Physica D 470 (2024) 134358

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Arbitrary sensitive transitions in recurrent neural networks

Muhammed Fadera *, Peter Ashwin

Department of Mathematics and Statistics, University of Exeter, Exeter, EX4 4QF, Devon, UK

ARTICLE INFO

Communicated by Boumediene Hamzi

MSC:
34C23
68T07

Keywords:
Interpretability
Excitable network attractor
Neural network
Almost complete

ABSTRACT

An Excitable Network Attractor (ENA) is a forward-invariant set in phase space that can be used to explain input-driven behaviour of Recurrent Neural Networks (RNNs) trained on tasks involving switching between a discrete set of states. An ENA is composed of two or more attractors and excitable connections that allow transitions from one attractor to another under some input perturbation. The smallest such perturbation that makes a connection between two attractors is called the excitability threshold associated with that connection. The excitability threshold provides a measure of sensitivity of the connection to input perturbations. Errors in performance of such trained RNNs can be related to errors in transitions around the associated ENA. Previous work has demonstrated that ENAs of arbitrary sensitivity and structure can be realised in a RNN by suitable choice of connection weights and nonlinear activation function. In this paper we show that ENAs of arbitrary sensitivity and structure can be realised even using a suitable fixed nonlinear activation function, i.e. by suitable choice of weights only. We show that there is a choice of weights such that the probability of erroneous transitions is very small.

1. Introduction

Recurrent Neural Networks are a class of machine learning models suitable for learning sequential tasks. They have been successfully used in chaotic time series forecasting (see [1] for a recent review), adaptive filtering [2,3], pattern generation [4] and classification to name a few. This flexibility of RNNs is not surprising as they have been shown to be universal approximators — they can approximate, to any degree of accuracy, any smooth bounded dynamical system on a compact time interval [5,6]. However, like most machine learning models, RNNs are hard to interpret and are usually considered as black box models. This means it is hard to understand how RNNs make decisions. Due to an increasing societal need to make machine learning models more transparent [7], opening this black box may increase trust in RNNs and allow for their flexibility to be taken advantage of in high-stakes applications such as medicine [8] and policy making [9].

As RNNs become larger and increasingly more sophisticated, fine-grained analysis of their behaviour becomes computationally expensive. Since RNNs are input-driven dynamical systems, their input-driven behaviour can be explained and interpreted using nonlinear dynamics. Previous research — most notably the works of Beer [10,11], Hopfield [12] and more recently Steinberg and Sompolinsky [13] – have highlighted that neural networks in general use lower dimensional attractors to encode information. Furthermore, recent research [14,15] have shown that the computation in echo state networks (ESNs), an important class of RNNs, performing tasks involving learning transitions between finite set of states can be interpreted as a set of attractors – encoding the states – and transitions between them. Such dynamical system structures are characteristics of classification problems in artificial neural networks and by extension cognitive task that involves different modes of computations [16,17]. An ENA is composed of a number of attractors representing these computational states/modes and input-driven (excitable) connections between them. With a suitable notion of distance, ENA can be formed between any type of attractors e.g stable periodic orbits or strange attractors [18,19]. We say there is an excitable connection (or connection) between two attractors if perturbations larger than some threshold δ_{th} in some phase space direction in the vicinity of one of the attractors makes a "jump" to the basin of attraction of another. The smallest such perturbation is called the excitability threshold.

Using tools from nonlinear dynamics, the authors of [15] extracted an Excitable Network Attractors from the dynamics of a trained Echo State Network (ESN) on a simple benchmark tasks that involves switching between four states. ESN are an important class of RNNs that are able to circumvent the famous vanishing/exploding [20,21] gradient problem by focusing training only in the output/readout layer. The results of [15] showed that Excitable Network Attractors are effective in capturing how ESN solves classification problems. The extracted ENA is also able to explain the emergence of structural errors in imperfectly trained systems beyond what performance measures like mean squared

* Corresponding author.
E-mail address: M.Fadera@exeter.ac.uk (M. Fadera).

https://doi.org/10.1016/j.physd.2024.134358
Received 25 January 2024; Received in revised form 9 July 2024; Accepted 5 September 2024
Available online 11 September 2024

error can reveal. These errors usually manifest themselves as transitions outside the set of the required number of attractors needed to solve the task. Furthermore, other properties of the originally trained system such as robustness to noise and the echo index [22] can be deduced from related properties of the extracted ENA.

In part, the results of [14,15] can be explained by the fact that ENAs are models of sequential time-dependent finite state computations [16, 23–25]. Furthermore, the ability of ENAs to capture finite state computations are ground in theoretical research in language recognition and theory of computation [26]. For example, [24] showed that for an RNN to reliably capture a discrete finite state computation, it must divide the phase space into n distinct closed invariant sets where n is at least the number of different states and learn transitions between them. Surprisingly, due to finite precision arithmetic of modern computers, the results of [24] does not exclude the possibility of interpreting any computation in RNNs using Network Attractors, even those trained on regression tasks.

We consider the question of realising arbitrary sensitive transitions between finite sets of states as an Excitable Network Attractor in a Continuous Time Recurrent Neural Network (CTRNN). These transitions are represented as a directed graph with vertices representing equilibria and edges representing allowable transitions between. The input-driven dynamics of a CTRNN with N hidden units is given by the system of ordinary differential equations (ODEs)

$$\tau_i \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} w_{ij}\phi(x_j) + u_i(t) \quad \text{for } i = 1, \dots, N \tag{1}$$

where $W = [w_{ij}] \in \mathbb{R}^{N\times N}$ is the recurrent matrix between the hidden units $x_i \in \mathbb{R}$, $\tau_i > 0$ are time constants, ϕ the activation function and input $u_i(t)$ is the input to the ith hidden unit.

Results of [23] has shown that a large class of graphs can be realised as ENA at arbitrarily small sensitivity in a CTRNN by choosing both the weights and the activation function. In the realisation in [23], the activation function is piecewise linear and depends on the sensitivity of the network to be realised. Here, we extend these results to smooth activation functions and showed that ENAs of arbitrary sensitivity and graph topology can be realised in the autonomous dynamics of a CTRNN by choosing only the recurrent matrix. The activation function can be chosen independently of the sensitivity of the network to input perturbations. We also show that this realisation can be made "almost complete" in the sense that the number of transitions that do not tend to some equilibrium in the network through an allowable transition is small. We also give an upper bound on the dimension of CTRNN with respect to the number of vertices in the graph of the ENA. The realisation allows a given connection to be made arbitrarily sensitive to input perturbation. Our analysis showed that cycles in the graph can be realised as periodic orbits in phase space allowing transitions between periodic orbits and between periodic orbits and equilibria.

2. Background

An Excitable Network Attractor consist of a set of equilibria and parts of their basin of attraction that allow input-driven transitions between them. ENAs can be defined for both flows and maps. Here we give the definition for flows followed from [23,27]. In what follows, we assume that ξ_i is an equilibrium of the flow

$$\dot{x} = f(x,t)$$

where $x \in \mathbb{R}^N$ and f is at least C^1 in x.

Definition 2.1. An excitable connection of amplitude $\delta > 0$ is said to exist from equilibrium ξ_i to another equilibrium ξ_j if

$$B_\delta(\xi_i) \cap W^s(\xi_j) \neq \emptyset.$$

where $B_\delta(\xi_i)$ is an open ball of radius δ around ξ_i and $W^s(\xi_j)$ is the basin of attraction of the equilibrium ξ_j.

Definition 2.2. The quantity

$$\delta_{th}^{ij} = \inf\{\delta > 0 : B_\delta(\xi_i) \cap W^s(\xi_j) \neq \emptyset\}$$

is called the excitability threshold of the excitable connection from ξ_i to ξ_j.

Definition 2.3. An **Excitable Network Attractor** at amplitude δ between a set of equilibria $\{\xi_i\}_{i=1}^N$ is defined as the set

$$\Sigma = \bigcup_{i,j} \{\Phi_t(x) : x \in B_\delta(\xi_i), t > 0\} \cap W^s(\xi_j).$$

where $\Phi_t(x)$ is the solution at time t associated with the initial condition x.

Remark. In the above definitions, the equilibria ξ_i will mostly be taken to be a sink for each i. However, this is not necessary. For example excitable connections starting from saddles can be considered as having an excitability threshold of 0.

There is a natural way to encode vertices of a directed graph as the set of equilibria in Σ that have excitable connections between them. Consider the directed graph G with vertices numbered from 1 to N. Identifying vertex k of G with equilibrium ξ_k of Σ, there is an edge from vertex i to vertex j of G if $B_\delta(\xi_i) \cap W^s(\xi_j) \neq \emptyset$.

Definition 2.4. We say $A \in \mathbb{R}^{N\times N}$ is the *adjacency matrix* of a directed graph G if $a_{ij} = 1$ if and only if there is a directed edge from vertex i to vertex j and $a_{ij} = 0$ otherwise. We say an adjacency matrix is *admissible* if the corresponding graph contains no 1-cycles, 2-cycles or Δ-cliques (see Fig. 1).

Definition 2.5. Let Σ be an Excitable Network Attractor of amplitude δ between the set of equilibria $\{\xi_i\}$. Let G be a directed graph with an associated adjacency matrix A. Σ is said to realise G at amplitude δ if and only if $a_{ij} = 1$ whenever

$$B_\delta(\xi_i) \cap W^s(\xi_j) \neq \emptyset.$$

G is referred to as the graph of Σ.

We can also consider G as a weighted directed graph where each edge in G is weighted with the corresponding excitability threshold δ_{th}^{ij}. The matrix $D \in \mathbb{R}^{N\times N}$ of edge weights of G has components $D_{ij} = \delta_{\text{th}}^{ij}$ if $a_{ij} = 1$ and 0 otherwise. When $\delta_{\text{th}}^{ij} = 0$ but $a_{ij} = 1$, then their is a connection from ξ_i to ξ_j and thus transitions can happen without inputs. These transitions are called *spontaneous transitions* and maybe due to a heteroclinic connection from ξ_i to ξ_j. Another example of transitions other than excitable transition is *spurious transitions*. A spurious transition occur when there is a transition from $\xi_i \in \Sigma$ to equilibria or other attractors outside Σ. If spurious transitions can be avoided, then all initial conditions explore the network following a path on the directed graph G. Such a realisation of G is said to be *complete*. In the case that the set of all initial conditions $x \in \Sigma$ that results in spurious transitions is a measure zero set, the realisation of G is said to be *almost complete*. This is a generalisation of the concept of almost complete realisation for heteroclinic networks from [28] to excitable network attractors.

Definition 2.6. Let $A \in \mathbb{R}^{N\times N}$ be a given adjacency matrix with G as the associated directed graph and $\delta > 0$. Suppose Σ is a realisation of G as an ENA at amplitude δ between attracting equilibria ξ_k. Let the equilibrium $\xi_k \in \Sigma$ be identified with vertex k of the graph G. Then the node ξ_k is said to be **almost complete** if

$$\mu(\{x \in B_\delta(\xi_k) |\ x \notin W^s(\xi_k) \text{ and } \forall l \text{ with } a_{kl} = 1,\ x \notin W^s(\xi_l)\}) = 0$$

where μ is Lebesgue measure on $\mathbb{R}^N$. Σ is said to be an **almost complete realisation** if for all $\xi_k \in \Sigma$, ξ_k is almost complete.

