

# Learning of Spatio–Temporal Codes in a Coupled Oscillator System

Gábor Orosz, Peter Ashwin, and Stuart Townley

**Abstract**—In this paper, we consider a learning strategy that allows one to transmit information between two coupled phase oscillator systems (called teaching and learning systems) via frequency adaptation. The dynamics of these systems can be modeled with reference to a number of partially synchronized cluster states and transitions between them. Forcing the teaching system by steady but spatially nonhomogeneous inputs produces cyclic sequences of transitions between the cluster states, that is, information about inputs is encoded via a “winnerless competition” process into spatio–temporal codes. The large variety of codes can be learned by the learning system that adapts its frequencies to those of the teaching system. We visualize the dynamics using “weighted order parameters (WOPs)” that are analogous to “local field potentials” in neural systems. Since spatio–temporal coding is a mechanism that appears in olfactory systems, the developed learning rules may help to extract information from these neural ensembles.

**Index Terms**—Adaptive learning, coupled oscillator system, heteroclinic network, spatio–temporal code, winnerless competition.

## I. INTRODUCTION AND MOTIVATION

NEURONS communicate with electrical and chemical signals with finite signal transmission times. These transmission delays may impede full synchronization but allow polysynchronization [10] leading to partially synchronized cluster states. In a variety of neural systems, an important mechanism at play is the use of spatio–temporal codes involving partially synchronized cluster states to represent information [28], [29]. In particular, there is biological evidence that such spatio–temporal codes may be an important part of the encoding of spatially nonhomogeneous inputs in olfactory systems such as the insect antennal lobe and mammalian olfactory bulb [22], [23], [27]. (By spatio–temporal code, we rather mean identity–temporal code since information depends on which neurons are acting rather than where they lie in physical space.)

A plausible nonlinear dynamical mechanism that robustly produces a spatio–temporal code is a process of winnerless competition between unstable cluster states [1], [3]. This paper

takes these ideas further; we study how a variety of different inputs can be converted into different spatio–temporal codes. Furthermore, we derive a non-Hebbian learning rule that allows one to transmit information between two systems (that are capable of winnerless competition) via synchronization and frequency adaptation. Our motivation for the proposed algorithm is to be able to extract information from a biological neural network (teaching system) by connecting it to an artificial neural network (learning system) that is able to learn spatio–temporal codes. The adaptive learning algorithm proposed in this paper is similar to the algorithm developed in [32] and it guarantees global convergence. This is in contrast to earlier biologically motivated examples where only local convergence was achieved; see, e.g., [30].

The winnerless competition dynamics of the neural network we consider is governed by a robust attracting heteroclinic network that is subject to spatially nonhomogeneous time-independent inputs (that detune the natural frequencies of oscillators) and background noise (that is uncorrelated in space and time). The heteroclinic network consists of partially synchronized cluster states of saddle type (the nodes) that are connected by their unstable manifolds (the edges). The graph structure of the heteroclinic network—designed by the dynamics—is implicit in the neural network itself but is much more complex [2]. Consequently, the time evolution consists of a sequence of quasi-steady residences near partially synchronized cluster states interspersed with fast transitions, called switches, between different cluster states. A spatio–temporal code is then a cyclic sequence of switches between cluster states, i.e., a cyclic path along the graph of the heteroclinic network. In this sense, the heteroclinic network acts as a skeleton for sequential switching dynamics. (From now on we use sequential switching as a synonym for winnerless competition.)

Since the studied dynamics is not reducible to any low-dimensional subspace there exist a large variety of cyclic sequences for even a small numbers of neurons. Applying different spatially nonhomogeneous time-independent inputs one may observe a large variety of different spatio–temporal codes (while the characteristic switching time depends on the magnitude of inputs). We will show that the spatio–temporal coding can be reduced to a combinatorial problem. Similar results may be obtained by reducing neural models to discrete dynamics [12], [31]. The obtained coding scenario is robust against small changes in initial conditions and against noise while it is sensitive to changes in input configurations. The dynamics is also robust against small changes in parameters, i.e., structurally stable.

