

From neural network to psychophysics of time: Exploring emergent properties of RNNs using novel Hamiltonian formalism

Rakesh Sengupta, Anindya Pattanayak, Raju Surampudi Bapi,

Abstract—The stability analysis of dynamical neural network systems generally follows the route of finding a suitable Liapunov function after the fashion Hopfield’s famous paper on content addressable memory network or by finding conditions that make divergent solutions impossible. For the current work we focused on biological recurrent neural networks (bRNNs) that require transient external inputs (Cohen-Grossberg networks). In the current work we have proposed a general method to construct Liapunov functions for recurrent neural network with the help of a physically meaningful Hamiltonian function. This construct allows us to explore the emergent properties of the recurrent network (e.g., parameter configuration needed for winner-take-all competition in a leaky accumulator design) beyond that available in standard stability analysis, while also comparing well with standard stability analysis (ordinary differential equation approach) as a special case of the general stability constraint derived from the Hamiltonian formulation. We also show that the Cohen-Grossberg Liapunov function can be derived naturally from the Hamiltonian formalism. A strength of the construct comes from its usability as a predictor for behavior in psychophysical experiments involving numerosity and temporal duration judgements.

Index Terms—Recurrent neural networks, Liapunov function, Winner-take-all, Time’s subjective expansion, Temporal Oddball

I. INTRODUCTION

Recurrent neural network (RNN) consists of interconnected neurons that have some feedback loops built-in between the nodes. The feedback can come from the same or different node at each time point of the network’s evolution and the behavior of the network resembles a nonlinear dynamical systems. These networks have been used for constructing neural models for memory [1], decision making [2], visual sense of numbers [3] to name a few. Some dynamic vision algorithms include recurrent neurons as building blocks for their usefulness in integrating and spreading local and global influences across the network [4]. The biological RNNs (bRNN) are mostly imagined as a single layer of neurons.

The on-center off-surround kind of recurrent networks have been very popular in the literature involving short-term mem-

ory, decision making, contour enhancement, pattern recognition, and several other problems because of their versatility and ability to produce self-organized outputs based on underlying non-linear mathematical properties of the network. One of the standard approaches towards the stability of these networks has always been to find a suitable Liapunov function or finding the condition for which network trajectories do not diverge [5], [6]. A detailed review of Liapunov and other stability approaches can be found in [7]. In the current paper we show a novel intuitive general purpose method for constructing a Liapunov function for recurrent networks in general and show several specific cases to show the effectiveness of the formalism. We also compare the stability criterion obtained from Hamiltonian formalism with ordinary differential equation approach. Towards the end of the paper we show how such general purpose construct can be used to generate predictions in actual biological systems.

II. HAMILTONIAN FUNCTION FOR GENERAL RECURRENT NETWORK

A. Cohen-Grossberg Liapunov function derived from Hamiltonian formalism

We begin by considering a single layer of fully connected recurrent neural nodes. The activation of i -th node is given by x_i . If we consider a general recurrent shunting networks with dynamics given by¹

$$\dot{x}_i = -A_i x_i + (B_i - x_i)(I_i + \mathcal{S}(x_i)) - (x_i + C_i) \left(J_i + \sum_{j \neq i} w_{ji} \mathcal{S}(x_j) \right) \quad (1)$$

where I and J are excitatory and inhibitory inputs to the node i and \mathcal{S} is a sigmoid function. B and $-C$ are constants determining the upper and lower bound for the network activation respectively. We can transform Eq. 1 with simple variable change ($y_i = x_i + C_i$) to the form

$$\dot{y}_i = y_i \left(b_i(y_i) - \sum_{j=1}^n w_{ji} \mathcal{S}(y_j - C_j) \right) \quad (2)$$

where,

$$b_i(y_i) = \frac{1}{y_i} [A_i C_i - (A_i + J_i) y_i + (B_i + C_i - y_i) (I_i + \mathcal{S}(y_i - C_i))] \quad (3)$$

¹In this paper we will ignore the noise terms in the analysis.

R. Sengupta is with the Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, 4700, Keele Street, North York, Ontario, CA, PIN - M3J1P3. e-mail: rakesh@cse.yorku.ca

Anindya Pattanayak is with Visva-varati University. email: pattanayanindya07@gmail.com

Raju Surampudi Bapi is with Cognitive Science Lab, International Institute of Technology, Gachibowli, Hyderabad Telangana 500032 and School of Computer and information Sciences, University of Hyderabad, Hyderabad Telangana 500046. e-mail: raju.bapi@iiit.ac.in

As x_i is the activation of a particular node i of a recurrent network with n nodes, the general time evolution of all Cohen-Grossberg systems can be written as

$$\frac{dx_i}{dt} = a_i(x_i) \left[b_i(x_i) - \sum_{j=1}^n c_{ij} d_j(x_j) \right] \quad (4)$$

where c_{ij} are symmetric coefficient. The above formulation is in fact quite general and the equation can be used for both additive and shunting model networks, continuous-time McCulloch-Pitts model, Boltzmann machines, mean field model among others².