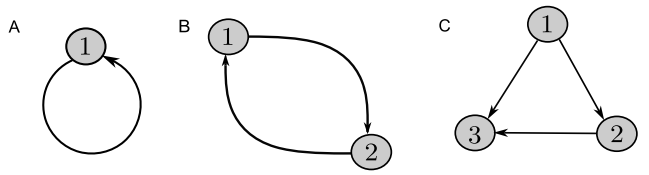

Fig. 1. From left to right: graph of 1 cycle, 2 cycle and Δ-clique. A directed graph that does not have any of these structures as a subgraph is referred to as admissible.

There is a key difference between the above definition for ENAs of almost completeness and that given in [28] for heteroclinic networks. The definition in [28] uses Lebesgue measure on the unstable manifolds of a saddle in the heteroclinic network; we make statements about some finite neighbourhood of stable equilibrium with respect to Lebesgue measure on the full space (i.e $\mathbb{R}^N$) and will depend on neighbourhood size. This is most appropriate when all ξ_i are sinks.

The input-driven dynamics of ENAs are intermittent with trajectories spending long times near one equilibrium before switching to another in a short time. This provides a natural way to describe the dynamics of the nodes of Σ using the relationship between the corresponding vertex and its neighbours in G. For example if a realisation is almost complete, the ordered sequence of nodes visited by a given input/noise-driven trajectory is known as its *itinerary*. The *itinerary* provides a natural symbolic dynamics with respect to the network associated with a given input/noise driving the network.

Definition 2.7. Let A be the adjacency matrix of a directed graph G associated with the ENA Σ. Let $\xi_k \in \Sigma$. For this equilibrium we defined the leading nodes $L(k) := \{i : a_{ki} = 1\}$, the trailing nodes $T(k) := \{i : a_{ik} = 1\}$ and the disconnected nodes $D(k) := \{i : a_{ik} = a_{ki} = 0\}$.

In terms of transitions between nodes of Σ, $L(k)$ contains the indices of the set of nodes accessible from ξ_k through an excitable transition while $T(k)$ contain the indices of the set of nodes that can make transition to ξ_k. Note that for graphs with admissible adjacency matrix this partition covers all cases.

The remainder of this paper is organised as follows. In Section 3, we explain previous attempts to realise directed graphs as excitable networks attractors in the dynamics of the CTRNN (1) by choosing both the weight and the activation function to depend on the amplitude of the ENA [23]. We discuss how this is useful in understanding some of the global bifurcations that emerge as the amplitude changes. Using properties of general sigmoidal functions explained in the Appendix A, we prove how these realisations can be extended to a smooth activation function by choosing the weight matrix W to depend on four parameters and properties of the adjacency matrix A. We use the fact that the equilibria that are needed for this realisation when $\delta > 0$ are all hyperbolic to remark that the this realisation holds for an open sets of parameters in $\mathbb{R}^4$. In Section 3.1, we explain two possible ways of making the realisation in Theorem 3.1 almost complete and gave a proof for one of these methods.

In Section 4, we prove that it is possible to choose the weight matrix W to be invertible. This allows us to extend the results of Theorem 3.1 through topological equivalence to an alternative formulation of the CTRNN model used in the machine learning literature. Section 5 explains the bifurcations that are possible at $\delta = 0$. Using bifurcation analysis and numerical simulations, we showed that the results of Theorem 3.1 can be extended to all one-cycle free graphs and provide a bound on the dimension of the CTRNN needed for this realisations. We explain how these bifurcations can turn cycles in the graph into corresponding periodic orbits and thus allowing realisation of excitable network of periodic orbits. We conclude the paper with possible applications and future directions of the ENA model developed here.

3. Realisation of graphs as ENAs

Ashwin and Postlethwaite [23], described a method of realising graphs with admissible adjacency matrices as Excitable Network Attractors between stable equilibria in the input-free dynamics of a CTRNN. These continuous time analogues of conventional discrete time recurrent neural networks are used in theoretical neuroscience as a model for episodic memory [29] or as minimal models for cognition [11], as well as in evolutionary computing [10,30] and robotics [31,32]. Although CTRNNs are not as popular as their discrete counterparts, their power to model continually evolving systems has recently gained traction in the machine learning community through work on liquid time-constant neural networks [31,32].

The (deterministic) input-driven dynamics of a CTRNN with N hidden units is described by the system of ODEs in (1). The system may also be driven with Wiener noise $W_i(t)$ with the following stochastic differential equation

$$\tau_i dx_i = \left(-x_i + \sum_{j=1}^{N} w_{ij}\phi(x_j)\right) dt + s^{\text{noise}} dW_i(t) \tag{2}$$

where $W_i(t)$ is independent Wiener noise with noise amplitude s^{noise}.

Assuming that A has no 1 cycles, 2 cycles or Δ-cliques (see Fig. 1), the authors of [23] showed that Eq. (1) realises A as an ENA Σ at amplitude δ by choosing the activation function to be ϕ_P with

$$\phi_P(x) = \begin{cases} 0 & x - \theta < -2\varepsilon, \\ 1/2 + (x-\theta)/(4\varepsilon) & |x-\theta| \le 2\varepsilon, \\ 1 & x - \theta > 2\varepsilon \end{cases}$$

where $0 < \varepsilon \ll 1$ controls the "speed" of transitions and (in [23]) depends on δ; θ is a location parameter and the weight matrix W is chosen using

$$W_{ij} = w_t + (w_s - w_t)\delta_{ij} + (w_p - w_t)a_{ji} + (w_m - w_t)a_{ij}$$

where a_{ij} are the entries of the adjacency matrix of the graph of Σ, δ_{ij} is the Kronecker delta and $w_p, w_m, w_s, w_t \in \mathbb{R}$ are parameters which are used to control the excitability threshold and the location of the equilibria.

In order to control the sensitivity of Excitable Network Attractors to inputs, we need a constructive way of realising ENAs of arbitrary amplitude where the excitability threshold can be controlled for each connection. In the construction described above, there is no direct way to control the excitability amplitude without changing the parameter ε. This means that both the activation function and the weight matrix needs to change as the amplitude changes. Although this is useful in understanding global bifurcation that may arise as ε is varied, there is no way of controlling the excitability threshold for a given connection.

This means that in their construction the excitability threshold is a global property of the constructed ENA and not for each connection. Numerical bifurcation analysis in w_p revealed that in an ENA with two nodes and one leading direction, the excitable connection is destroyed on increasing w_p in a saddle–node bifurcation (see Fig. 7C). Our simulation with other graph topologies also revealed the same bifurcations are responsible for destroying excitable connections. It can be shown that this corresponds to the value of w_p where two of the fixed points of the map $\phi_P(x) + w_p$ for $x \le 0.5$ merge. As w_p gets close to this value, these fixed points get closer to each and the amplitude of the resulting network becomes smaller. These observations revealed that it may be possible to use w_p as the only parameter for controlling the amplitude and by extension, the excitability threshold between any two nodes of Σ.

Using this idea, we extend the results of [23] to show that with the activation function ϕ_S (see Fig. 2 for $\phi_S(x)$ shifted by 0.25) where

$$\phi_S(x) = \gamma\left(2x - \frac{1}{2}\right)$$

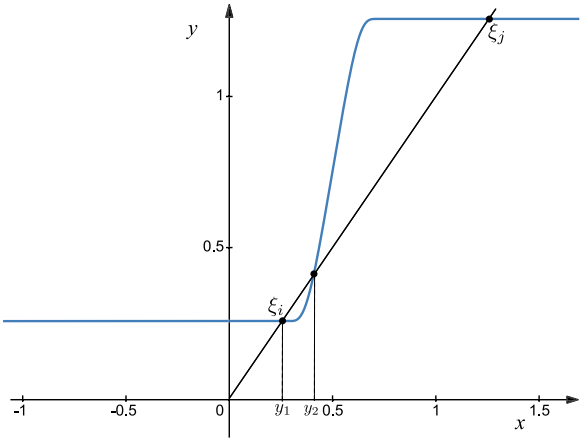

Fig. 2. The graph of $w_s\phi_S(x)+w_p$ for $w_p = 0.25$ and $w_s = 1$ (in blue) showing the fixed points y_1 and y_2. The distance between y_1 and y_2 along the x-axis translates to the excitability threshold in phase space. This can be made arbitrary small by shifting the $\phi_S(x)$ up or down appropriately by choosing w_p. y_1 would corresponds to the active equilibria ξ_i and once the thresholds between y_1 and y_2 is exceeded in the corresponding coordinate direction, a transition occur from ξ_i to ξ_j.

with γ defined as the smooth transition function

$$\gamma(z) = \begin{cases} 0 & \text{if } z \le 0, \\ \frac{e^{-1/z}}{e^{-1/z}+e^{-1/(1-z)}} & \text{if } 0 < z < 1, \\ 1 & \text{if } z \ge 1. \end{cases}$$

The transition function γ can in general be used to join two smooth function f and g in the interval (a,b) where f is defined on $(-\infty,a]$ and g is defined on $[b,\infty)$ by setting

$$h(x) = \left(1-\gamma\left(\frac{x-a}{b-a}\right)\right) f(x) + \gamma\left(\frac{x-a}{b-a}\right) g(x).$$

Hence $h(x) = f(x)$ for $x \le a$ and $h(x) = g(x)$ for $x \ge b$ and h is smooth in the interval (a,b) matching derivatives of all orders of f at a and derivatives of all orders of g at b. The function ϕ_S smoothly interpolates between the functions $f(x) = 0$ for $x \le \frac{1}{4}$ and $g(x) = 1$ for $x \ge \frac{3}{4}$.

It should be noted that γ is a non-analytic smooth function — it is everywhere smooth but does not have a convergent Taylor series in any neighbourhood of the origin. Furthermore, computing γ close to 0 or 1 can be problematic because either $1/x$ or $1/(1-x)$ tend to infinity. These properties are inherited by ϕ_S. An alternative choice of activation is the smooth activation function ϕ_ε introduced in [23]

$$\phi_\varepsilon(x) = \frac{1}{1+\exp(-(x-0.5)/\varepsilon)} \quad \text{for } 0 < \varepsilon \ll 1.$$

If the adjacency matrix A is admissible and $\delta > 0$ is sufficiently small, it can be shown that (1) with activation function ϕ_S realises A as an Excitable Network Attractor with amplitude $\delta > 0$ by choosing components of W as follows

$$w_{ij} = \begin{cases} w_s & \text{if } i = j, \\ w_m & \text{if } a_{ij} = 1, \\ w_p & \text{if } a_{ji} = 1, \\ w_t & \text{if } a_{ij} = a_{ji} = 0. \end{cases} \tag{3}$$

where

- $w_t = 0$,
- $w_s = \max_{i\in\{1,\dots,N\}}\{2, |T(i)|\}$.
- w_p is chosen so that $w_s\phi_S(x)+w_p$ have two fixed points y_1, y_2 with $y_1 < y_2 \le \frac{1}{2}$ that are at most δ apart (see Fig. 2). We prove in Appendix A that this is always possible if δ is sufficiently small and illustrated it in Fig. 2. The value of w_p corresponding to $\delta = 0$ is denoted as β, and y_1 and y_2 merge into a single fixed point denoted α at this value of w_p.
- $w_m = -w_s(1+2w_p)$.

