of the neuron's equilibrium points as a function of the net input $I+\theta$. Note that the SSIO becomes folded for $w > 4$, indicating the existence of three equilibrium points. When the SSIO is folded, the left and right edges of the fold are given by $I_L(w)$ and $I_R(w)$, respectively (black points in B). The ranges of synaptic inputs received from other neurons are indicated by gray rectangles. The lower (min \mathcal{I}) and upper (max \mathcal{I}) limits of this range play an important role in the analysis described in this paper. In both plots, two synaptic input ranges are shown; one for which the neuron is saturated off (left rectangle) and one for which the neuron is saturated on (right rectangle). The dashed line in A shows the piecewise linear SSIO approximation utilized in Section 4.2, which suggests using the intersections of the linear pieces (black points) as the analogues of the fold edges in part B.

Figure 2

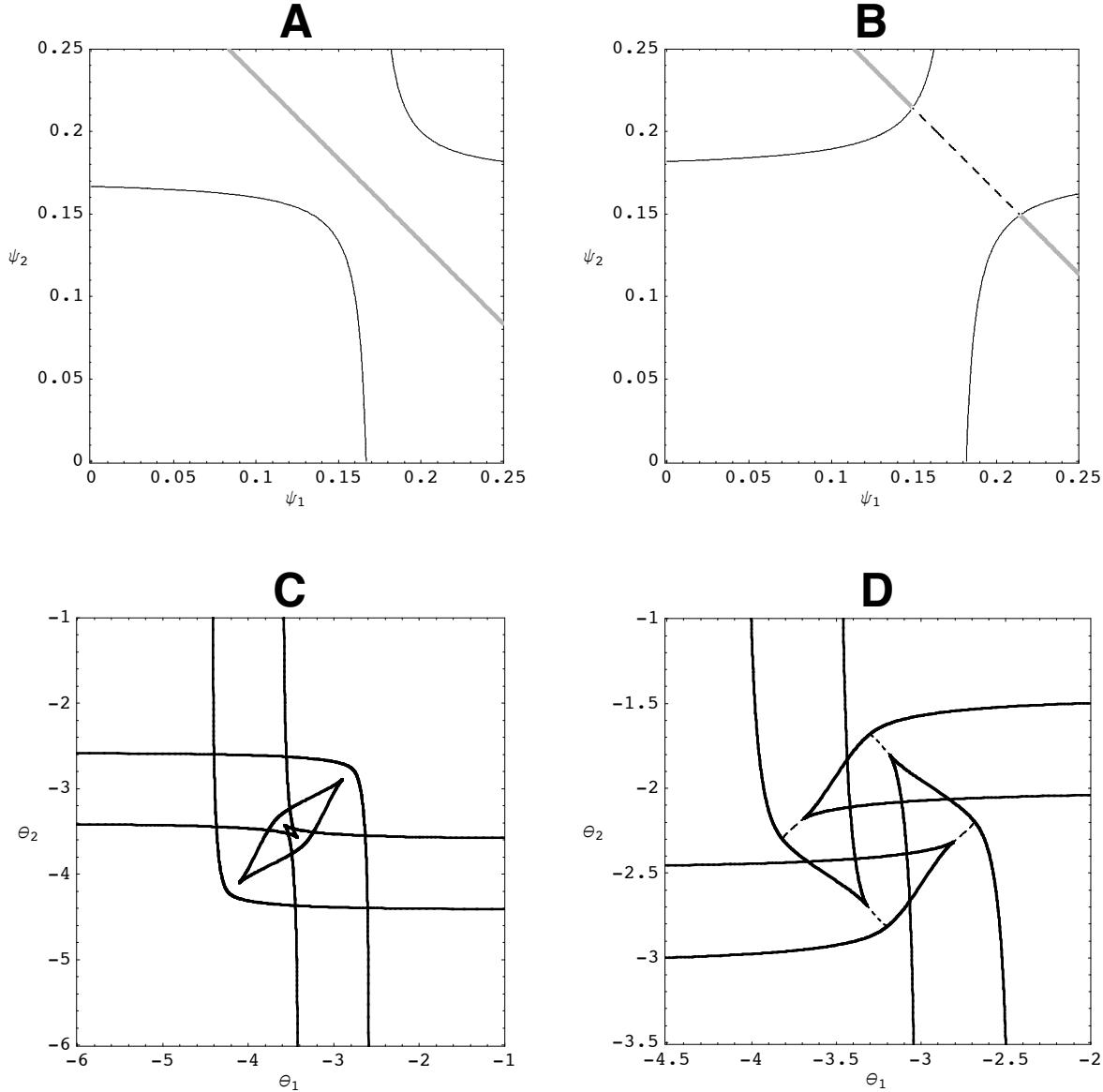


Figure 2: Representative local bifurcation curves of 2-neuron CTRNNs. In all cases, solid black curves represent saddle-node bifurcations, dashed curves represent Hopf bifurcations, and solid gray curves represent neutral saddles. (A) Bifurcation curves in the space (ψ_1, ψ_2) of activation function derivatives for a circuit with $\mathbf{W} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$ and $\tau_1 = \tau_2 = 1$. (B) Bifurcation curves in (ψ_1, ψ_2) for a circuit with $\mathbf{W} = \begin{pmatrix} 5.5 & 1 \\ -1 & 5.5 \end{pmatrix}$ and $\tau_1 = \tau_2 = 1$. (C) The bifurcation curves from Part A in net input space (neutral saddle curves have been removed). Note that each curve in (ψ_1, ψ_2) can produce multiple curves in (θ_1, θ_2) . (D) The bifurcation curves from Part B in net input space (neutral saddle curves have been removed).