The global Liapunov function used by Cohen and Grossberg is given by [5], [6]

$$V = - \sum_{i=1}^n \int_0^{x_i} b_i(\xi_i) d'_i(\xi_i) d\xi_i + \frac{1}{2} \sum_{j,k=1}^n c_{jk} d_j(x_j) d_k(x_k) \quad (6)$$

The steady-state or equilibrium solution in the time evolution of the network implies that the activations of the nodes will be unable to influence each other in the long-term, i.e., all the local instabilities will decay. It allows us to consider the set $\{x_i\}$ as a set of generalized co-ordinates describing the state of the network. It follows then that we can use the Hamiltonian principle from classical mechanics and say

$$\frac{\partial x_i}{\partial t} \propto \frac{\partial H_i}{\partial \dot{x}_i} \quad (7)$$

$$\frac{\partial \dot{x}_i}{\partial t} \propto - \frac{\partial H_i}{\partial x_i} \quad (8)$$

where $\dot{x}_i = \frac{dx_i}{dt}$. If the proportionality constants in Eq. 7 and 8 are unequal (the equality leading to a trivial case), the Hamiltonian for a particular node i will follow

$$dH_i = \frac{\partial H}{\partial \dot{x}_i} d\dot{x}_i + \frac{\partial H}{\partial x_i} dx_i \propto \dot{x}_i d\dot{x}_i \quad (9)$$

In the above we have used the identity $\dot{x}_i d\dot{x}_i = \ddot{x}_i dx_i$, anticipating the final derivation.

Using Eq. 4, we have

$$d\dot{x}_i = \left[a_i(x_i) [b'_i(x_i) - c_{ii} d'_i(x_i)] + \frac{a'_i(x_i)}{a_i(x_i)} \dot{x}_i \right] dx_i \quad (10)$$

$$\begin{aligned} \dot{x}_i d\dot{x}_i &= (a_i(x_i))^2 [b'_i(x_i) b_i(x_i) \\ &- c_{ii} b_i(x_i) d'_i(x_i) + c_{ii} \sum_{j=1}^n c_{ij} d'_i(x_i) d_j(x_j) \\ &- \sum_{j=1}^n c_{ij} b'_i(x_i) d_j(x_j)] dx_i + \mathcal{O}(\dot{x}_i^3) \end{aligned} \quad (11)$$

where $\mathcal{O}(\dot{x}_i^3)$ represents terms that are of the order to \dot{x}_i^3 .

²For example, continuous-time Hopfield networks with network model [8]

$$\frac{dx_i}{dt} = -\lambda_i x_i(t) + \sum_j c_{ij} d_j(x_j) + I_i \quad (5)$$

is a special case of Eq. 4 with $a_i(x_i(t)) = 1$ and $b_i(x_i(t)) = -\lambda_i x_i(t) + I_i$

Near equilibrium ($\dot{x}_i \rightarrow 0$), we can safely ignore the terms $\mathcal{O}(\dot{x}_i^3) = \left(\frac{a'_i(x_i)}{a_i(x_i)} \right) \dot{x}_i^3 dt$ and also as $\dot{x}_i \rightarrow 0$, we have $b_i(x_i) \rightarrow \sum c_{ij} d_j(x_j)$ and thus the first and last terms in the sum within parenthesis cancel each other. Ignoring multiplicative coefficients, we can say that near equilibrium,

$$\sum_i^n dH_i \propto - \sum_i^n b_i(x_i) d'_i(x_i) dx_i + \frac{1}{2} \sum_{i,j}^n c_{ij} d(d_i(x_i) d_j(x_j)) \quad (12)$$

and thus the full Hamiltonian for the system can be written as (with some changes in the dummy indices),

$$\begin{aligned} H &= \sum_i^n \int dH_i \\ &\propto - \sum_i^n \int_0^{x_i} b_i(\xi_i) d'_i(\xi_i) d\xi_i + \frac{1}{2} \sum_{j,k=1}^n c_{jk} d_j(x_j) d_k(x_k) \end{aligned} \quad (13)$$

Thus we can show that the Cohen-Grossberg Liapunov function is a special case of and can be derived from general Hamiltonian formalism. This has a far reaching consequence for neural networks. As Eq. 4 is applicable for a large variety of neural networks, we can now derive energy functions based on the first principles rather than inspired guesses.