The equilibrium identified with vertex k of the graph G of A (i.e corresponding node of the network Σ) is given by

$$[\xi_k]_i = \begin{cases} Y_A^k & \text{if } k = i, \\ Y_T^{ki} & \text{if } a_{ik} = 1, \\ Y_L & \text{if } a_{ki} = 1, \\ Y_D^{ki} & \text{if } a_{ik} = a_{ki} = 0. \end{cases} \tag{4}$$

where

$$\begin{aligned} Y_L &= y_1, \\ Y_A^k &= w_s + |L(k)|w_m\phi_S(Y_L), \\ Y_T^{ki} &= w_m + |L(k)\cap T(i)|w_p\phi_S(Y_L), \\ Y_D^{ki} &= w_t + \phi_S(Y_L)\big(w_m|L(k)\cap L(i)| + w_p|L(k)\cap T(i)|\big) \end{aligned}$$

where Y_A^k, Y_L Y_T^{ki} and Y_{ki}^D are associated with active, leading, trailing and disconnected directions respectively. It is easy to see that $\phi_S(Y_A^k) = 1$ and $\phi_S(Y_T^{ki}) = \phi_S(Y_D^{ki}) = 0$. Thus ϕ_S in some way tells us which node of Σ we are closed to. We carry the naming convention of Definition 2.7 to a generic point x in the phase space containing Σ. A cell k is **active** if $\phi_S(x_k) = 1$. A cell j is said to be

- **leading** from k if $\phi_S(x_j) \in (0,1)$ and $a_{kj} = 1$,
- **trailing** to k if $a_{jk} = 1$ and $\phi_S(x_j) = 0$ and
- **disconnected** from k if $a_{kj} = a_{jk} = 0$ and $\phi_S(x_j) = 0$.

Now we state the main result of this paper.

Theorem 3.1. *Let $0 \le \delta \ll 1$ and $A \in \mathbb{R}^{N\times N}$ be an admissible adjacency matrix. Let the components of W be chosen using (3). Then the input-free dynamics of (1) with the activation function ϕ_S realises A as an Excitable Network Attractor at amplitude δ.*

We used the following lemma in the proof of Theorem 3.1. It gives an upper bound on the difference between y_1 and w_p.

Lemma 3.2. *Let $0 < \delta \ll 1$, $w_s \ge 1$ and $w_p = w_p(\delta, w_s)$ be chosen so that $w_s\phi_S(y)+w_p$ has two fixed points in the interval $[0,\frac{1}{2}]$ that are δ apart. Let $y_1 = y_1(\delta, w_s)$ be the smaller of the two fixed points. Then* $\sup_{w_s,\delta,} w_s\phi_S(y_1) = \sup_{\delta,w_s}(y_1 - w_p) \le \frac{1}{4}$.

Proof. We may assume that for some choices of w_s and δ, $y_1 > \frac{1}{4}$. Otherwise there is nothing to prove. First let w_s be fixed. The largest value of $w_s\phi_S(y_1)$ occurs when y_1 is at its maximum. This occurs at $\delta = 0$ when $y_1 = \alpha$ where α satisfies

$$w_s\phi_S'(\alpha) = 1.$$

Now since $\phi_S''(y) > 0$ for $\frac{1}{4} < y < \frac{1}{2}$, the corresponding solution α to the above equation decreases as w_s increases. In particular, the largest possible value of α occurs when $w_s = 1$. Thus

$$\sup_{\delta,w_s}(y_1 - w_p) = w_s\phi_S(\alpha) \quad \text{when } w_s = 1.$$

It can be shown that $\phi_S'(0.36) \ge 1$ and thus we have $\alpha \le 0.36$ in which case $\phi_S(\alpha) \le \phi_S(0.36) < \frac{1}{4}$. □

From (3), $w_s > 1$ and thus the inequality above is strict. Since y_1 and y_2 tend to each other as $\delta \to 0$, we can take δ sufficiently small so that

$$\sup_{w_s}(y_2 - w_p) \le \frac{1}{4} \tag{5}$$

Proof of Theorem 3.1. To realise A as an ENA at amplitude δ using (1), we choose the parameters w_s, w_p, w_t, w_m using (3). It is sufficient to show it for the case $\tau_i = 1$ for all i and the results follows by topological

equivalence for all other choice of τ_i satisfying $\tau_i > 0$ [33] with the transformation $z_i = \tau_i x_i$. The remainder of the proof is divided into three main parts.

1. *We show that ξ_k is a stable equilibrium of Eq. (1)*

We will show that ξ_k (4) is a stable equilibrium of (1) identified with vertex k of the graph of A. First we note that $Y_A^k \geq \frac{3}{4}$ and $Y_T^{ki}, Y_D^{ki} \leq \frac{1}{4}$ for all values of k and i for which they are defined. Let $f : \mathbb{R}^N \to \mathbb{R}^N$ be the function with components $f_i(x) = -x_i + \sum_{j=1}^N w_{ij}\phi_S(x_j)$ for $i = 1, \ldots, N$. Since the trailing and disconnected directions of k are both less than $\frac{1}{4}$ at ξ_k, the non zero contributions to $f_i(\xi_k)$ comes from k and its leading directions. Thus

$$\begin{aligned} f_i(\xi_k) &= -[\xi_k]_i + \sum_{j=1}^N w_{ij}\phi_S([\xi_k]_j) \\ &= -[\xi_k]_i + w_{ii}\phi_S([\xi_k]_i) + (1-\delta_{ki})w_{ik}\phi_S([\xi_k]_k) \\ &\quad + \sum_{\{a_{kj}=1\}} w_{ij}\phi_S([\xi_k]_j). \end{aligned}$$

There are four cases to consider.

(a) For $i = k$,

$$\begin{aligned} f_k(\xi_k) &= -Y_A^k + w_s\phi_S(Y_A) + \sum_{\{a_{kj}=1\}} w_m\phi_S(Y_L) \\ &= -Y_A^k + w_s + w_m|L(k)|\phi_S(Y_L) \\ &= 0. \end{aligned}$$

(b) For the case $a_{ki} = 1$, note that $i \neq k$ since A has no one cycles. Furthermore, $a_{ij} = a_{ji} = 0$ for all $j \in L(k)\backslash\{i\}$ by lack of Δ-cliques. Thus

$$\begin{aligned} f_i(\xi_k) &= -[\xi_k]_i + w_{ii}\phi_S([\xi_k]_i) + w_{ik}\phi_S([\xi_k]_k) \\ &\quad + \sum_{\{a_{kj}=1, j\neq i\}} w_{ij}\phi_S([\xi_k]_j) \\ &= -Y_L + w_s\phi_S(Y_L) + w_p\phi_S(Y_A) \\ &= -Y_L + w_s\phi_S(Y_L) + w_p \\ &= 0. \end{aligned}$$

(c) For the case $a_{ik} = 1$. Again i must be different from k. For $j \in L(k)$, we must have that $j \in T(i) \cup D(i)$ since A has no Δ-cliques in A. Hence

$$\begin{aligned} f_i(\xi_k) &= -[\xi_k]_i + w_{ii}\phi_S([\xi_k]_i) + w_{ik}\phi_S([\xi_k]_k) \\ &\quad + \sum_{\{a_{kj}=1\}} w_{ij}\phi_S([\xi_k]_j) \\ &= -[\xi_k]_i + w_{ii}\phi_S([\xi_k]_i) + w_{ik}\phi_S([\xi_k]_k) \\ &\quad + \sum_{\{a_{kj}=1\}} w_{ij}\phi_S([\xi_k]_j) \\ &= -Y_T^{ki} + w_s\phi_S(Y_T^{ki}) + w_m\phi_S(Y_A) \\ &\quad + \sum_{\{a_{kj}=1\}} w_{ij}\phi_S([\xi_k]_j) \\ &= -Y_T^{ki} + w_m + |L(k)\cap T(i)|w_p\phi_S(Y_L) \\ &= 0. \end{aligned}$$

(d) For $a_{ki} = a_{ik} = 0$,

$$\begin{aligned} f_i(\xi_k) &= -[\xi_k]_i + w_{ii}\phi_S([\xi_k]_i) \\ &\quad + w_{ik}\phi_S([\xi_k]_k) + \sum_{\{a_{kj}=1\}} w_{ij}\phi_S([\xi_k]_j) \\ &= -Y_D^{ki} + w_s\phi_s(Y_D^{ki}) + w_{ik}\phi_S(Y_A) + \sum_{\{a_{kj}=1\}} w_{ij}\phi_S([\xi_k]_j) \\ &= -Y_D^{ki} + w_t + \phi_S(Y_L)\big(w_p|L(k)\cap T(i)| \\ &\quad + w_m|L(k)\cap L(i)|\big) \\ &= 0. \end{aligned}$$

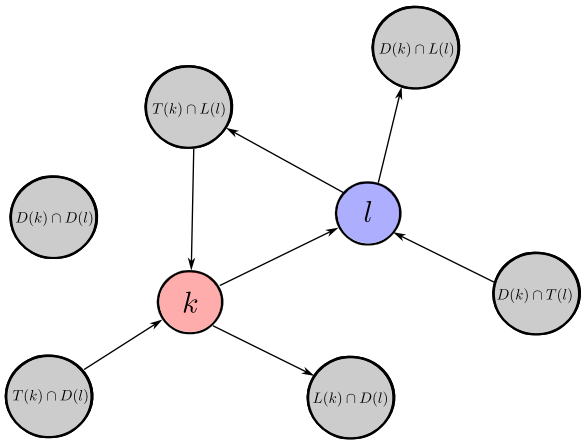

Fig. 3. A schematic diagram showing the classification of all the vertex in the graph of Σ during a transition from active cell k to a leading cell l. The remaining cells depending on whether they trailing, leading or disconnected with respect to k before the transitions and with respect to l after the transitions. Due to lack of one-cycles, two-cycles and Δ-cliques, this classification is exhaustive.
Source: Image adopted from [23].

Thus ξ_k is an equilibrium for each k. Now we will show that ξ_k is linearly stable. Let J^k be the Jacobian evaluated at ξ_k. Then

$$\begin{aligned} J^k_{il} &= -\delta_{il} + w_{il}\frac{\partial\phi_S([\xi_k]_l)}{\partial x_l} \\ &= -\delta_{il} + w_{il}\phi'_S([\xi_k]_l) \\ &= \begin{cases} -\delta_{il} & \text{for } l \notin L(k) \\ -\delta_{il} + w_{il}\phi'_S(Y_L) & \text{for } l \in L(k). \end{cases} \end{aligned}$$

It is easy to see that -1 is an eigenvalue of J^k(repeated $N - |L(k)|$ times). In particular, for $j \notin L(k)$, the corresponding eigenvector for the eigenvalue $\lambda_j = -1$ is the standard unit vector $\mathbf{e}_j$. If the set $L(k)$ is empty, then we are done. Otherwise, for each $j \in L(k)$, $\lambda_j = -1 + w_s\phi'_S(Y_L)$ is an eigenvalue of J^k. This is because the components of $J^k - \lambda_j I_N$ where I_N is the $N \times N$ identity matrix are

$$\begin{aligned} J^k_{jl} - \lambda_j\delta_{jl} &= J^k_{jl} - (-1 + w_s\phi'_S(Y_L))\delta_{jl} \\ &= \begin{cases} -w_s\phi'_S(Y_L)\delta_{il} & \text{for } l \notin L(k), \\ w_{il}\phi'_S(Y_L)(1-\delta_{il}) & \text{for } l \in L(k). \end{cases} \end{aligned}$$

Due to lack of Δ-clique, for each l with $l \neq j$, either

(a) $l \notin L(k)$ or
(b) $l \in L(k)$ but disconnected from j (see Fig. 4).