Figure 3

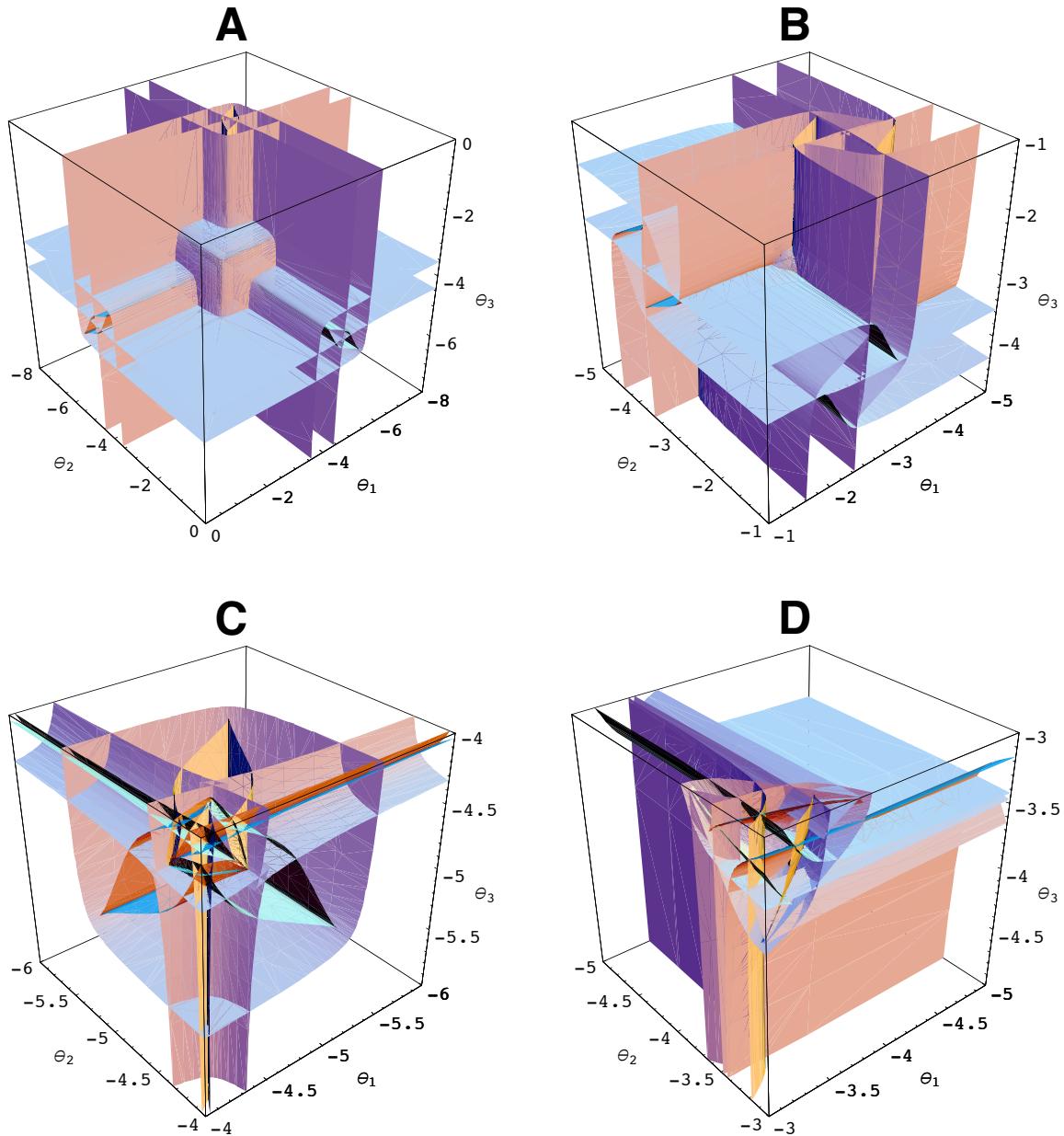
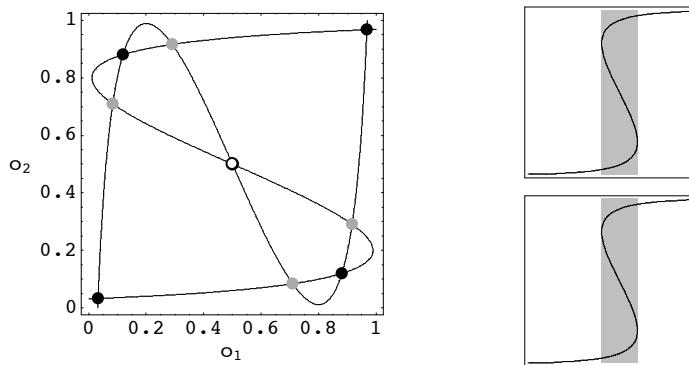