When leaving the parameter domain of sequential switching, the heteroclinic structure may be “destroyed” leading to chaos [2]. Such edge-of-criticality behavior is claimed to be typical

Manuscript received April 30, 2008; revised November 21, 2008; accepted February 16, 2009. First published May 27, 2009; current version published July 09, 2009. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/C510771/1.

G. Orosz is with the Department of Mechanical Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA and also with the School of Engineering, Computing and Mathematics, University of Exeter, Exeter EX4 4QF, U.K. (e-mail: gabor@engineering.ucsb.edu).

P. Ashwin and S. Townley are with the School of Engineering, Computing and Mathematics, University of Exeter, Exeter EX4 4QF, U.K. (e-mail: p.ashwin@exeter.ac.uk; s.b.townley@exeter.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNN.2009.2016658

in neural systems [6], [7], [14], [20]. We remark that, beside heteroclinic structures, there are other proposed nonlinear dynamical mechanisms that can cause transitions between cluster states and so result in spatio-temporal codes. “Chaotic itinerancy” occurs when destabilized attractors (called “attractor ruins”) are interconnected by unstable manifolds [5], [13], [33]. Such attractor ruins are closely related to Milnor attractors [33], while interconnected Milnor attractors may become heteroclinic cycles when the noninvertibility of the underlying dynamical system is removed [15]. “Spatio-temporal chaos” may also lead to the appearance of (even more complicated) spatio-temporal patterns in simple coupled oscillator models [21] and in hierarchical brain models [18], [19]. Note that chaotic itinerancy and spatio-temporal chaos are sensitive to initial conditions and noise in contrast to sequential switching.

In this paper, we introduce a weighted order parameter (WOP) to represent the state of the system. This is a scalar observable that distinguishes between different cluster states, meaning that the spatio-temporal code can be represented as a single time series. The WOP mimics the local field potential (LFP) for neural ensembles.

To transmit information about the code from one system to another, an appropriate learning rule has to be derived. We consider an idealized teaching system that has its frequencies detuned by the input and produces the corresponding spatio-temporal code. The derived learning rule forces the learning system to synchronize with the teaching one and adapt its frequencies to those of the teaching system. When the learning phase is completed, the learning system continues to shadow the teaching system. In this way, the learning system encodes information about the spatio-temporal code, i.e., about the input provided to the teaching system. This synchronize-and-adapt learning scenario is required since the spatio-temporal codes cannot be represented by simple point attractors in state space. Note that during the learning procedure only the intrinsic frequencies of oscillators are changed in the learning system while the couplings between those oscillators are left intact. In other words, a non-Hebbian learning algorithm is proposed in this paper.

The paper is organized as follows. In Section II, we introduce the coupled oscillator system consisting of two copies of the system discussed in [2] with coupling from the teaching to the learning system and adaptation of the frequencies in the learning system. The robust dynamics of the teaching system is described in Section III. The dynamics of the adaptive learning is discussed in Section IV. Here we give a proof of stability of the synchronized-adapted state, and then by numerical simulation, we demonstrate that learning can be done effectively even over periods shorter than one period of the spatio-temporal code. Learning is illustrated using a WOP. Finally, in Section V, we outline some of the consequences and outstanding problems of this approach.

## II. MODELING

We consider two coupled phase oscillator systems of the type studied in [2], each containing  $N$  oscillators: a *teaching system* with phases  $\theta = \text{col}[\theta_1, \dots, \theta_N]$  and a *learning system* with

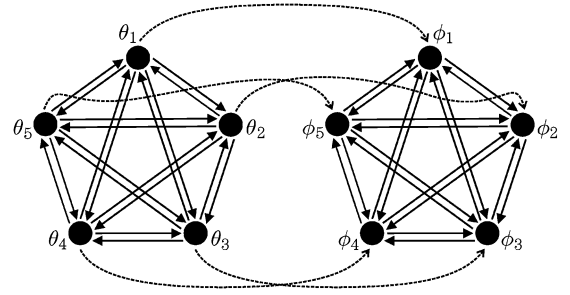


Fig. 1. Sketch of the teaching and learning systems in (1) when each system contains  $N = 5$  oscillators. Couplings inside a system are represented by solid arrows while couplings between systems are represented by dashed arrows.