In both cases, the entry $J^k_{jl} - \lambda_j\delta_{jl}$ must be 0. Furthermore, $J^k_{jj} = -1 + w_s\phi'_S(Y_L)$. Thus row j of $J^k - \lambda_j I_N$ are all zeros from which it follows that λ_j is an eigenvalue of J^k. Since $\lambda_j < 0$ (see Lemma A.1 for the proof of this), ξ_k is a linearly stable.

Associated with the edge $a_{kl} = 1$ of A is the saddle η_{kl} in phase space which differ from ξ_k at the index l and (possibly) k. In particular

$$[\eta_{kl}]_i = \begin{cases} [\xi_k]_i & i \neq l, k; \\ y_2 & i = l; \\ Y_A^k + \frac{w_m}{w_s}(y_2 - Y_L) & i = k. \end{cases} \tag{6}$$

Following similar arguments above, it can be shown that η_{kl} is an equilibrium of the system and that -1 and $\lambda_l = -1 + w_s\phi'_S(y_2) > 0$ are eigenvalues of the Jacobian at η_{kl}.

2. *If $a_{kl} = 1$, there is an excitable connections at amplitude δ from ξ_k to ξ_l*

Assume $a_{kl} = 1$. We define the following region:

$$R^k := \{x \in \mathbb{R}^N | x_k \geq \frac{3}{4}; x_i \geq 0, i \in L(k); x_i \leq \frac{1}{4}, i \in T(k) \cup D(k)\}$$

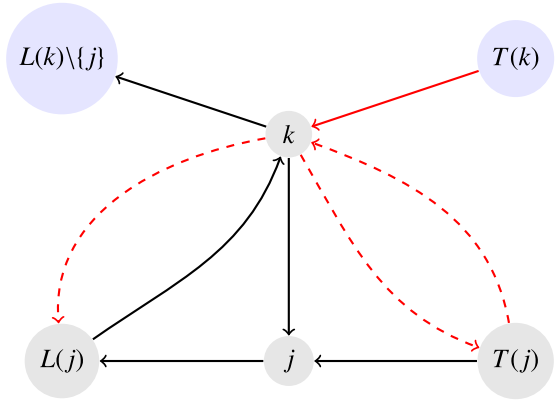

Fig. 4. A schematic diagram showing admissible connections involving $T(i), T(k), L(k), L(j)$ and the vertex k. The black arrows are admissible connections while the red dashed arrows are not allowed. In particular, in phase space directions corresponding to the red dashed arrows, $\phi_S([\xi_k]_j) = 0$.

R^l is defined similarly. By virtue of there being no Δ-cliques or two-cycles, each $i \neq k, l$ can be classified as one of six possibilities depending whether it is trailing, leading or disconnected from k and l. These possibilities are $L(k)\cap D(l), T(k)\cap L(l), D(k)\cap L(l), T(k)\cap D(l), D(k)\cap T(l)$ and $D(k)\cap D(l)$ (see Fig. 3). These are interpreted as how a cell switches from one type to another during a transition. As all leading cells of l are either disconnected or trailing from k, we have that $T(k)\cap L(l)\cup D(k)\cap L(l) = L(l)$. Furthermore, all leading cells of k are disconnected from l. Hence $L(k)\backslash\{l\} \subseteq D(l)$ from which we have $L(k)\cap D(l) = L(k)\backslash\{l\}$. Now let $\mathcal{P}(k,l) := T(k)\cap D(l)\cup D(k)\cap T(l)\cup D(k)\cap D(l)$.

Consider the initial condition $x(0) = \xi_k + \delta \mathbf{e}_l \in B_\delta(\xi_k)$. Let the associated trajectory be $x(t)$. In R^k,

$$\dot{x}_i = 0, \quad i \in L(k)\backslash\{l\}; \tag{7}$$

$$\dot{x}_l = -x_l + w_s\phi_S(x_l) + w_p > 0 \quad \text{since } x_i(0) > y_2; \tag{8}$$

$$\dot{x}_i = -x_i + [\xi_k]_i + w_{il}(\phi_S(x_l) - \phi_S(Y_L)), \quad i \in \{k\}\cup L(l). \tag{9}$$

Hence $x_i(t) = Y_L$ for $i \in L(k)\backslash\{l\}$ so long as $x_k(t) \geq \frac{3}{4}$ and $x_i(t) \leq \frac{1}{4}$ for $i \in L(l)\cap\mathcal{P}(k,l)$. Using Eqs. (7), (8) and (9), it can be shown that $x_i(t) = -\frac{w_{il}}{w_s}\delta e^{-t} + [\xi_k]_i + \frac{w_{il}}{w_s}(x_l(t) - Y_L)$ is a solution for $i \in \{k\}\cup L(l)$ and for all values of t for which $x(t)$ stays in R^k. From this, it follows that $x_k(t)$ is strictly decreasing and $x_l(t)$ is strictly increasing. At $x_l(t) = \frac{3}{4}$, we have that $x_k(t) \geq \frac{3}{4}$ and

$$\begin{aligned} x_i &= -\frac{w_p}{w_s}\delta e^{-t} + [\xi_k]_i + \frac{w_p}{w_s}(x_l(t) - Y_L), \\ &\leq [\xi_k]_i + \frac{w_p}{w_s}\left(\frac{3}{4} - Y_L\right), \\ &\leq [\xi_k]_i + \frac{w_p}{2w_s}, \\ &\leq [\xi_k]_i + \frac{w_p}{4}, \\ &\leq [\xi_k]_i + \frac{1}{8} \leq \frac{1}{4} \end{aligned}$$

for all $i \in L(l)$.

In addition, $x_i \leq [\xi_k]_i$ for $i \in \mathcal{P}(k,l)$ since $w_{il} \leq 0$ and

$$\begin{aligned} \dot{x}_i &= -x_i + w_s\phi_S(x_i) + w_{il}\phi_S(x_l) + w_{ik}\phi_S(x_k) + \sum_{j\in L(k)\backslash\{l\}} w_{ij}\phi_S(Y_L), \\ &\leq -x_i + w_{il}\phi_S(Y_L) + w_{ik} + \sum_{j\in L(k)\backslash\{l\}} w_{ij}\phi_S(Y_L), \\ &= -x_i + w_{ik} + \sum_{j\in L(k)} w_{ij}\phi_S(Y_L), \\ &= -x_i + [\xi_k]_i. \end{aligned} \tag{10}$$

Consequently, $x_l(t) = \frac{3}{4}$ before $x(t)$ exits R^k.

When $x_l \geq \frac{3}{4}$ and $x_k \geq 0$, $\dot{x}_k$ satisfies the following bound

$$\begin{aligned} \dot{x}_k &= -x_k + w_s\phi_S(x_k) + w_m\phi_S(x_l) + \sum_{j\neq l,k} w_{kj}\phi_S(x_j) \\ &\leq w_s\phi_S(x_k) + w_m + w_p|T(k)| \\ &\leq w_s + w_m + w_p w_s < 0. \end{aligned}$$

Hence $x_k(t)$ decreases monotonically in finite time to 0. This also means that when $x_l \geq \frac{3}{4}$, the region where $x_k \leq \frac{1}{4}$ and $x_k \leq \frac{3}{4}$ are positively invariant. Furthermore for any other $i \in T(l)$, the region $R^i \cap \{z \in \mathbb{R}^N : z_l \geq \frac{3}{4}, z_i \leq \frac{1}{4}\}$ is also positively invariant. Note that k is one such i since $k \in T(l)$. Furthermore $\phi_S(x_i)$ remains bounded above by 0 since

$$\begin{aligned} \dot{x}_i &= -x_i + w_s\phi_S(x_i) + w_{il}\phi_S(x_S) + w_{ik}\phi_S(x_k) + \sum_{\{j\neq i,l,k\}} w_{ij}\phi_S(x_j), \\ &\leq -x_i + w_s + w_m + w_p|T(i)|, \\ &\leq -x_i + w_s + w_m + w_p w_s, \\ &< -x_i - w_s w_p. \end{aligned}$$

We may now assume that $x_i(t) \leq \frac{1}{4}$ for $i \in T(l)$ and $x_l(t) \geq \frac{3}{4}$.

For a disconnected direction $i \in D(l)$, we have

$$\begin{aligned} \dot{x}_i &= -x_i + w_s\phi_S(x_i) + w_{il}\phi_S(x_l) + w_{ik}\phi_S(x_k) + \sum_{\{j\neq i,l,k\}} w_{ij}\phi_S(x_j), \\ &\leq -x_i + w_s\phi_S(x_i) + \sum_{\{j\in T(i)\backslash\{k\}\}} w_p\phi_S(x_j) + w_{ik}\phi_S(x_k), \\ &\leq -x_i + w_s\phi_S(x_i) + |T(i)|w_p\phi_S(y_2), \\ &\leq -x_i + w_s\phi_S(x_i) + w_s w_p\phi_S(y_2), \\ &\leq -x_i + w_s\phi_S(x_i) + \frac{w_p}{4} \end{aligned}$$

where we used the fact that $x_i(t) \leq y_2$ for all $i \neq k, l$ and (5). Since $\dot{x}_i < 0$ for $x_i \in \left[\frac{1}{4}, y_1\right]$, $\phi_S(x_i) = 0$ in finite time.

Finally for $i \in L(l)$, we have

$$\begin{aligned} \dot{x}_i &= -x_i + w_s\phi_S(x_i) + w_{il}\phi_S(x_l) + w_{ik}\phi_S(x_k) + \sum_{j\neq i,l,k} w_{ij}\phi_S(x_j) \\ &\geq -x_i + w_s\phi_S(x_i) + w_p + \sum_{\{j\in L(i)\}} w_m\phi_S(x_j) \\ &\geq -x_i + w_s\phi_S(x_i) + w_p + \sum_{\{j\in D(l)\}} w_m\phi_S(x_j) \end{aligned}$$

So the inequality

$$\dot{x}_i \geq -x_i + w_s\phi_S(x_i) + w_p > 0$$

will eventually hold for $x_i < y_1$.

Putting it all together, the trajectory of $x(0)$ enters R^l in finite time and since all initial conditions in R^l tend to ξ_l, we have that $x(t) \to \xi_l$ as $t \to \infty$. Fig. 5 schematically illustrates the existence of this connection.