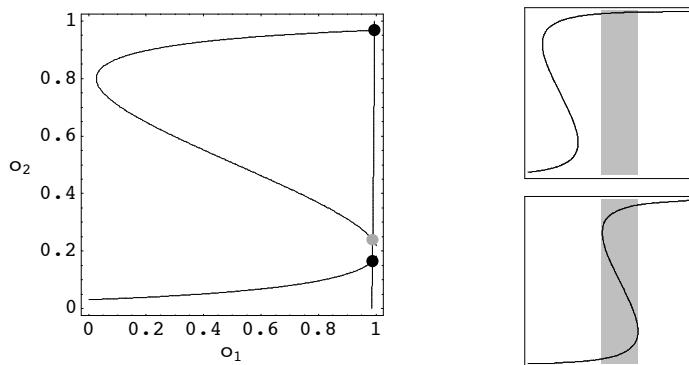
Figure 3: Representative local bifurcation surfaces of 3-neuron CTRNNs in net input space. (A) A circuit with $\mathbf{W} = \begin{pmatrix} 6 & 1 & 1 \\ 1 & 6 & 1 \\ 1 & 1 & 6 \end{pmatrix}$ and $\tau_i = 1$. (B) A circuit with $\mathbf{W} = \begin{pmatrix} 6 & -1 & 1 \\ 1 & 6 & -1 \\ -1 & 1 & 6 \end{pmatrix}$ and $\tau_i = 1$. (C) A cut-away view of the interior of the plot in Part A. (D) A cut-away view of the interior of the plot in Part B. Note that the bifurcation manifolds in the space (ψ_1, ψ_2, ψ_3) of activation function derivatives are not shown in this figure and saddle-node and Hopf bifurcation surfaces are not distinguished in these plots.

Figure 4

A



B



C

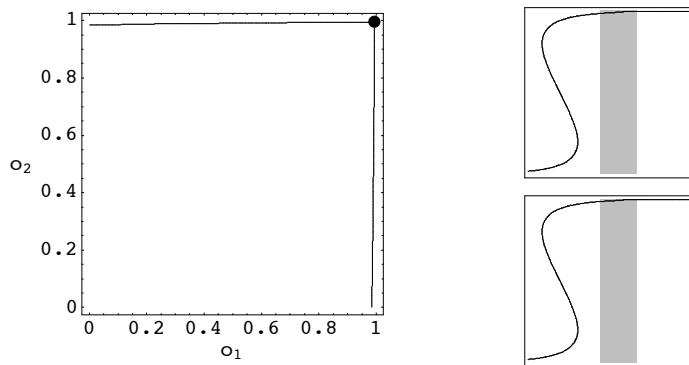


Figure 4: Phase portraits and SSIO diagrams of 2-neuron CTRNNs with dynamics of different effective dimension. In the phase portrait plots, stable equilibrium points are shown in black, saddle points are shown in gray, unstable equilibrium points are shown as circles and the nullclines are shown as black curves. In the SSIO plots, the range of synaptic input that each neuron receives from the other is indicated by a gray rectangle. (A) A phase portrait with an effective dimension of 2. Note that the folds of both neurons intersect their range of synaptic inputs. (B) A phase portrait with an effective dimension of 1. Note that neuron 1 is saturated on, confining the dynamics to the right edge of output space after transients have passed. (C) A phase portrait with an effective dimension of 0. Note that both neurons are saturated on, confining the dynamics to the upper righthand corner of output space after transients have passed.