phases  $\phi = \text{col}[\phi_1, \dots, \phi_N]$  that are coupled to each other. For each  $n = 1, \dots, N$  we have

$$\dot{\theta}_n = \Omega_n + \frac{1}{N} \sum_{m=1}^N g(\theta_n - \theta_m) + \eta \xi_n, \quad (\text{teach})$$

$$\dot{\phi}_n = \omega_n + \frac{1}{N} \sum_{m=1}^N g(\phi_n - \phi_m) + \eta \zeta_n + u(\theta_n, \phi_n), \quad (\text{learn})$$

$$\dot{\omega}_n = v(\theta_n, \phi_n), \quad (\text{adapt}) \quad (1)$$

where the dot denotes differentiation with respect to time  $t$  and  $\theta_n(t), \phi_n(t) \in [0, 2\pi)$ . The frequencies of the teaching system  $\Omega_n$  are time independent while the frequencies of the learning system  $\omega_n(t)$  will adaptively evolve in time. The quantities  $\xi_n(t), \zeta_n(t)$  stand for uncorrelated white noise such that the associated random walk has unit growth of variance per unit time. The noise is scaled by the noise strength  $\eta$ . In (1), the first  $N$  equations describe the time evolution of phases in the teaching system while the last  $2N$  equations describe the time evolution of phases and adaptation of frequencies in the learning system.

A sketch of the teaching and learning systems is depicted in Fig. 1 in case of  $N = 5$  where solid arrows represent the coupling  $g$  between oscillators inside a system while dashed arrows represent the adaptive couplings  $u$  and  $v$  between oscillators of two different systems. Note that  $g(\theta_n - \theta_m)$  represents the forcing of the  $n$ th oscillator by the  $m$ th oscillator in the teaching system while  $g(\phi_n - \phi_m)$  means the same in the learning system. Furthermore,  $u(\theta_n, \phi_n)$  represents the forcing of the  $n$ th oscillator in the learning system by the  $n$ th oscillator in the teaching system and  $v(\theta_n, \phi_n)$  governs the adaptation of the frequency of the  $n$ th oscillator in the learning system. We will specify the adaptive couplings  $u$  and  $v$  in Section IV.

In this paper, we consider the coupling function

$$g(\varphi) = -\sin(\varphi + \alpha) + r \sin(2\varphi + \beta) \quad (2)$$

and use the parameters

$$\Omega = 1.0 \quad r = 0.2 \quad \alpha = 1.8 \quad \beta = -2.0. \quad (3)$$

The  $2\pi$ -periodic coupling function can be obtained as a Fourier series by performing a phase reduction of a realistic neural model, e.g., Hodgkin–Huxley model with electrotonic or synaptic coupling [8], [34]. In (2), we consider a truncation of the coupling function up to second order that permits clustering

up to three clusters. Indeed, to study clustering that involves higher number of clusters, more harmonics need to be included [16], [17]. The parameters  $\Omega$ ,  $r$ ,  $\alpha$ , and  $\beta$  are originated from neural parameters. In particular, transmission delays between neurons alter the phase shifts  $\alpha$  and  $\beta$  [11]. It will be shown below that the heteroclinic network is hyperbolic and so the switching dynamics is robust to small changes in parameters (structurally stable). So one may vary the parameters around the values given in (3) and obtain the same qualitative dynamics. The finite parameter domain corresponding to switching dynamics changes as  $N$  is increased. However, such a domain can be found, so there exists structurally stable switching dynamics for large  $N$ , as will be demonstrated in Section IV-B.