In the next section we will show certain special cases of the general formalism and how stability criteria could be found for several different kinds of recurrent networks. We also compare the stability criterion from the Hamiltonian with stability criterion derived from divergence tests.

B. Coupled oscillatory brain network from the Hamiltonian formalism

Another interesting contribution of the Hamiltonian formalism is its application towards oscillatory neural theories. From Eq. 4 we can say that

$$\begin{aligned} \dot{x}_i^2 &= (a_i(x_i))^2 [b_i(x_i)^2 - 2 \sum_{j=1}^n b_i(x_i) c_{ij} d_j(x_j) \\ &+ \sum_j^n \sum_k^n c_{ij} c_{ik} d_j(x_j) d_k(x_k)] \end{aligned} \quad (14)$$

Since near equilibrium, we have $b_i(x_i) \rightarrow \sum c_{ij} d_j(x_j)$, we can write,

$$(a_i(x_i))^2 \left(\sum_j^n \sum_k^n c_{ij} c_{ik} d_j(x_j) d_k(x_k) \right) = \dot{x}_i^2 + \mathcal{F}_1(x_1, \dots, x_n) \quad (15)$$

If $\mathcal{C} = ||c_{ij}||$ is a symmetric matrix ($c_{ij} = c_{ji}$), if the following property is satisfied

$$\sum_i^n c_{ij} c_{ik} = \sum_i^n c_{ji} c_{ik} = m_i c_{jk} \quad (16)$$

then m_i are the diagonal elements of the diagonal matrix D which satisfies the following matrix relation

$$\mathcal{C}^2 = DC \quad (17)$$

So using the multiplicative coefficients in Eq. 11, and using Eq. 15, 16 and 13, we can write the full Hamiltonian as the following

$$H = \frac{1}{2} \sum_{i=1}^n m_i \dot{x}_i^2 + \mathcal{F}_2(x_1, \dots, x_n) \quad (18)$$

where m_i are proportionality constants. Interestingly, the form of the Hamiltonian given in Eq: 18 is the Hamiltonian for multiple coupled nonlinear oscillators with equation of motion of the form [9]

$$m_i \ddot{x}_i + f_i(x_1, \dots, x_n) = 0 \quad (19)$$

It is interesting that for the Hamiltonian to generate meaningful oscillatory solution, we need transient input to the system as well as a period where the network will be allowed to settle or stabilize. Thus for oscillatory dynamics to be biologically feasible in the brain we need a recurrent layer to get transient input from a feed-forward network. The above formulation gives very important insights and constraints of deriving oscillatory models of the brain function.

III. SPECIAL CASES OF HAMILTONIAN FORMALISM

A recurrent shunting network with the range $[-D, B]$ follows the dynamics

$$\begin{aligned} \frac{dx_i}{dt} &= -Ax_i + (B - x_i)f(x_i) \\ &- (D + x_i) \sum_{k=1, k \neq i}^N f(x_k) + I_i \end{aligned} \quad (20)$$

A general additive recurrent network is given by

$$\frac{dx_i}{dt} = -\lambda x_i + \alpha F(x_i) - \beta \sum_{j=1, j \neq i}^N F(x_j) + I_i \quad (21)$$

For the additive network we assume a decay constant of λ .

A. Additive recurrent network with slower than linear activation function

Let us assume that in Eq. 21,

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x}{1+x} & \text{for } x > 0 \end{cases} \quad (22)$$

The network should reach steady state activity when the external input is taken away. If we disregard noise, at steady state, i.e. when, $\frac{dx_i}{dt} = 0$,

$$\lambda x_i = \alpha F(x_i) - \beta \sum_{j=1, j \neq i}^N F(x_j) \quad (23)$$

As the equation is symmetric under permutation of units, the system should have symmetric solutions characterized by

number of active units n , and their activation $x(n)$, all other units having 0 activation.

$$x(n) = \left(\frac{\alpha - (n-1)\beta}{\lambda} \right) F(x(n)) \quad (24)$$

Using Eq. 22, we get

$$x(n) = \left(\frac{\alpha - (n-1)\beta}{\lambda} \right) - 1 \quad (25)$$

Noise can bring in additional fluctuation that can destabilize the solution for a pair of active modes (with equal activation according to Eq. 25), unless the difference of activations between the said nodes $\Delta x = x_i - x_j$ decays. Using Eq. 21 & 25 we get

$$\frac{d\Delta x}{dt} = \Delta x \left[-\lambda + \lambda^2 \left(\frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} \right) \right] \quad (26)$$

Thus the fluctuation decays only if $\frac{d\Delta x}{dt} \leq 0$, i.e.,

$$\frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} \leq \frac{1}{\lambda} \quad (27)$$

As we can see that the decay parameter, excitation parameter and inhibition parameter are not completely independent for stable solutions. For the present purposes we use $\lambda = 1$ (we will keep using this value for the rest of the paper for simplicity).