Figure 5

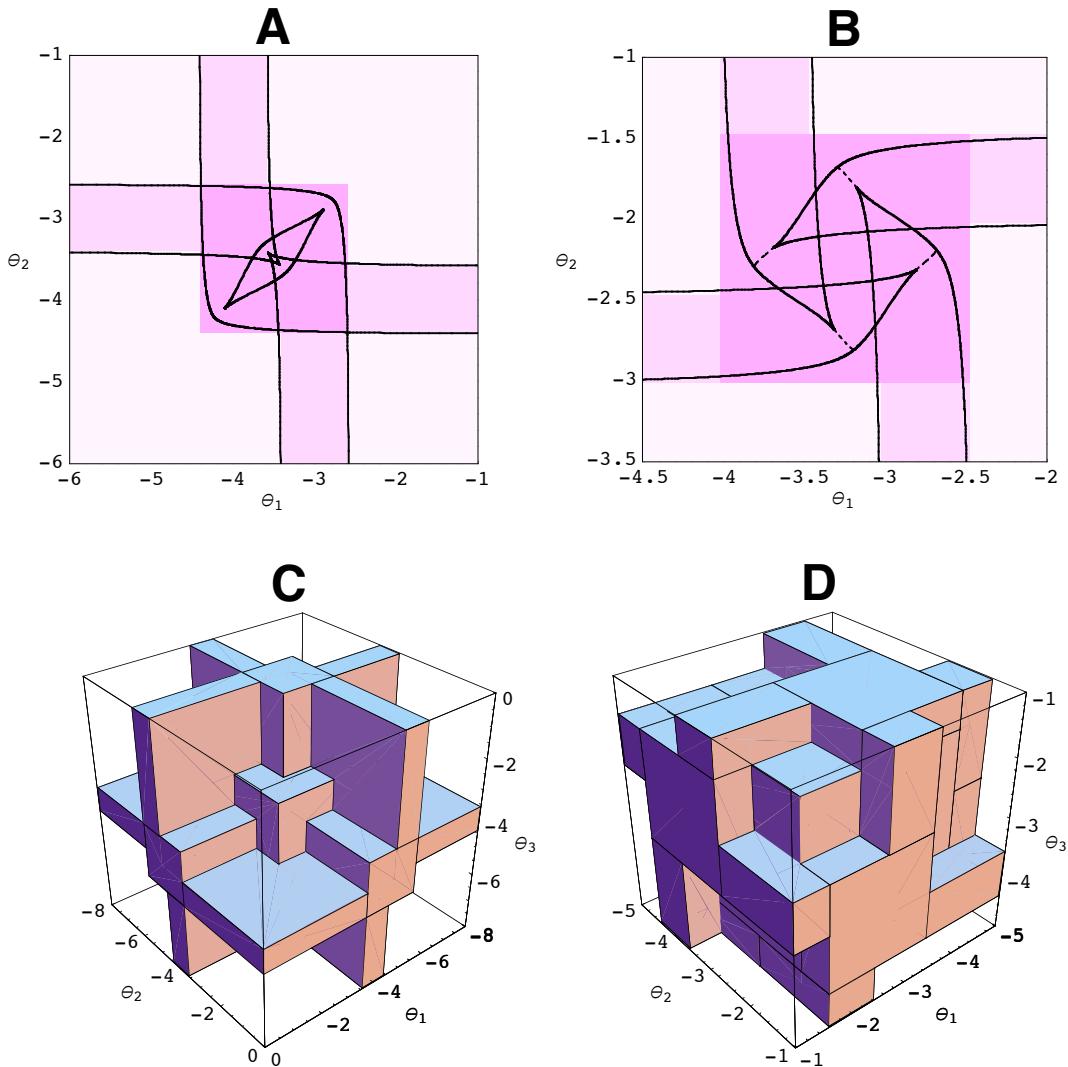


Figure 5: Asymptotic hyperplane approximations to the extremal saddle-node bifurcation manifolds shown in Figures 2 and 3. (A) Region approximations corresponding to Figure 2C. The darkest square is \mathcal{R}_2^2 , the lighter rectangles are \mathcal{R}_1^2 and the lightest regions are \mathcal{R}_0^2 . (B) Region approximations corresponding to Figure 2D. (C) Region approximations corresponding to Figure 3A. (D) Region approximations corresponding to Figure 3B.

Figure 6

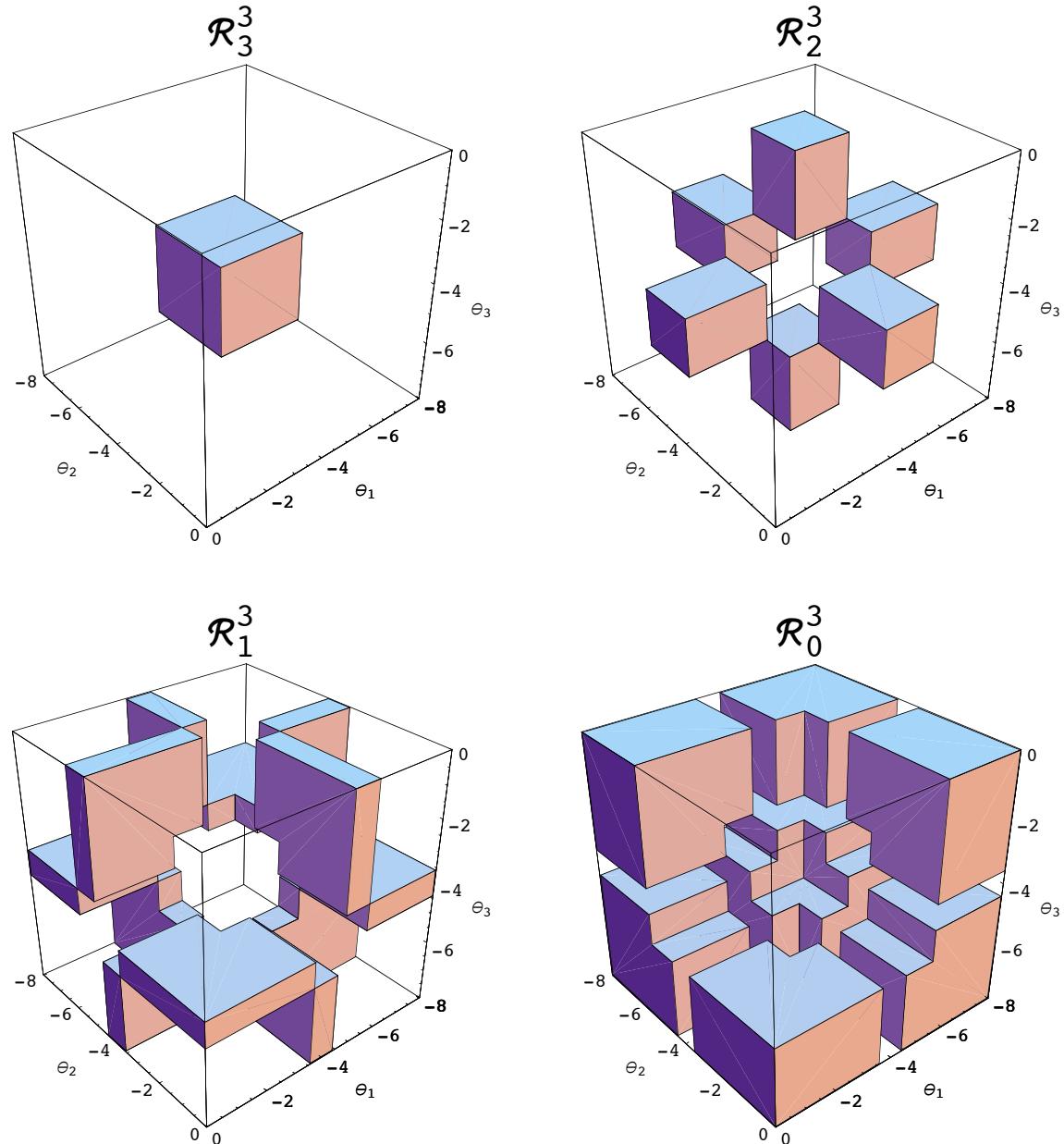


Figure 6: The individual components of the region approximation shown in Figure 5C. Note that the frontmost polytope ${}^3Q^{123}$ has been removed from \mathcal{R}_0^3 in order to make the interior visible.