### III. SWITCHING DYNAMICS BETWEEN CLUSTER STATES

In this section, we summarize the sequential switching dynamics through an isolated system when its frequencies are modulated by spatially nonhomogeneous constant inputs. That is, we consider the first  $N$  equations in (1)

$$\dot{\theta}_n = \Omega_n + \frac{1}{N} \sum_{m=1}^N g(\theta_n - \theta_m) + \eta \xi_n. \quad (4)$$

Note that the right-hand side only depends on phase differences, so it is sufficient to examine these to determine the behavior of the system.

We assume that the natural frequencies  $\Omega_n$  are close to the average frequency

$$\Omega = \frac{1}{N} \sum_{n=1}^N \Omega_n. \quad (5)$$

In particular, we consider that the frequencies are evenly distributed

$$\Omega_n = \Omega + p \left( I_n - \frac{N+1}{2} \right) \quad (6)$$

where  $p \ll \Omega$  and the set  $[I_1, \dots, I_N]$  is a permutation of  $[1, \dots, N]$ . The second term in (6) can be interpreted either as a detuning to the average frequency  $\Omega$  or as a constant external input added to a system of identical oscillators. For any *input configuration*  $[I_1, \dots, I_N]$  each oscillator receives a different constant stimulus of the order of the *input magnitude*  $p \ll \Omega$ . We will use the abbreviation: oscillator  $n$  receives stimulus  $I_n$  meaning that oscillator  $n$  receives stimulus  $p(I_n - (N+1)/2)$ . Note that there exist  $N!$  different input configurations. These configurations are special in the sense that the stimuli  $I_n$  are equidistant while the qualitative dynamics do not change for near-equidistant stimuli as will be demonstrated in Section IV-B.

For  $N = 2k + 1$  oscillators, considering  $p = 0$  and  $\eta = 0$  in (4) and (6), there exists a partially synchronized cluster state

$$[\theta_1(t), \dots, \theta_N(t)] = \tilde{\Omega}t[1, \dots, 1] + \underbrace{[y, \dots, y]}_k, \underbrace{[w, b, \dots, b]}_k \quad (7)$$

with constant frequency  $\tilde{\Omega}$  and phases  $y$ ,  $w$ , and  $b$ . Since  $\theta$  describes the phases of oscillators, (7) describes a periodic orbit in the state space of those oscillators. Along this periodic orbit, the

TABLE I  
LIST OF CLUSTER STATES FOR  $N = 5$  OSCILLATORS. (EACH BLOCK MAY BE GENERATED FROM ANOTHER BY CYCLIC PERMUTATION OF THE COMPONENTS)

$[y, y, w, b, b] = s_1$	$[w, b, b, y, y] = s_{19}$
$[b, b, w, y, y] = s_2$	$[w, y, y, b, b] = s_{20}$
$[y, b, w, y, b] = s_3$	$[w, y, b, y, b] = s_{21}$
$[b, y, w, b, y] = s_4$	$[w, b, y, b, y] = s_{22}$
$[y, b, w, b, y] = s_5$	$[w, b, y, y, b] = s_{23}$
$[b, y, w, y, b] = s_6$	$[w, y, b, b, y] = s_{24}$
$[b, y, y, w, b] = s_7$	$[y, w, b, b, y] = s_{25}$
$[y, b, b, w, y] = s_8$	$[b, w, y, y, b] = s_{26}$
$[b, y, b, w, y] = s_9$	$[b, w, y, b, y] = s_{27}$
$[y, b, y, w, b] = s_{10}$	$[y, w, b, y, b] = s_{28}$
$[y, y, b, w, b] = s_{11}$	$[b, w, b, y, y] = s_{29}$
$[b, b, y, w, y] = s_{12}$	$[y, w, y, b, b] = s_{30}$
$[b, b, y, y, w] = s_{13}$	
$[y, y, b, b, w] = s_{14}$	
$[y, b, y, b, w] = s_{15}$	
$[b, y, b, y, w] = s_{16}$	
$[b, y, y, b, w] = s_{17}$	
$[y, b, b, y, w] = s_{18}$	