From Eq. 21 and 22 we have

$$d\dot{x}_i = -dx_i + \alpha \left(\frac{F(x_i)}{x_i} \right)^2 dx_i \quad (28)$$

It is easy to show that from the definition given in 9,

$$H = \sum_i H_i \propto - \sum_i \int \left(1 - \alpha \left(\frac{F(x_i)}{x_i} \right)^2 \right) \dot{x}_i^2 dt \quad (29)$$

From Eq. 25 we can substitute terms in steady state to get

$$H \propto - \sum_i \int \left(1 - \frac{\alpha}{(\alpha - (n-1)\beta)^2} \right) \dot{x}_i^2 dt \quad (30)$$

Now if $dH < 0$ and thus a monotonically decreasing Liapunov type function in absence of external input, we have the stability condition as

$$\frac{\alpha}{(\alpha - (n-1)\beta)^2} \leq 1 \quad (31)$$

Comparing this to Eq. 27, we can see that the conditions derived from the energy value is slightly different and diverges greatly for higher β . This is due to the fact that Eq. 27 excludes winner-take-all mechanisms operating at higher inhibition, whereas the energy function does not. And it is evident that for all β ,

$$\frac{\alpha}{(\alpha - (n-1)\beta)^2} \leq \frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} \leq 1 \quad (32)$$

and thus the energy function is a very suitable candidate for the network as it is in line with the stability analysis derived from the dynamics of the network.

B. Shunting recurrent network with constant activation function

Here we assume in Eq. 20,

$$f(x) = \begin{cases} 0 & x \leq 0 \\ k & x > 0 \end{cases} \forall k \in \mathbb{R} \quad (33)$$

A search for a symmetric solution at steady state leads to steady state activation value

$$x(n) = \frac{(B - D(n-1))f(x(n))}{A + nf(x(n))} = \frac{(B - D(n-1))k}{A + nk} \quad (34)$$

The stability criterion is determined considering the decay of $\Delta x = x_i - x_j$ and is calculated to the following condition

$$nk \leq -A \quad (35)$$

Constructing the energy function using

$$d\dot{x}_i = \left(-A - \sum_{j=1}^N f(x_j) \right) dx_i \quad (36)$$

and Eq. 9, we have

$$H = \sum_i H_i \propto - \sum_i \int \left(A + \sum_{j=1}^N f(x_j) \right) \dot{x}_i^2 dt \quad (37)$$

As $\sum_{j=1}^N f(x_j) = nk$ near steady state, we have the same criterion for stability from the energy function as Eq. 35.

C. Shunting network with linear activation function

The activation function for such networks is given by

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \quad (38)$$

Using the above activation function in Eq. 20 we have at steady state condition,

$$x(n) = \frac{(B - D(n-1))f(x(n))}{A + nf(x(n))} = \frac{(B - D(n-1))x(n)}{A + nx(n)} \quad (39)$$

This leads to the non-trivial solution

$$x(n) = \frac{B - A - D(n-1)}{n} \quad (40)$$

Looking at the decay of $\Delta x = x_i - x_j$ leads to the stability criterion,

$$D \leq 0 \quad (41)$$

The Hamiltonian constructed using

$$d\dot{x}_i = -dx_i + (B - x_i)dx_i - \sum_{j=1}^N f(x_j)dx_i \quad (42)$$

and Eq. 9 leads to,

$$B - A - D(n^2 - 1) < 0 \quad (43)$$

which upon substituting $B - A = nx(n) + D(n-1)$ (from Eq.40) yields $nx(n) - nD(n-1) < 0$. For the situation where total network activation $nx(n) \geq 0$, we have $D \geq 0$ for $n > 1$. In fact this leads to a better stability criterion combining the stability criterion given in Eq. 41 with the condition $D \geq 0$ obtained from the Hamiltonian

$$D = 0 \quad (44)$$

This is a sensible result as by definition of shunting networks B and D are positive real quantities.