Figure 7

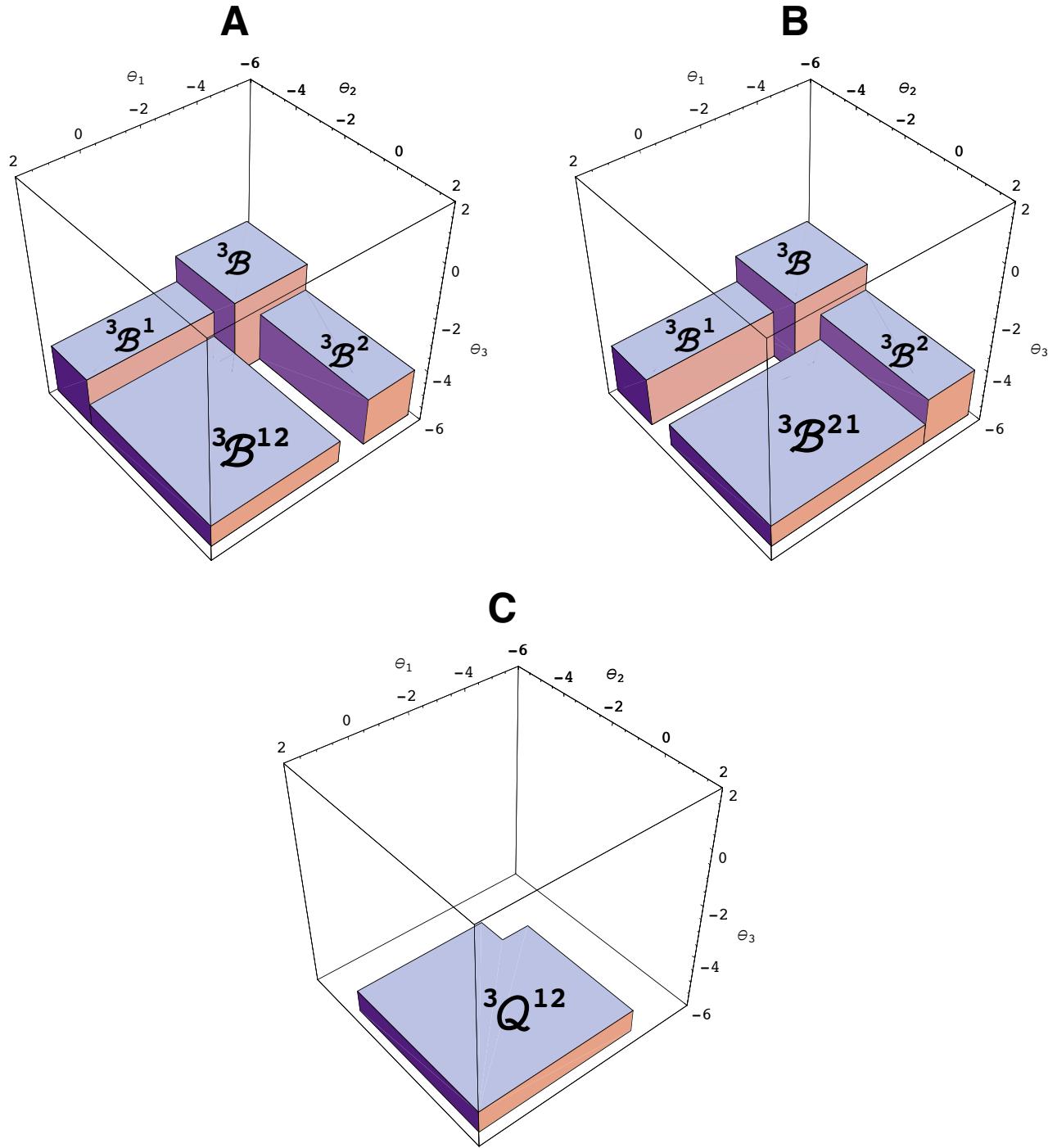


Figure 7: An illustration of the structure of Q and \mathcal{B} regions. ${}^3\mathcal{B}^{12}$ (A) and ${}^3\mathcal{B}^{21}$ (B) differ in the order in which their defining inequalities are constructed. (C) The nonconvex polytope ${}^3Q^{12}$ is formed by the union of the rectangular solids ${}^3\mathcal{B}^{12}$ and ${}^3\mathcal{B}^{21}$.