oscillators are frequency synchronized with frequency  $\tilde{\Omega}$  and the clustering is due to the phases  $y$ ,  $w$ , and  $b$ . In this paper, we will use the abbreviated names  $b$  cluster/oscillators,  $y$  cluster/oscillators, and  $w$  oscillator. One may substitute the cluster state (7) into (4) and the obtained algebraic equations determine the frequency  $\tilde{\Omega}$  and the phase differences  $b - w$  and  $y - w$  that are unique for parameters (3); see (25) in part A of the Appendix. Furthermore, there are additional cluster states with different cluster configurations obtained by permuting the components of  $[y, \dots, y, w, b, \dots, b]$ . (There are  $N!/(k!1!k!)$  cluster states.) Table I lists all (30) states for  $N = 5$  oscillators so that the term  $\tilde{\Omega}t[1, \dots, 1]$  is not spelled out (and this convention is adopted in the remainder of this paper).

One may linearize (4) about the cluster state (7) (or about any of its symmetrical copies) and calculate the corresponding eigenvalues  $\lambda_1, \dots, \lambda_5$  and the eigenvectors  $v_1, \dots, v_5$ ; see (26)–(31) in part A of the Appendix. For parameters (3), there are eigenvalues on both sides of the imaginary axis such that

$$\lambda_1 = 0 \quad \lambda_2, \text{Re}\lambda_4, \text{Re}\lambda_5 < 0 \quad 0 < \lambda_3 < \max_{j \in \{2,4,5\}} |\lambda_j| \quad (8)$$

that is, the above cluster states are saddle type and the most attracting eigenvalue is stronger than the repelling one. [Notice that  $\lambda_2$  and  $\lambda_3$  are real while  $\lambda_4$  and  $\lambda_5$  can be real or a complex conjugate pair; see (26) and (27).] Since  $\text{Re}\lambda_4 < 0$  and  $\text{Re}\lambda_5 < 0$ , the cluster states are stable in the subspace spanned by  $v_4$  and  $v_5$  where both the  $y$  and the  $b$  clusters remain synchronized. Since the eigenvalue  $\lambda_2 < 0$ , the eigenvector  $v_2$  shows that the  $y$  cluster is “stable,” that is, perturbations desynchronizing the  $y$  cluster decay in time. The  $k$  different vectors appearing in  $v_2$  show that there are  $k$  different perturbations depending on which  $y$  oscillator has its phase advanced with respect to the other  $y$  oscillators. Since the eigenvalue  $\lambda_3 > 0$ , the eigenvector  $v_3$  shows that the  $b$  cluster is “unstable,” that is, perturbations desynchronizing the  $b$  cluster grow in time and lead the system away from the cluster state in state space. The  $k$  vectors appearing in  $v_3$  correspond to the  $k$  different perturbations that can be applied to the  $b$  cluster.

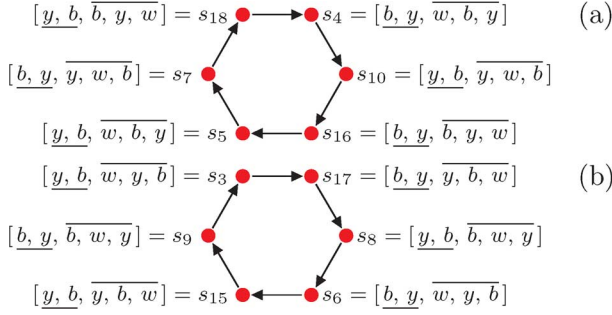


Fig. 2. Two six-cycles (spatio-temporal codes) in case of  $N = 5$  oscillators for input configuration  $[I_1, I_2, I_3, I_4, I_5] = [1, 2, 3, 4, 5]$ . At each switch, the phase of the  $b$  oscillator receiving the larger stimulus changes to  $w$ , the phase of the other  $b$  oscillator changes to  $y$ , and the phases of the  $w$  and  $y$  oscillators change to  $y$  and  $b$ , respectively. Thus, the underlined oscillators (receiving stimuli  $[1, 2]$ ) swap their phases  $y$  and  $b$ , while the overlined oscillators (receiving stimuli  $[3, 4, 5]$ ) cyclically permute their phases  $b, y$ , and  $w$ .