3. *If $a_{kl} = a_{lk} = 0$, there is no excitable connection of amplitude δ from ξ_k to ξ_l*

Assume $a_{kl} = 0$. Let $x(0) \in B_\delta(\xi_k)$. We first note that $x_i(0) \leq \frac{1}{4}$ for all $i \in T(k)\cup D(k)$ and $x_k(0) \geq \frac{3}{4}$. There are three possibilities for the trajectory with initial condition $x(0)$.

(a) If all $x_j(0) < y_2$ for all $j \in L(k)$, then $x(0) \in W^s(\xi_k)$.

(b) If for some j_1, $x_j(0) \leq y_2$ for all $j \in L(k)\backslash\{j_1\}$ and $x_{j_1}(0) > y_2$, we have by (8) that $x \in W^s(\xi_{j_1})$.

(c) Finally if $x_i > y_2$ for some set of indices $i \in I \subseteq L(k)$ where $|I| > 1$, then each $x_i(t)$ satisfy the ODE (8) and it can be shown that the trajectory with initial condition $x(0)$ will enter, in finite time, an invariant set in the region where all indices in I are active i.e $\phi_S(x_i) = 1$ for all $i \in I$ (see Appendix B for proof of this).[1]

[1] The four cell levels of (4) cannot be used to describe the dynamics with multiple active cells. Whether or not a transition will happen to a node of Σ

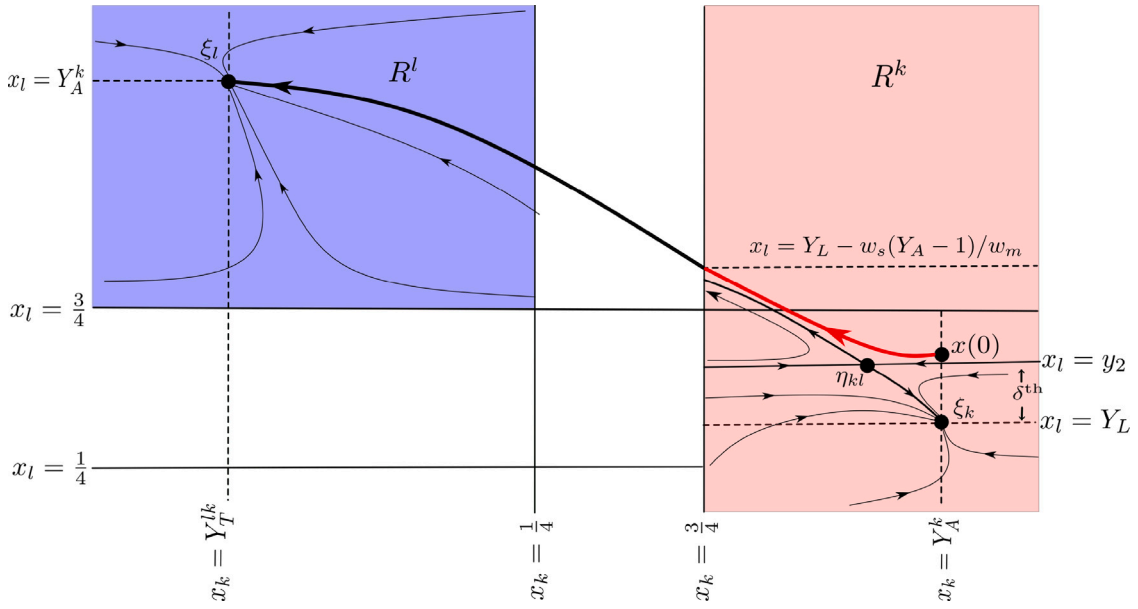

Fig. 5. A schematic diagram showing the dynamics in the region containing R^k (shaded light red) and R^l (shaded light blue). Initial conditions starting in R^k in the region where $x_l > y_2$ enters R^l in finite time. The trajectories for some of these initial conditions enters the region where $x_l \geq \frac{3}{4}$ before leaving R^k. The thick red trajectory for the initial condition $x(0) = \xi_k + \delta \mathbf{e}_l$ is one such trajectories. Trajectories starting in R^l asymptotes towards ξ_l in forward time.

In all cases, $x(0) \notin W^s(\xi_l)$. □

Remark 1. Since all of the equilibria involve in the realisation above are hyperbolic for $\delta > 0$, the realisation holds for an open sets of parameters around $(w_p, w_m, w_s, w_t) \in \mathbb{R}^4$. Furthermore, the proof can easily be extended to the case with different excitability threshold for each excitable connection by choosing w_p^{ij} to depend on δ_{th}^{ij} when $a_{ij} = 1$. Here we take $w_m = -w_s \max_{i,j}(1 + 2w_p^{ij})$ so that

$$w_{ij} = \begin{cases} w_s & \text{if } i = j, \\ w_m & \text{if } a_{ij} = 1, \\ w_p{}^{ji} & \text{if } a_{ji} = 1, \\ w_t & \text{if } a_{ij} = a_{ji} = 0. \end{cases}$$

Remark. The choice of ϕ_S as the activation function in (1) makes the proof above much easier. However, the construction described in the proof of Theorem 3.1 can be extended to a large class of activation functions. In particular, for $w_s \geq 2$, we prove in Appendix A that it is possible to choose the constant w_p so that $w_s\phi(y) + w_p$ has two fixed points that can be made arbitrarily close to each other for a large class of sigmoidal functions ϕ.

3.1. Multiple active cells and almost complete realisations

A δ perturbation within Σ may not always tend to an equilibrium in Σ. The question of almost complete realisation is concern with conditions under which this occurs only for a small set of initial conditions (in the sense of Lebesgue). As an example, when multiple directions from the current active cell cross the threshold (i.e bigger than y_2) in the above realisation, there will be a transition to an equilibrium outside the network attractor where all cells corresponding to these directions are active. Consider the realisation of the adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{11}$$

with the first vertex having two leading directions. The dynamics of the system is governed by

$$\begin{aligned} \dot{x}_1 &= -x_1 + w_s\phi_S(x_1) + w_m\phi_S(x_2) + w_m\phi_S(x_3) \\ \dot{x}_1 &= -x_2 + w_s\phi_S(x_2) + w_p\phi_S(x_1) + w_t\phi_S(x_3) \\ \dot{x}_1 &= -x_3 + w_s\phi_S(x_3) + w_p\phi_S(x_1) + w_t\phi_S(x_2) \end{aligned}$$

Apart from the equilibria ξ_1, ξ_2 and ξ_3 associated with the vertices of the graph of A, there is an equilibrium s_{23} in the region where $\phi_S(x_1) = 0$ and $\phi_S(x_2) = \phi_S(x_3) = 1$ (see Fig. 6). Both the x_2 and x_3 coordinates of s_{23} is $w_s + w_t$. Thus if $w_t + w_s = y_2$, then s_{23} merges with the equilibrium at (y_2, y_2) in a saddle–node bifurcation. Notices that for the same choice of w_t, the stable equilibria ξ_1, ξ_2 and ξ_3 still exist. Thus for $w_t \leq y_2 - w_s$, ξ_1 is almost complete.

Extending this method to an arbitrary adjacency matrix may require choosing w_t to be different for the sets of leading directions belonging to different active directions and understanding the bifurcations for these choices of w_t.

Remark. In general, there is a stable equilibrium in the region where $\phi_S(x_i) = \phi_S(x_j) = 1$ when $a_{ij} = a_{ji} = 0$ even if i and j do not belong to the same leading directions $L(k)$ of k. However, these equilibria will not be visited for $\delta < \frac{1}{4}$ from ξ_k since one of i or j is in $D(k) \cup T(k)$.

Alternatively, a condition can be imposed on w_p and δ so that transitions leading to multiple active cells do not occur. Suppose for a given $w_p \in (0, \beta)$, we know that $y_2 - y_1 = \tilde{\delta}$, then for all amplitudes $\delta \in (\tilde{\delta}, \sqrt{2}\tilde{\delta})$ and for this fixed choice of w_p, the realisation of Theorem 3.1 is almost complete. We use the following observation.

Lemma 3.3. *Let $0 < w_p < \beta$ be fixed so that $y_2 - y_1 = \tilde{\delta}$. Then for all $\delta \in (\tilde{\delta}, \tilde{\delta}\sqrt{2})$ and for any two distinct leading directions j_1 and j_2 of node k, we have that*

$$B_\delta(\xi_k) \cap \{x : x_{j_1}, x_{j_2} \geq y_2\} = \emptyset. \tag{12}$$

Proof. For the given w_p and $\delta \in (\tilde{\delta}, \tilde{\delta}\sqrt{2})$, if $x \in B_\delta(\xi_k)$ and $x_{j_1}, x_{j_2} \geq y_2$ for some distinct leading directions j_1 and j_2, then

$$\|\xi_k - x\|^2 \geq 2(y_2 - y_1)^2 \geq 2\tilde{\delta}^2$$

which is a contradiction. □

from this region will depends on interactions between the active cells in I and their neighbours in A.

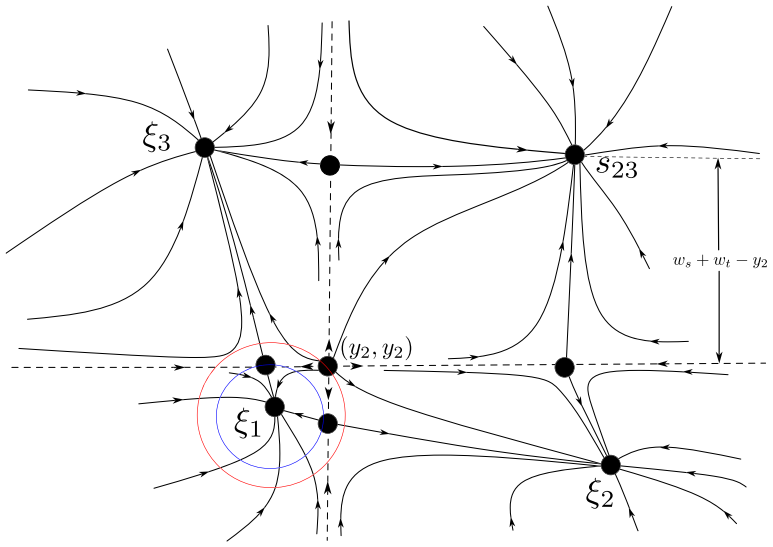

Fig. 6. A two dimensional projection of the phase portrait of the realisation of (11) in the region where $\phi_S(x_1) = 0$. Equilibria are marked with black dots with ξ_1, ξ_2 and ξ_3 denoting the nodes of Σ. The dashed straight lines are $x_1 = y_2$ and $x_2 = y_2$. The equilibrium s_{23} is in a region where both cell 2 and 3 are active. All perturbations from ξ_k inside the annular region between the red circle and the blue circle will either tend to ξ_1 or one of ξ_2 and ξ_3. On the other hand, if $w_t + w_s = y_2$, then the equilibrium s_{23} will disappear in a saddle–node bifurcation. In both cases, ξ_1 will be almost complete.