Figure 8

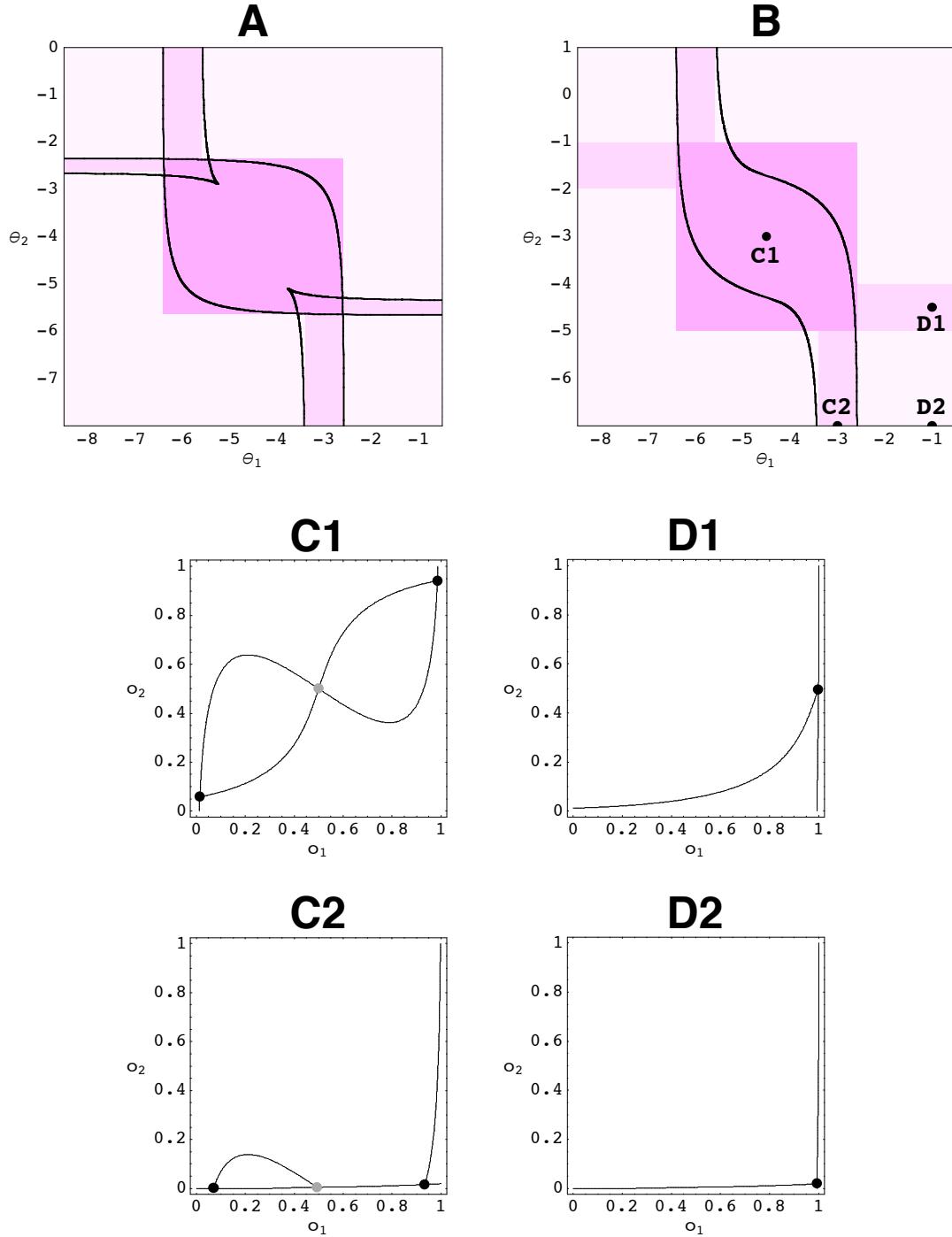


Figure 8: An illustration of the extension of the definition of \mathcal{R} to CTRNNs containing self-weights less than 4. (A) Local bifurcation curves and region approximations for a 2-neuron circuit with $\mathbf{W} = \begin{pmatrix} 6 & 3 \\ 3 & 5 \end{pmatrix}$. (B) The same circuit as A, but with $w_{22} = 3$. Note that the saddle-node bifurcations involving the bistability of neuron 2 have now disappeared, since $w_{22} < 4$. However, the extended definition $\tilde{\mathcal{R}}$ can still be used to calculate the regions shown. C1, C2, D1 and D2 show the phase portraits at the four indicated points in Part B, with the nullclines drawn as black curves, stable equilibrium points colored blue (or black), and saddle points colored green (or gray). Although no saddle-node bifurcations separate C1 from C2 or D1 from D2 in this circuit, these regions do differ in their effective dimensionality.