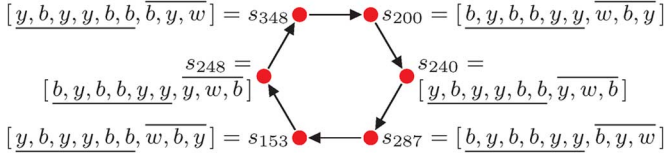


Fig. 3. One of the 20 six-cycles (spatio-temporal codes) in case of  $N = 9$  oscillators for the input configuration  $[I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9] = [1, 2, 3, 4, 5, 6, 7, 8, 9]$ . At each switch, the phase of the  $b$  oscillator receiving the largest stimulus changes to  $w$ , the phase of the other  $b$  oscillators changes to  $y$ , and the phases of the  $w$  and  $y$  oscillators change to  $y$  and  $b$ , respectively. Thus, the underlined oscillators (receiving stimuli  $[1, 2, 3, 4, 5, 6]$ ) swap their phases  $y$  and  $b$ , while the overlined oscillators (receiving stimuli  $[7, 8, 9]$ ) cyclically permute their phases  $b, y$ , and  $w$ .

Following the unstable directions in  $v_3$  for each cluster state (by numerical simulation) shows that the cluster states are connected by their unstable manifolds to form an attracting heteroclinic network. This network governs the switching dynamics between cluster states, and it can be represented by a graph where nodes represent cluster states and directed edges represent switches. Each node has  $k$  incoming edges (corresponding to the  $k$  different vectors in  $v_2$ ) and  $k$  outgoing edges (corresponding to the  $k$  different vectors in  $v_3$ ). The state reached after a switch is determined by which  $b$  oscillator has its phase advanced with respect to the other  $b$  oscillators in the current state. The phase of this oscillator changes to  $w$ , the phase of the other  $b$  oscillators changes to  $y$ , and the phases of the  $w$  and  $y$  oscillators change to  $y$  and  $b$ , respectively; see switches in Figs. 2 and 3. More details about the simplest nontrivial case of  $N = 5$  oscillators can be found in [2].

Note that one may find parameter domains where

$$\lambda_1 = 0 \quad \lambda_3, \text{Re}\lambda_4, \text{Re}\lambda_5 < 0 \quad 0 < \lambda_2 < \max_{j \in \{3, 4, 5\}} |\lambda_j| \quad (9)$$

[cf., (8)]. In this case, the appearing switching dynamics remains qualitatively similar except that the roles of  $y$  and  $b$  groups are swapped. Also note that having one of the eigenvalue configurations (8) or (9) is necessary but not sufficient for the existence of the attracting heteroclinic network that also requires the appropriate connections by the unstable manifolds.

Considering (4) and (6) for  $p = 0$  and  $\eta = 0$  with eigenvalues (8), the time interval between switches increases exponentially

with exponent  $\ln(\lambda_m/\lambda_3)$ , where  $\lambda_m = \max_{j \in \{3, 4, 5\}} |\lambda_j|$ , that is, oscillations slow down. For  $p = 0$  and  $\eta > 0$ , each switch is randomly chosen by the noise, thus a random walk along the heteroclinic network is observed. In this case, there exists a characteristic switching period  $T_s \approx -(\ln \eta)/\lambda_3$ . For  $p > 0$  and  $\eta = 0$ , the input configuration  $[I_1, \dots, I_N]$  determines in each state which  $b$  oscillator's phase is advanced the most, i.e., determines the switch. Consequently, each node has only one outgoing edge. The heteroclinic network reduces to cyclic paths of six switches (six-cycles) that are approached after a finite number of transient switches. These cyclic paths correspond to stable limit cycles in state space. Which six-cycle is approached, and via which transient path, depends on the initial condition (the transients are not discussed in detail in this paper). The characteristic switching period is now  $T_s \approx -(\ln p)/\lambda_3$ . Finally, for  $p > \eta > 0$ , the input dominates the dynamics, i.e., six-cycles persist, as will be demonstrated in Section IV-B.