D. Shunting network using reciprocal activation function

The activation function for such networks is given by

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{x} & x > 0 \end{cases} \quad (45)$$

The steady state solution (assuming the positive root of a quadratic equation) is given by,

$$x(n) = \frac{-n + k}{2A} \quad (46)$$

where $k = \sqrt{n^2 + 4A(B - D(n-1))}$.

The stability criterion from decay condition is derived to be

$$\frac{4(B + D) + 2n(-n + k)}{(-n + k)^2} \geq -1 \quad (47)$$

Constructing the energy function in the now familiar way, we get

$$\frac{(2B + n - k)2 + 2n(-n + k)}{(-n + k)^2} \geq -1 \quad (48)$$

It can be shown that both the criteria given by Eq. 47 and 48 are same if $n - k = 2D$.

IV. ANALYTICAL PREDICTION OF WINNER-TAKE-ALL

In section III-A we have shown that there are constraints on the stability of the network response in an additive recurrent model, i.e., the condition that allows for the activation of two nodes to not diverge during the simulation. The constraint was given by

$$\frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} < 1 \quad (49)$$

where n is the number of nodes active after the simulation is run. We also obtained another similar stability condition after consideration of the novel Hamiltonian formulation, mainly,

$$\frac{\alpha}{(\alpha - (n-1)\beta)^2} < 1 \quad (50)$$

However, for a very important behavior of the network, the winner-take-all dynamics requires activation differences between nodes to be amplified until only one node emerges as winner. Now if $n = 1$, as in standard WTA interaction, Eq. 50 transforms into

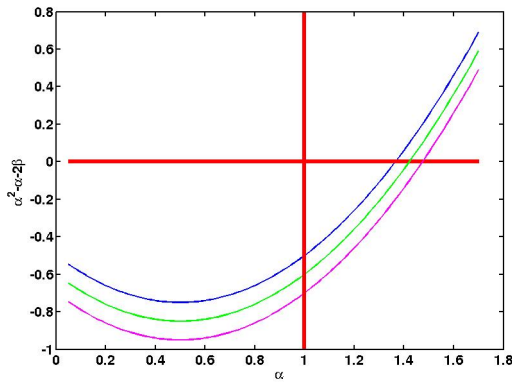


Fig. 1. A plot of $\alpha^2 - \alpha - 2\beta$ for different β values. Blue line for $\beta = 0.25$, green for $\beta = 0.3$ and magenta for $\beta = 0.35$. In the bottom right quadrant of the plot bound by the lines $\alpha^2 - \alpha - 2\beta = 0$ and $\alpha = 1$ we get the α values desired for WTA interaction, mainly $\alpha \geq 1$ and $\alpha^2 - \alpha - 2\beta \leq 0$. For $\beta = 0.25$ the range is $1 \leq \alpha \leq 1.37$, for $\beta = 0.3$, $1 \leq \alpha \leq 1.42$, for $\beta = 0.35$, $1 \leq \alpha \leq 1.47$.

Parameter	Value
N	10
α	0.5:0.01:1.5
β	0.3
No. of nodes receiving input	2
Total duration of simulation in time steps	2500

TABLE I
SIMULATION PARAMETERS

$$\alpha - 1 > 0 \quad (51)$$

This gives the lower limit of α for WTA. The upper limit will be set by Eq. 49, i.e. when $n = 2$ reaches stability, i.e.,

$$\alpha^2 - \alpha - 2\beta > 0 \quad (52)$$

Thus WTA behavior will be supported by a range of α that satisfies the conditions $\alpha \geq 1$ and $\alpha^2 - \alpha - 2\beta \leq 0$. This is shown in Fig. 1. It shows how the α ranges should be calculated for different β values. For $\beta = 0.25$ the range is $1 \leq \alpha \leq 1.37$, for $\beta = 0.3$, $1 \leq \alpha \leq 1.42$, for $\beta = 0.35$, $1 \leq \alpha \leq 1.47$.

We also confirmed the parameters with actual simulation. We chose a network of 10 nodes and gave inputs to 2 nodes for each simulation. The input level was clamped at 0.3. The probability of WTA interaction was calculated as the fraction of time out of 1000 simulations, that only one node survives. We varied the α between 0.5 and 1.5. Both stimuli were presented for 255 time steps each and the total duration of simulation was 2500 time steps. Noise was sampled from normal distribution of mean 0 and standard deviation 0.1. The results are shown in Fig. 2. The analytic limits obtained in Fig. 1 are closely confirmed by the simulation as well. The specific parameter values for the simulation are given in Table I.