Figure 9

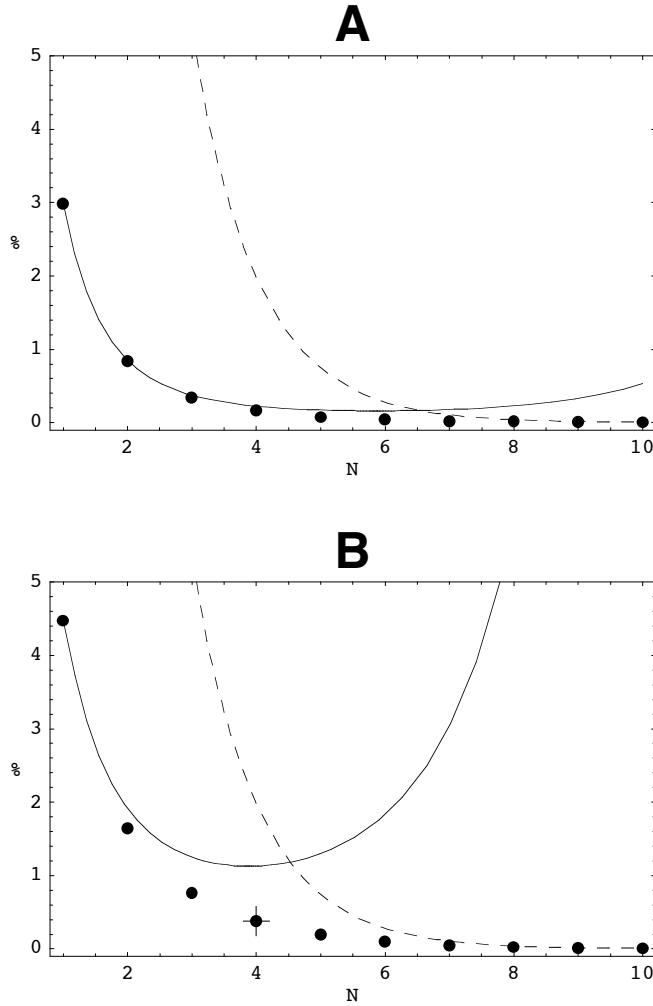


Figure 9: Plots of the $P(\mathcal{R}_N^N)$ approximations $P_0(\mathcal{R}_N^N)$ (solid curve) and $P_\infty(\mathcal{R}_N^N)$ (dashed curve) compared to data (black points) from 10^6 random parameter space samples at each N , with $w_i \in [-16, 16]$ and $\theta_i \in [-24, 24]$ (A) or $\theta_i \in [-16, 16]$ (B). Note that P_0 is most accurate for small N , while P_∞ is most accurate for large N . The maximum error occurs at the crossover point between the two approximations, and can be significant for smaller θ ranges. The crosshairs at $N = 4$ in Part B corresponds to the exact value calculated in Appendix A.1.