Theorem 3.4. *Let $A \in \mathbb{R}^{N\times N}$ be an admissible adjacency matrix. Let w_p be as in Lemma 3.3 and $\delta \in (\tilde{\delta}, \tilde{\delta}\sqrt{2})$. With this choice of w_p, the realisation of A using Theorem 3.1 as an ENA at amplitude δ is almost complete.*

Proof. Let $x \in B_\delta(\xi_k)$. We know that $\phi_S(x_i) = 0$ for all $i \notin \{k\} \cup L(k)$. By Lemma 3.3, there are two cases to consider.

1. If for some $j \in L(k)$, $x_j \geq y_2$, then $x_i < y_2$ for all $i \in L(k)\backslash\{j\}$. Consequently $x \in W^s(\xi_j)$ except possibly if $x_j = y_2$.
2. On the other hand, if $x_j < y_2$ for all $j \in L(k)$ then $x \in W^s(\xi_k)$. □

4. Equivalent formulation of Theorem 3.1

Two dynamical systems are said to be smoothly equivalent if there exist a smooth invertible transformation mapping the orbit of one to the orbit of the other. Smooth equivalence preserve equilibria and their stability. If W is invertible, it can be shown using the transformation $z = Wx$ that the input-free dynamics of the CTRNN (1) is smoothly equivalent to

$$\dot{z}_i = -z_i + \phi\left(\sum_{j=1}^{N} w_{ij} z_j\right) \quad \text{for } i = 1, \dots, N. \tag{13}$$

Orbits of system (1) are mapped to the orbits of system (13) through the invertible transformation $z = Wx$. Under some mild conditions on $u(t)$, this can be extended to the input-driven dynamics [34]. So if (1) exhibit an ENA in its phase space, then (13) will also have a network attractor in its phase space although the amplitude may be different. We use the following definition of *smooth equivalence* from [33].

Definition 4.1. Let F and G be smooth functions from $\mathbb{R}^N$ to $\mathbb{R}^N$. Two systems

$$\dot{x} = F(x),$$
$$\dot{y} = G(y)$$

are said to be **smoothly equivalent** if there exist a diffeomorphism $h : \mathbb{R}^N \to \mathbb{R}^N$ such that $F(x) = [Dh(x)]^{-1} G(h(x))$ where is $Dh(x)$ is the Jacobian matrix of $h(x)$ evaluated at x.

Remark. In terms of solutions between the two systems, two systems are smooth equivalent if the h is a diffeomorphism and the solutions are related by $y(t) = h(x(t))$.

Remark. By the inverse function theorem, a sufficient condition for h to be a diffeomorphism is that h is smooth and the determinant of the Jacobian $Jh(x)$ be non-vanishing at all points $x \in \mathbb{R}^N$. For the CTRNN (1) with $h = Wx$, this is equivalent to W being invertible.

Lemma 4.1. *Let $\delta > 0$ and $A \in \mathbb{R}^{N\times N}$ be an admissible adjacency matrix. Let W be chosen so that (1) has an ENA of amplitude δ in its phase space. Then there is an invertible choice $\tilde{W} \in \mathbb{R}^{N\times N}$ of W such that the dynamics of (1) realises A as an ENA at amplitude δ.*

Proof. As noted in Remark 1, the realisation of A using Theorem 3.1 holds for a sufficiently small open ball around (w_p, w_m, w_s, w_t) in $\mathbb{R}^4$. The results follows from the fact that for any singular matrix W, the matrix $\tilde{W} = W - \epsilon \mathbb{I}_N$ is non-singular for all $\epsilon > 0$ sufficiently small [35, p. 54]. □

Another way of viewing Lemma 4.1 is that we can make an arbitrary small perturbation to the choice of w_s to make W invertible without changing the arguments of the proof of Theorem 3.1. This leads to the following corollary which is an alternative formulation of Theorem 3.1.

Corollary 4.2. *Let $\delta > 0$ be sufficiently small and $A \in \mathbb{R}^{N\times N}$ be admissible. Suppose W is chosen using (3) with Lemma 4.1 applied so that it is invertible. Then the dynamics of*

$$\dot{z}_i = -z_i + \phi_S\left(\sum_{j=1}^{N} w_{ij} z_j\right) \quad \textit{for } i = 1, \dots, N$$

realises A as an Excitable Network Attractor at amplitude $h(\delta) > 0$ where h depends smoothly on δ and $h(0) = 0$.

Proof. We can write (1) in matrix form,

$$\dot{\mathbf{x}} = -\mathbf{x} + \phi_S(W\mathbf{x}).$$

Using the transformation $\mathbf{z} = W\mathbf{x}$, we have

$$\dot{\mathbf{z}} = -\mathbf{z} + W\phi_S(\mathbf{z}).$$

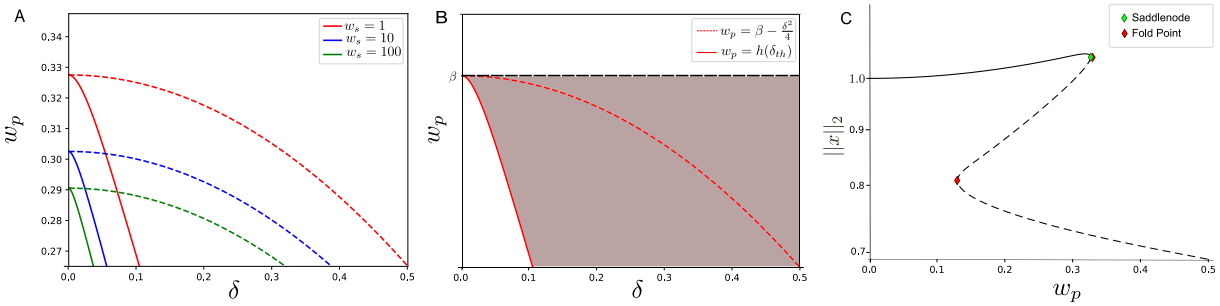

Fig. 7. A. A graph showing the function $w_p = \beta - \frac{\delta^2}{4}$ for $w_s = 1, 10, 100$ (dashed lines) and the corresponding excitability threshold for the each value of w_p (solid lines). B. For any fixed δ, the shaded region shows alternative choices of w_p where a choice of this w_p between two connected vertex in a graph will result in excitable connections of amplitude δ. The dashed line where $w_p = \beta$ corresponds to the choice of w_p with excitability threshold 0. Values of w_p above this will result in spontaneous transitions. C. A bifurcation diagram varying w_p. The solid black curve showing the stable equilibrium ξ_k and dashed curve showing the saddle η_{kl}. The excitable connection is destroyed in a saddle–node bifurcation.

Since W is invertible, the system (13) is smoothly equivalent to the system (1). Thus associated with any excitable connection $a_{kl} = 1$, the equilibria ξ_k, ξ_l and η_{kl} for the system (1) corresponds equilibria to $\tilde{\xi}_k = W\xi_k, \tilde{\xi}_l = W\xi_l$ and $\tilde{\eta}_{kl} = W\eta_{kl}$ respectively in the system (13) associated with the same excitable connection. Using (6), the excitability threshold $\tilde{\delta}_{\text{th}}$ associated with the connection $a_{kl} = 1$ in system (13) satisfies

$$\begin{aligned}
\tilde{\delta}_{\text{th}}^2 &\leq \|\tilde{\xi}_k - \tilde{\eta}_{kl}\|_2^2, \\
&\leq \|W(\eta_{kl} - \xi_k)\|_2^2, \\
&\leq \|W\|_2^2 \|\eta_{kl} - \xi_k\|_2^2, \\
&\leq C\|W\|_\infty^2 \|\eta_{kl} - \xi_k\|_2^2 \quad \text{for some } C > 0, \\
&= C\|W\|_\infty^2 \left((y_2 - Y_L)^2 + \frac{w_m^2}{w_s^2}(y_2 - Y_L)^2 \right), \\
&\leq C\|W\|_\infty^2 \left(\delta^2 + \delta^2(1 + 2w_p)^2 \right), \\
&= C\|W\|_\infty^2 \delta^2 \left(1 + (1 + 2w_p)^2 \right)
\end{aligned}$$

where $\|W\|_\infty$ is the maximum absolute row sum norm of W. For $\delta > 0$, W is invertible and so $\|W\|_2 > 0$. Furthermore,

$$\|W\|_\infty \leq |w_m| N = -w_m N$$

Now, define $h(\delta) = -C\delta N w_m \sqrt{\left(1 + (1 + 2w_p)^2\right)}$. Since w_p and w_m depends continuously on δ, h is a continuous function of δ with $h(0) = 0$. Furthermore, as $h'(0) \neq 0$, their is neighbourhood of the origin where h has a continuous inverse. □

Remark. Corollary 4.2 may not hold at $\delta = 0$ as W may be singular. In this case, adjustments of Lemma 4.1 will not be possible since any ϵ perturbation to w_s will either annihilate the fixed point α at $\delta = 0$ or $y_2 - y_1 > 0$.

5. Bifurcations creating ENAs

One of the main contribution of this paper is the way in which w_p is chosen for the fixed activation function ϕ_S. For $w_p = \beta - \frac{\delta^2}{4}$, it can be shown that $y_2 - y_1 < \delta$ (see Appendix A) and thus Excitable Network Attractors of amplitude δ exists for this choice of w_p. However, there are infinitely many such choices for w_p. Fig. 7 shows the exact nature of the relationship between w_p and δ for $w_s \in \{1, 10, 100\}$ and a bifurcation diagram for w_p when $w_s = 1$. In phase space, this bifurcations are responsible for destroying excitable connections and sometimes creating new connections.

Consider for example the dynamics of the leading cell l of k in the region R^k. From Eq. (8), x_l satisfies

$$\dot{x}_l = -x_l + w_s \phi_S(x_l) + w_p.$$

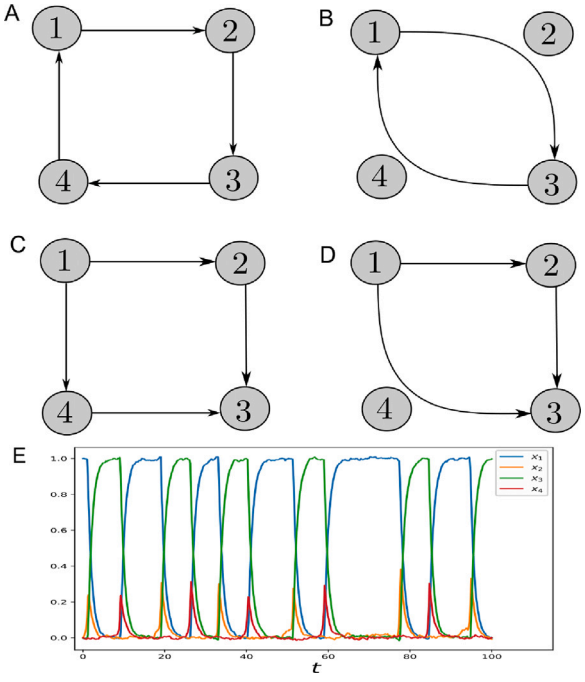

Fig. 8. A two cycle (B) and Δ-clique (D) realised in a four dimensional system. The two cycle is realised by severing the connections from vertex 2 to vertex 3 and from vertex 4 to vertex 1 in the four cycle A and setting the corresponding w_p large. For the Δ-clique, the graph C is used and the connection from vertex 4 to vertex 3 is severed. (E) Shows noise-driven trajectories of the two cycle in (B) using Eq. (13) with $s^{\text{noise}} = 0.07, w_p = 0.3$ and $w_s = 2$ initialised close to node 1 of the network. Notice that node 2 and 4 are only non-zero during transitions from node 1 to 3 and 3 to 1.