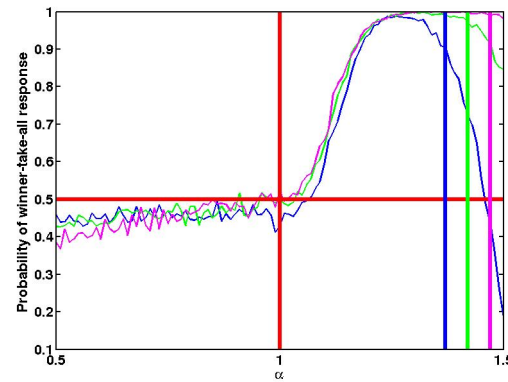


Fig. 2. Probability of winner take all interaction plotted against α for three different β values 0.25 (blue line), 0.30 (green line) and 0.35 (magenta line). For $\alpha > 1$, the WTA probability increases beyond chance level (0.5). The analytical limits on α obtained in Fig. 1 are shown in correspondingly colored lines.

V. APPLICATION OF WTA IN TEMPORAL ODDBALL EXPERIMENTS IN PSYCHOLOGY

Subjective expansion of time (TSE) is the phenomenon where human beings perceive an oddball stimulus in a stream/series of identical standard stimuli to have lasted longer in terms of percept duration [10]. There are several competing theories that have tried to explain this phenomenon. However, they can be roughly classified into - a) dedicated internal clock b) information based decision mechanism [11]–[13]. Here we have used a computational model to explore the hypothesis that recurrent on-center off-surround network [3] (with the hypothesis that oddball is more perceptually salient than standard stimuli) executing a winner-take-all decision can account for the oddball effect.

In the context of perceptual decision making, a winner-take-all mechanism is a prime candidate for a biologically plausible neuro-computational approach [14]. In case of temporal oddball paradigm, we observed that the experiment follows a two-alternative forced choice (2AFC) task. The 2AFC task led to the idea that time's subjective expansion factor observed in temporal oddball paradigm might arise from a flexible decision boundary arising from a neural WTA mechanism.

The on-center off-surround architecture we have used here has been described in more detail in [3], [15]. Here we will just mention the salient points. The differential equation governing the time-evolution of the network of N nodes is given by

$$\frac{dx_i}{dt} = -x_i + \alpha F(x_i) - \beta \sum_{j=1, j \neq i}^N F(x_j) + I_i + noise \quad (53)$$

$x_i(t)$ is the activation of node i at time t . I_i represents the intensity of external input ($\forall i, 0 \leq I_i \leq 1$), it is zero if the stimulus is absent for a particular node at that time point. Input is only presented for a finite amount of time, typically much less than total time of simulation. $F(x)$ is the activation function given by the formula,

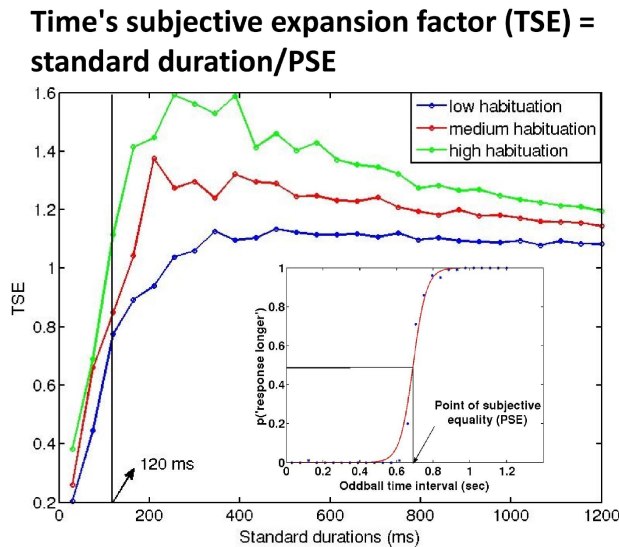


Fig. 3. TSE values for different standard durations at different levels of oddball input. Blue line has oddball input at 0.3, red line has oddball input at 0.4, green line has oddball input at 0.5. The different levels were interpreted as arising from different levels of habituation, i.e., if the oddball appears in a stream of 6 standard stimuli, its salience would be less than that appearing in a stream of 11 stimuli, and thus the former will have lower stimulus level (see appendix). The inset shows an example psychophysical curve generated in order to calculate the point of subjective equality.