Figure 10

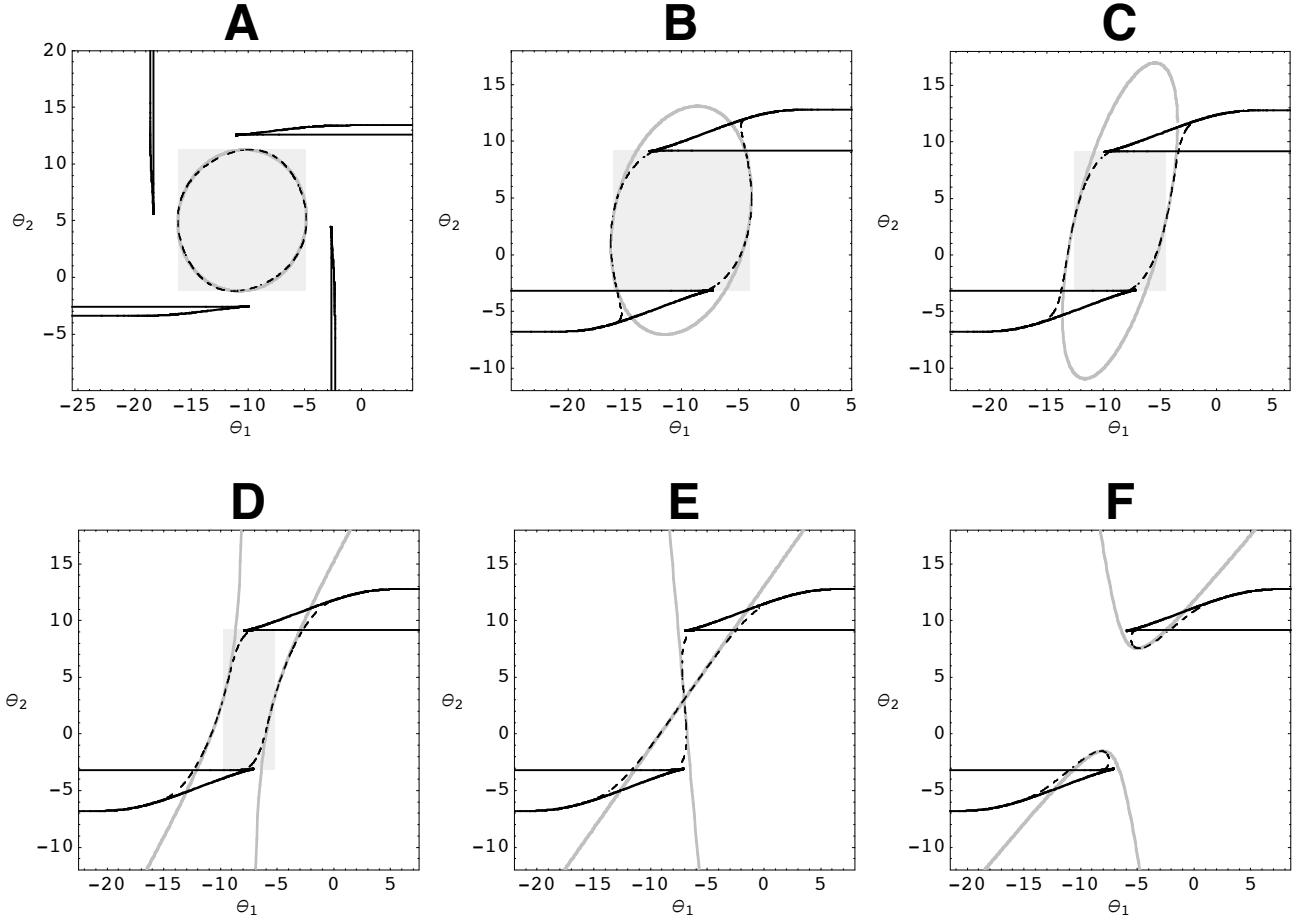


Figure 10: An illustration of an approximation to the region of central oscillation phase portraits P^2_{1LC} in 2-neuron CTRNNs. The Hopf bifurcation curve approximations $\theta^* + H^\pm(\Delta\theta; \mathbf{W}, \tau)$ are shown as solid gray curves. The gray rectangles approximate the oscillatory regions themselves. These rectangles are defined by $[\theta_1^* + H_1^\pm(0; \mathbf{W}, \tau)] \times [\theta_2^* + H_2^\pm(0; \mathbf{W}, \tau)]$ clipped to the saddle-node bifurcation curves. Here $\tau_1 = \tau_2 = 1$. In Part A, $\mathbf{W} = \begin{pmatrix} 5 & 16 \\ -16 & 6 \end{pmatrix}$ and in Parts B-F, $\mathbf{W} = \begin{pmatrix} w_{11} & 16 \\ -16 & 10 \end{pmatrix}$, with $w_{11} = 4$ (B), $w_{11} = 1$ (C), $w_{11} = -1$ (D), $w_{11} = -2$ (E) and $w_{11} = -3$ (F). Note that, although no central oscillations exist in Parts E and F, $\theta^* + H^\pm(\Delta\theta; \mathbf{W}, \tau)$ still approximate the Hopf bounds of regions of noncentral oscillations. A more accurate nonrectangular approximation to the oscillatory region would be able to take advantage of this fact.

Figure 11

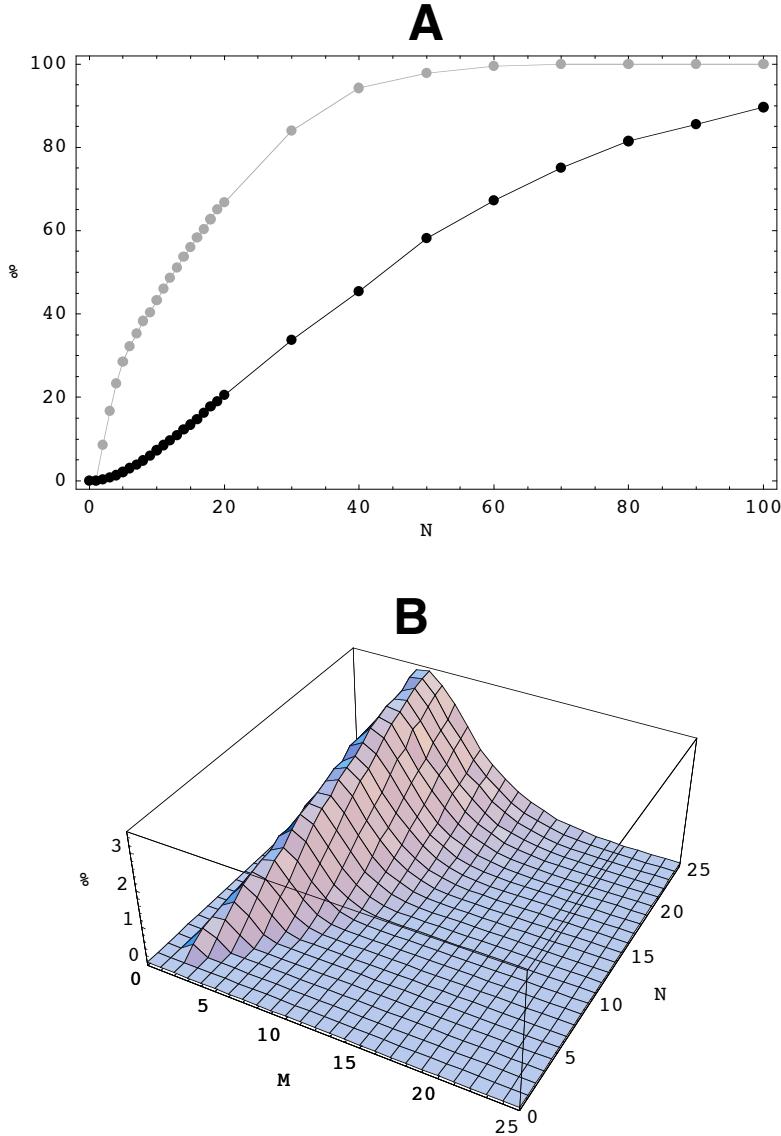


Figure 11: Empirical studies of oscillation probability in CTRNNs. (A) The probability $P(O^N)$ of observing "oscillatory" (nonequilibrium) dynamics in general (black) and center-crossing (gray) N -neuron circuits with 10^5 ($N \leq 20$) or 10^4 ($N > 20$) random samples from $w_i, \theta_i \in [-16, 16]$, $\tau_i \in [0.5, 10]$. Note that, although oscillations are clearly much more common in center-crossing networks, both curves increase with N . Samples were obtained by randomly generating 10 initial conditions in the range $y_i \in [-16, 16]$, integrating each with the forward Euler method for 2500 integration steps of size 0.1 to skip transients, and then integrating for an additional 500 integration steps. If the output of any neuron varied by more than 0.05 during this second integration for any initial condition, then the circuit was classified as oscillatory. (C) The probability $P(O_M^N)$ of observing M -dimensional oscillatory dynamics in an N -neuron circuit, using 10^5 parameter space samples for each N and the same sampling protocol as in Part A. Note that the distribution shifts to the right, broadens and rises with increasing N .