Taking the Taylor's expansion around $x = \alpha$ (see Appendix A for the definitions of α and β), we have that

$$\begin{aligned}
\dot{x}_l &= -x_l + w_s \phi_S(\alpha) + w_s(x_l - \alpha)\phi_S'(\alpha) \\
&\quad + w_s \frac{(x_l - \alpha)^2}{2} \phi_S''(\alpha) + w_p + \cdots \\
&= -x_l + w_p - \beta + w_s \phi_S(\alpha) + \beta + w_s(x - \alpha)\phi_S'(\alpha) \\
&\quad + \frac{w_s(x_l - \alpha)^2}{2} \phi_S''(\alpha) + \cdots \\
&= w_p - \beta - x_l + \alpha + w_s(x_l - \alpha)\phi_S'(\alpha) \\
&= w_p - \beta + (x_l - \alpha)(-1 + w_s \phi_S'(\alpha)) + \frac{w_s(x_l - \alpha)^2}{2} \phi_S''(\alpha) + \cdots \\
&= w_p - \beta + \frac{w_s(x_l - \alpha)^2}{2} \phi_S''(\alpha) + \cdots
\end{aligned}$$

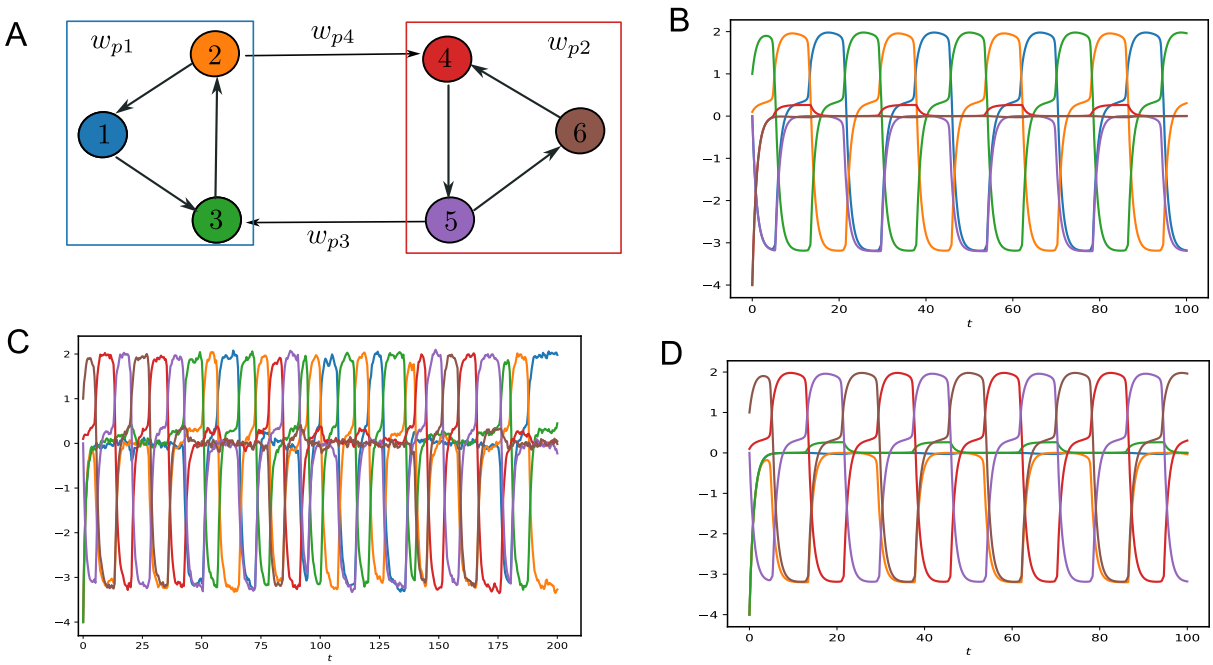

Fig. 9. Periodic orbits from two three cycles. In A, the cycle containing the node 1, 2 and 3 share w_{p1} while the cycle containing the nodes 4, 5 and 6 share w_{p2}. The edges along which transitions happens from the first cycle to the second have excitability parameters w_{p4} and w_{p3}. Setting $w_{p1} = w_{p2} = 0.3$ and $w_{p3} = w_{p4} = 0.2$, we observed periodic orbits C. and D. close to the first and the second cycle respectively. In the noisy case with $s_{\text{noise}} = 0.07$, we observe a sequence of transitions from the first cycle to the second (at $t = 40$) and back (at $t = 130$).

If $\phi_S''(\alpha) \neq 0$, the right hand side is the normal form of a saddle node bifurcation. It follows there is a saddle–node bifurcation at $w_p = \beta$ with two equilibria when $w_p < \beta$ which merge at $w_p = \beta$ and disappear when $w_p > \beta$. If w_p is shared between a set of connections, this bifurcation occurs for all leading directions in this set. If the leading directions are connected in a cycle, these global bifurcations are in fact a multiple saddle–node on an invariant circle (SNIC) bifurcation.

5.1. Generalisation to arbitrary directed graphs

If for $l \in L(k)$, we set w_p^{lj} much larger than β, no equilibrium exists in the region where ξ_k was located and all excitable transitions to ξ_k will tend to ξ_l. Hence an excitable connection between node k and node l of Σ becomes a connections from the trailing directions of k and node l. This approach can be used to realise a two cycle Fig. 8B using the adjacency matrix for the four cycle Fig. 8A by setting w_p^{32} and w_p^{14} much larger than β. Similarly, we can get the Δ-clique Fig. 8D in a four dimensional network attractor from the graph Fig. 8C using the same trick. Replacing two-cycles and Δ-cliques this way, we can realise any adjacency matrix $A \in \mathbb{R}^{N\times N}$ with no 1 cycle as an Excitable Network Attractor in no more than $N + 2\binom{N}{2} = N^2$ dimensions.

Corollary 5.1. *Given any 1-cycle free adjacency matrix $A \in \mathbb{R}^{N\times N}$ and $0 < \delta \ll 1$, there exist an admissible adjacency matrix $A' \in \mathbb{R}^{M\times M}$ where $M \leq N^2$ such that A can be realised as an ENA at amplitude δ using Theorem 3.1 with the adjacency matrix A'.*

The ENA in Corollary 5.1 can be thought of as identifying vertices k of the adjacency matrix A with equilibrium k^2 in N^2-dimensional system. Furthermore, some vertices in A' may not have corresponding equilibria in the ENA. These are only non-zero and sometimes active during transitions (see Fig. 8E). This realisation is similar to the coupled two-cell realisation described in [27] where p-cells classify which equilibria we are currently close to and y-cells become active only during transitions.

5.2. Network attractors of periodic orbits

We can use the simultaneous SNIC bifurcations that emerge when the vertices of A are in a cycle to realise excitable networks of periodic orbits. Consider the graph in Fig. 9A involving two connected three cycles. To avoid numerical instabilities in computing ϕ_S, we used ϕ_ε with $\varepsilon = 0.05$. The excitability parameter w_{p1} is shared between the nodes in the cycle on the left and w_{p2} is shared between the nodes of the cycle of the right. w_{p3} and w_{p4} controls transitions between the two cycles and are set to $w_{p3} = w_{p4} = 0.2$. Setting $w_{p1} = w_{p2} = 0.3$ turns each of the cycles into periodic orbits. Starting close to the first cycle in the noise-free, we see the sequence of spontaneous transitions $1 \to 2 \to 3 \to 1$ (see Fig. 9B). Similarly, we observe the sequence $4 \to 5 \to 6 \to 4$ in Fig. 9D. In Fig. 9C, the system is driven by independent Wiener noise with noise amplitude $s^{\text{noise}} = 0.07$. In this case, we observe $1 \to 2 \to 3 \to 1$ sequence of transitions and then $4 \to 5 \to 6 \to 4$ around $t = 130$ in Fig. 9C showing transitions between the periodic orbit in the region where x_1, x_2 and x_3 are active to the periodic orbit in the region where x_4, x_5 and x_6 are active. From Section 5.1, this can be extended to any sequence of transitions involving periodic orbits and equilibria.

6. Discussion and future directions

An important generalisation of Theorem 3.1 would be to find a method to realise a directed graph A so that each vertex of A is associated with a node $\xi_k \in \Sigma$ with multiple active cells. This will require a detailed bifurcation analysis to understand the role of w_p and w_t in the region with multiple active cells. In particular, if w_t is sufficiently large, it is possible to have realisation where ξ_k is in the region where cell k and all or some of its leading cells are active. It may even be possible to realise A in a way that ξ_k can selectively activate some of its leading directions if they satisfy some pre-specified condition. It would also be interesting to extend the result of Theorem 3.1 to the class of activation functions in Appendix A. It seems, at least for the case of ϕ_ε, that system (1) with the activation function ϕ_S and ϕ_ε are smoothly equivalent except in a small region containing $B_\epsilon(y_1) \cup B_\epsilon(y_2) \in \mathbb{R}^N$ for ϵ small. This is not trivial to prove as one needs to show that there is a

homeomorphism between the parameters in both systems. For example, for any $\tilde{\delta} > 0$ you may have to choose w_p in both systems so that $y_2 - y_1 = \tilde{\delta}$ for both activation functions.

In applications, ENA have been shown to be useful in explaining how trained RNNs perform sequential time-dependent finite state computation [14,15,24]. A finite state computer is an idealised model of computation where a system can be classified as belonging to one or more discrete sets of states at any given time. Classification problems in machine learning are essentially of this type. In such tasks, learning corresponds to choosing the require set of memory states and rules on how to switch from one state to another under the action of inputs. This takes into account the causal relationship between different memory states. As an example, consider part of speech tagging where a model needs to learn the part of speech of a word in a sentence using previously seen words in the sentence. Here, the model may learned that only certain parts of speech follows a noun.

Errors in RNNs with ENA in their phase space can be associated with transitions outside the set of equilibria relevant to the task — spurious transitions. If the ENA is an almost complete realisation at the largest excitability threshold, then such transitions do not occur and errors maybe due to noisy inputs that occasionally push the dynamics outside the set of "memory" states needed for the task. This may occur for example if other stable equilibria exists and the system is allowed to run for a very long time under small noise conditions [36]. On the other hand, if there are spurious equilibria with unusually small excitability threshold and "large" basin of attraction, structural errors may emerge in the prediction of the associated RNN. For example, it is observed that some errors in prediction in trained Echo State Networks are of this type [14,15].

A possible future direction of this research is to find ways to train such models to realise the ENA associated with a trained RNN. This will be computationally cheaper compared to for example the fixed point finding technique explained in [14,15]. Our experiments have shown that this is possible at least in the case of two-state finite state computations problems. For example it is possible to train the realisation of a two cycle to learn transitions in an Echo State Network trained to solve a 1-bit flip-flop task. A 1-bit flip-flop is a simple benchmark task for classification with discrete inputs and discrete outputs [14]. In a 1-bit flip-flop task, the input is mostly zero but at random times, it becomes either a plus one or minus one. However, it may not be appropriate to impose a particular graph structure on the ESN as other attractors maybe accessible from the required number of attractors needed to solve the task. Hence such a training process may involve allowing for A to be trainable possibly through genetic programming.