Parameter	Value
N	10
α	1.05
β	0.3
No. of nodes receiving input	2
Total duration of simulation in time steps	2500

TABLE II
SIMULATION PARAMETERS

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x}{1+x} & \text{for } x > 0 \end{cases} \quad (54)$$

The network has a stable dynamics, i.e., the activated states at equilibrium do not diverge from the steady state dynamics if

$$\frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} \leq 1 \quad (55)$$

Considering a Liapunov function for energy for the same network we get another stability criterion

$$\frac{\alpha}{(\alpha - (n-1)\beta)^2} \leq 1 \quad (56)$$

Considering these two we get a condition for winner-take-all (WTA) interaction for a range of α that satisfies the conditions $\alpha \geq 1$ and $\alpha^2 - \alpha - 2\beta \leq 0$. After some rigorous convergence tests we used the following modeling parameters In a standard oddball task for temporal judgements involve showing a participant a series of frequent standard stimuli followed by infrequent oddball stimulus. A few standard frequent stimuli might also be shown after oddball presentation. The participant makes a judgement whether the oddball stimulus was longer in duration than the standard or not. Our model

assumes that a normalized input is given to two nodes of varying durations simultaneously as the duration judgements are post-event judgements in temporal oddball, it is not such an unreasonable assumption.

Now we ran 100 simulations for each duration pairs (30 vs 660 ms steps, 75 vs. 600 ms steps,... 1200 vs. 660 ms steps) for the two nodes and calculated the probability of response that one duration will be judged longer than the other by counting the fraction of times the node with the variable input is the winner of WTA interaction. In reality it is rare that both the units to be excited with the same level of input. It is more likely that the salience of the inputs differ. The standard stimuli is more frequent and thus prone to habituation and should receive lesser level input than the oddball stimuli. So the temporal oddball should correspond to the case where the standard interval gets a clamped input at level 0.2 and oddball interval gets clamped input at 0.3.

Now we calculated the values for the probability that oddball duration would be perceived as longer for each of the standard durations in the set (30:45:1200) (to be given input of 0.2). The oddball durations are also (30:45:1200) (to be given input 0.3). We calculated the points of subjective equality (PSE) through sigmoid fits and time's subjective expansion by dividing the standard durations by corresponding PSE's. The result is given in Fig. 3. The overall pattern of the results is strikingly close to TSE pattern observed in experimental results given in [10].

VI. DISCUSSION

In the current work we have shown how analytical Hamiltonian formalism developed by the authors can derive Cohen-Grossberg Liapunov function for general single-layer shunting recurrent neural networks. This has a far reaching consequence the dynamics of such networks can be used to model a wide variety of neural networks including additive recurrent networks, continuous-time McCulloch-Pitts neurons, Boltzman machines, and Mean field models, etc.

We have also shown how the analytic stability criteria derived from the Hamiltonian function fares against the stability criterion emerging from the network analysis for a wide range of activation functions, as well as for both additive and shunting networks. The agreement between the two stability criteria is indeed quite close.

For the additive variant, the stability criterion derived from the Hamiltonian has the advantage of predicting onset of Winner-take-all (WTA) behavior in the additive recurrent networks. In fact the two stability criteria (derived from Hamiltonian and network analysis) appear to complement each other. We have also shown how the WTA decision making process be used to predict interesting consequences for studies involving subjective expansion of time in temporal oddball paradigm in psychology.

Furthermore in our previous work [3], we have shown how the Hamiltonian function for the network can be used to derive predictions for psychophysical attributes like reaction times in constructions of biological recurrent networks for humans. We used the Hamiltonian successfully to explain

reaction time distributions for visual sense of numbers. To derive the prediction for reaction time (RT) we used only the assumption, $RT \propto -H$, i.e., reaction time should be correlated with the energy that will be needed for resetting the network. Experimentally, the fMRI activation patterns for enumeration and visual working memory task were predicted from the model and subsequently verified through human experimentation [16].

In recent years the understanding of neural codes has provided us with insights that go beyond the concepts of rate coding and it is increasingly more commonplace to speak of temporal codes that use spike timing and phase information in order to transmit and process information reliably [17]–[20]. However, most of these attempts are to look at neural codes after the presentation of stimulus. Recently, some studies have looked at pre-stimulus brain states in MEG and EEG based studies and have found that it is possible to predict conscious detection of stimuli based on pre-stimulus oscillatory brain activity [21]–[24]. For instance in case of near threshold stimuli some researchers have found the pre-stimulus α frequency band modulation to be important [22]. These attempts have drawn a large amount of interest, but have revealed little towards a theoretical or physiological understanding of such phenomena. In section IIB, we have shown the possibility of oscillatory brain states under certain equilibrium conditions for a neural assemble consisting of both feed-forward and recurrent connections. Interestingly, we this leads to an important understanding that the delay between feed-forward and recurrent connections is very important towards understanding the distribution of oscillatory brain states. An alternative approach to oscillatory brain states has been explored in [25].