CRediT authorship contribution statement

Muhammed Fadera: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Peter Ashwin:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

We thank Claire Postlethwaite and Roman Borisyuk for their insightful comments. PA thanks EPSRC, United Kingdom for funding via EP/T017856/1 and MF thanks EPSRC, United Kingdom for studentship funding via EP/W523859/1. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Appendix A. Properties of sigmoidal functions

Definition A.1. A smooth function $\phi : \mathbb{R} \to \mathbb{R}$ is said to be sigmoidal if

1. $\phi(x)$ is increasing,
2. $\lim_{x\to-\infty} \phi(x) = 0$ and $\lim_{x\to\infty} \phi(x) = 1$.

Lemma A.1. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a sigmoidal function satisfying*

1. *$\phi'(0.5) \geq 1$ and $\phi'(0) < 1$,*
2. *$\phi''(x) > 0$ for $0 < x < 0.5$.*

Then for $0 < x \leq 0.5$, $\phi'(x) = 1$ has a unique solution.

Proof. Since $\phi'(x)$ is increasing for $x \leq 0.5$, the result follows from the fact that $\phi'(0) < 1$ and $\phi'(0.5) \geq 1$. □

Corollary A.2. *Let $w_s > 1$. Then $w_s\phi'(x) = 1$ has a unique solution for $0 < x \leq 0.5$.*

The remainder of Appendix A, we assume that ϕ is a sigmoidal function with $\phi'(0.5) \geq 1$ and $\phi'(0) < 1$.

Lemma A.3. *Suppose $w_s > 1$. Let α be the unique point in $(0, 0.5]$ such that $w_s\phi'(\alpha) = 1$. Define $\beta = \alpha - w_s\phi(\alpha)$. The function $g_\beta(y) = w_s\phi(y) + \beta - y$ satisfies $g_\beta(y) \geq 0$ for $y \leq 0.5$. Furthermore g_β attains its minimum at α.*

Proof. The results follows from the fact that $g'_\beta(y) < 0$ for $y < \alpha$, $g'_\beta(y) > 0$ for $\alpha < y \leq 0.5$ and $g(\alpha) = 0$. □

Lemma A.4. *Let $0 < \delta \ll 1$ and $w_s \geq 2$. Suppose ϕ satisfies $\phi''(\alpha) \geq 1$. Let $w_p = \beta - \frac{\delta^2}{4}$. The equation $g(y) = w_s\phi(y) + w_p - y$ for $0 < y \leq 0.5$ has two roots less than δ apart.*

Proof. Consider g evaluated at $r_1 = \alpha + \frac{\delta}{2}$ and $r_2 = \alpha - \frac{\delta}{2}$. Here

$$g(r_{1,2}) = w_s\phi(r_{1,2}) + \beta - \frac{\delta^2}{4} - r_{1,2} > 0$$

if $g(r_1) > 0$. Using Taylor expansion of g near α, we have

$$\begin{aligned} g(r_1) &= w_s\left(\phi(\alpha) + \phi'(\alpha)\frac{\delta}{2} + \phi''(\alpha)\frac{\delta^2}{8} + O(\delta^3)\right) + \beta - \frac{\delta^2}{4} - \alpha - \frac{\delta}{2} \\ &= \left(\frac{w_s\phi'(\alpha)}{2} - \frac{1}{2}\right)\delta + \left(\frac{w_s\phi''(\alpha)}{8} - \frac{1}{4}\right)\delta^2 + O(\delta^3) \\ &= \left(\frac{w_s\phi''(\alpha)}{8} - \frac{1}{4}\right)\delta^2 + O(\delta^3) \\ &> 0 \end{aligned}$$

if δ is sufficiently small. Since $g(\alpha) < 0$ and $g(r_{1,2}) > 0$, the statement follows. □

Appendix B. Invariant sets with multiple active cells

Proposition B.1. *Let Σ be an ENA at amplitude $\delta > 0$. Suppose $x \in B_\delta(\xi_k)$ with $x_i \geq y_2$ for $i \in I \subseteq L(k)$. Then the trajectory with enters, in finite time, an invariant set where $\phi_S(x_i) = 1$ for all $i \in I$.*

Proof. For each $i \in I$, x_i satisfies the ODE

$$\dot{x}_i = -x_i + w_s \phi_S(x_i) + w_p$$

in R^k. For cells in $\mathcal{P}(k,l)$, we have

$$\dot{x}_i \leq -x_i + [\xi_k]_i + \sum_{j \in I} w_{ij}(\phi_S(x_j) - \phi_S(y_1)).$$

Hence $\phi_S(x_i) = 0$ in R^k for $i \in \mathcal{P}(k,I)$ before $x_j = \frac{3}{4}$ for all $j \in I$. So x enters the region where $x_i \geq \frac{3}{4}$ for all $i \in I$ before leaving R^k. Following the remainder of the proof of the existence of an excitable connection in Theorem 3.1, $x(t)$ will enter an invariant region where all cells in I are active. □

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physd.2024.134358.

References

[1] H. Hewamalage, C. Bergmeir, K. Bandara, Recurrent neural networks for time series forecasting: Current status and future directions, Int. J. Forecast. 37 (1) (2021) 388–427.

[2] H. Jaeger, Adaptive nonlinear system identification with echo state networks, Adv. Neural Inf. Process. Syst. 15 (2002).

[3] Y. Xia, D.P. Mandic, M.M. Van Hulle, J.C. Principe, A complex echo state network for nonlinear adaptive filtering, in: 2008 IEEE Workshop on Machine Learning for Signal Processing, IEEE, 2008, pp. 404–408.

[4] D. Sussillo, L.F. Abbott, Generating coherent patterns of activity from chaotic neural networks, Neuron 63 (4) (2009) 544–557.

[5] K.-I. Funahashi, Y. Nakamura, Approximation of dynamical systems by continuous time recurrent neural networks, Neural Netw. 6 (1993) 801–806.

[6] L. Gonon, J.P. Ortega, Fading memory echo state networks are universal, Neural Netw. 138 (2021) 10–13, http://dx.doi.org/10.1016/j.neunet.2021.01.025.

[7] B. Goodman, S. Flaxman, European union regulations on algorithmic decision-making and a "right to explanation", AI Mag. 38 (3) (2017) 50–57.

[8] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[9] T. Brennan, W.L. Oliver, Emergence of machine learning techniques in criminology: implications of complexity in our data and in research questions, Criminol. Pub. Pol'y 12 (2013) 551.

[10] R.D. Beer, J.C. Gallagher, Evolving dynamical neural networks for adaptive behavior, Adapt. Behav. 1 (1) (1992) 91–122.

[11] R.D. Beer, et al., Toward the evolution of dynamical neural networks for minimally cognitive behavior, From Animals to Animats 4 (1996) 421–429.

[12] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. 79 (8) (1982) 2554–2558.

[13] J. Steinberg, H. Sompolinsky, Associative memory of structured knowledge, Sci. Rep. 12 (1) (2022) 21808.

[14] D. Sussillo, O. Barak, Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks, Neural Comput. 25 (3) (2013) 626–649.

[15] A. Ceni, P. Ashwin, L. Livi, Interpreting recurrent neural networks behaviour via excitable network attractors, Cogn. Comput. 12 (2) (2020) 330–356.

[16] J. Creaser, P. Ashwin, C. Postlethwaite, J. Britz, Noisy network attractor models for transitions between EEG microstates, J. Math. Neurosci. 11 (1) (2021) 1–25.

[17] M.I. Rabinovich, M.A. Zaks, P. Varona, Sequential dynamics of complex networks in mind: Consciousness and creativity, Phys. Rep. 883 (2020) 1–32.

[18] K. Kaneko, On the strength of attractors in a high-dimensional system: Milnor attractor network, robust global attraction, and noise-induced selection, Physica D 124 (4) (1998) 322–344.

[19] I. Tsuda, Chaotic itinerancy and its roles in cognitive neurodynamics, Curr. Opin. Neurobiol. 31 (2015) 67–71.

[20] R. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, in: International Conference on Machine Learning, PMLR, 2013, pp. 1310–1318.

[21] M. Arjovsky, A. Shah, Y. Bengio, Unitary evolution recurrent neural networks, in: International Conference on Machine Learning, PMLR, 2016, pp. 1120–1128.

[22] A. Ceni, P. Ashwin, L. Livi, Interpreting recurrent neural networks behaviour via excitable network attractors, Cogn. Comput. 12 (2) (2020) 330–356, http://dx.doi.org/10.1007/s12559-019-09634-2.

[23] P. Ashwin, C. Postlethwaite, Excitable networks for finite state computation with continuous time recurrent neural networks, Biol. Cybern. 115 (5) (2021) 519–538.

[24] M. Casey, The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction, Neural Comput. 8 (6) (1996) 1135–1178.

[25] P. Tiňo, B.G. Horne, C.L. Giles, P.C. Collingwood, Finite state machines and recurrent neural networks—automata and dynamical systems approaches, in: Neural Networks and Pattern Recognition, Elsevier, 1998, pp. 171–219.

[26] J.E. Hopcroft, R. Motwani, J.D. Ullman, Introduction to automata theory, languages, and computation, Acm Sigact News 32 (1) (2001) 60–65.

[27] P. Ashwin, C. Postlethwaite, Designing heteroclinic and excitable networks in phase space using two populations of coupled cells, J. Nonlinear Sci. 26 (2016) 345–364.

[28] P. Ashwin, S.B. Castro, A. Lohse, Almost complete and equable heteroclinic networks, J. Nonlinear Sci. 30 (1) (2020) 1–22.

[29] K. Nikiforou, The Dynamics of Continuous-Time Recurrent Neural Networks and Their Relevance to Episodic Memory (Ph.D. thesis), Imperial College London, 2019.

[30] J. Blynel, D. Floreano, Exploring the T-maze: Evolving learning-like robot behaviors using CTRNNs, in: Workshops on Applications of Evolutionary Computation, Springer, 2003, pp. 593–604.

[31] R.M. Hasani, M. Lechner, A. Amini, D. Rus, R. Grosu, Liquid time-constant recurrent neural networks as universal approximators, 2018, arXiv preprint arXiv:1811.00321.

[32] R. Hasani, M. Lechner, A. Amini, D. Rus, R. Grosu, Liquid time-constant networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 7657–7666.

[33] Y.A. Kuznetsov, I.A. Kuznetsov, Y. Kuznetsov, Elements of Applied Bifurcation Theory, Vol. 112, Springer, 1998.

[34] K.D. Miller, F. Fumarola, Mathematical equivalence of two common forms of firing rate models of neural networks, Neural Comput. 24 (1) (2012) 25–31.

[35] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 2012.

[36] M.I. Freidlin, A.D. Wentzell, M. Freidlin, A. Wentzell, Random Perturbations, Springer, 1998.