Overall, the current work not only brings a novel formalism to recurrent neural networks, but also shows way to connect the neural network properties for the purpose of predicting diverse psychophysical parameters in biological systems including but not limited to time perception, enumeration, working memory, oscillatory brain dynamics to name a few.

REFERENCES

- [1] M Usher and J D Cohen. Short term memory and selection processes in a frontal-lobe model. In *Connectionist models in cognitive neuroscience*, pages 78–91. Springer, 1999.
- [2] Rafal Bogacz, Marius Usher, Jiaxiang Zhang, and James L McClelland. Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 362(1485):1655–70, September 2007.
- [3] R Sengupta, S Bapi Raju, and D Melcher. A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network. *Brain Research*, 1582:114–124, 2014.
- [4] Alexander Andreopoulos and John K Tsotsos. 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding*, 117(8):827–891, 2013.
- [5] Michael A Cohen and Stephen Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE transactions on systems, man, and cybernetics*, (5):815–826, 1983.
- [6] Stephen Grossberg. Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural networks*, 1(1):17–61, 1988.
- [7] Huaguang Zhang, Zhanshan Wang, and Derong Liu. A comprehensive review of stability analysis of continuous-time recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7):1229–1262, 2014.
- [8] John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10):3088–3092, 1984.
- [9] Seher Durmaz, S Altay Demirbag, and MO Kaya. Hamiltonian Approach to Multiple Coupled Nonlinear Oscillators. *Acta Physica Polonica-Series A General Physics*, 121(1):47, 2012.
- [10] Peter Ulric Tse, James Intriligator, Josée Rivest, and Patrick Cavanagh. Attention and the subjective expansion of time. *Attention, Perception, & Psychophysics*, 66(7):1171–1189, 2004.
- [11] Åke Hellström. The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, 97(1):35, 1985.
- [12] Michel Treisman. Temporal discrimination and the indifference interval: Implications for a model of the “internal clock”. *Psychological Monographs: General and Applied*, 77(13):1, 1963.
- [13] Ewart AC Thomas and Wanda B Weaver. Cognitive processing and time perception. *Attention, Perception, & Psychophysics*, 17(4):363–367, 1975.
- [14] Kong-Fatt Wong and Xiao-Jing Wang. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.
- [15] Rakesh Sengupta, S Bapiraju, Prajit Basu, and David Melcher. Accounting for subjective time expansion based on a decision, rather than perceptual, mechanism. *Journal of Vision*, 14(10):1150–1150, 2014.
- [16] André Knops, Manuela Piazza, Rakesh Sengupta, Evelyn Eger, and David Melcher. A shared, flexible neural map architecture reflects capacity limits in both visual short-term memory and enumeration. *Journal of Neuroscience*, 34(30):9857–9866, 2014.
- [17] Garrett B Stanley. Reading and writing the neural code. *Nature neuroscience*, 16(3):259–263, 2013.
- [18] Rufin Van Rullen and Simon J Thorpe. Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural computation*, 13(6):1255–1283, 2001.
- [19] Timothée Masquelier. Relative spike time coding and STDP-based orientation selectivity in the early visual system in natural continuous and saccadic vision: a computational model. *Journal of computational neuroscience*, 32(3):425–441, 2012.
- [20] J Gautrais and S Thorpe. Rate coding versus temporal order coding: a theoretical approach. *BioSystems*, 48(1-3):57–65.
- [21] Kyle E Mathewson, Gabriele Gratton, Monica Fabiani, Diane M Beck, and Tony Ro. To see or not to see: prestimulus α phase predicts visual awareness. *The Journal of neuroscience*, 29(9):2725–2732, 2009.
- [22] Nathan Weisz, Anja Wühle, Gianpiero Monitola, Gianpaolo Demarchi, Julia Frey, Tzvetan Popov, and Christoph Braun. Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proceedings of the National Academy of Sciences*, 111(4):E417–E425, 2014.
- [23] Elisabeth S May, Markus Butz, Nina Kahlbrock, Nienke Hoogenboom, Meike Brenner, and Alfons Schnitzler. Pre-and post-stimulus alpha activity shows differential modulation with spatial attention during the processing of pain. *Neuroimage*, 62(3):1965–1974, 2012.
- [24] Julian Keil, Nadia Müller, Niklas Ihssen, and Nathan Weisz. On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cereb. Cortex*, 22(1):221–31, January 2012.
- [25] Rakesh Sengupta. Pre-stimulus oscillatory brain states and cognition: a theoretical approach. *arXiv preprint arXiv:1508.02257*, 2015.