We remark that along the cyclic paths each switch has a slightly different switching period. One may incorporate these time differences into the spatio-temporal code to give a one-to-one correspondence between the inputs and the codes [35]. In this paper, we consider relatively small signal-to-noise ratio so these small differences in switching period are not easily decodable. On the other hand, the coding scenario explained below is very robust against noise.

#### A. Six-Cycles as Spatio-Temporal Codes: Combinatorics of Encoding

The information about the input configuration  $[I_1, \dots, I_N]$  is encoded into the appearing six-cycles which are the spatio-temporal codes. Here we describe this coding scenario combinatorially. First, we discuss the simplest nontrivial case of  $N = 5$  oscillators. Fig. 2 shows the two six-cycles (spatio-temporal codes) for input configuration  $[1, 2, 3, 4, 5]$ . Notice that the underlined oscillators (receiving stimuli  $[1, 2]$ ) swap their phases  $y$  and  $b$  at each switch (a period two oscillation), and the overlined oscillators (receiving stimuli  $[3, 4, 5]$ ) cyclically permute their phases  $b, y$ , and  $w$  (a period three oscillation). Combining these period two and period three oscillations gives a cycle with minimal period six. (There are two different six-cycles since there are two different period three oscillations with cyclically permuted phases.)

In fact, the same spatio-temporal codes are obtained for all input configurations with permutations within the pair of small stimuli  $[1, 2]$  and within the triplet of large stimuli  $[3, 4, 5]$ . (There are  $2!3! = 12$  such cases.) The first row of the code table in Table II represents all these cases such that  $[a, b]$  and  $[X, Y, Z]$  represent the pair and the triplet, respectively. Permutations which mix elements of the pair and the triplet results in different spatio-temporal codes. Thus, the 120 input configurations are divided into  $120/12 = 10$  groups that are distinguished by which three oscillators receive the large stimuli  $[X, Y, Z]$ . Each group results in a different pair of spatio-temporal codes as shown in Table II.

In this way, all 20 possible six-cycles appear as spatio-temporal codes. Note that the total number of six-cycles can be calculated with the help of an adjacency matrix  $A$ . The elements of  $A$  are  $A_{ij} = 1$  if and only if there is a directed edge from

TABLE II  
CODE TABLE FOR  $N = 5$  OSCILLATORS. FOR ANY INPUT CONFIGURATION GROUP, THE PAIR  $[a, b]$  REPRESENTS ANY OF THE TWO PERMUTATIONS OF  $[1, 2]$  WHILE THE TRIPLET  $[X, Y, Z]$  REPRESENTS ANY OF THE SIX PERMUTATIONS OF  $[3, 4, 5]$ . THE SIX-CYCLES (a)–(f) ARE REPRESENTED AS TIME SERIES IN FIG. 4

input configuration group	spatiotemporal codes	
$[a, b, X, Y, Z]$		
$[a, X, b, Y, Z]$		
$[a, X, Y, b, Z]$		
$[a, X, Y, Z, b]$		
$[X, a, b, Y, Z]$		
$[X, a, Y, b, Z]$		
$[X, a, Y, Z, b]$		
$[X, Y, a, b, Z]$		
$[X, Y, a, Z, b]$		
$[X, Y, Z, a, b]$		

node  $s_i$  to  $s_j$  and  $A_{ij} = 0$ , otherwise. The number of cyclic paths of length  $l$  is given by  $\text{tr}(A^l)/l$ ; see [9]. The denominator  $l$  is needed since  $\text{tr}(A^l)$  counts each cycle  $l$  times (starting from different nodes of the cycle). Since  $\text{tr}(A^l)/l = 0$  for  $l = 2, 3, 4$  (there are no two-, three-, and four-cycles) the six-cycles are nonrepetitive.

The above coding strategy persists for larger numbers of oscillators with essentially the same combinatorial rules. In the general case, there are  $N!$  input configurations  $[I_1, \dots, I_N]$

TABLE III  
THE NUMBER OF SPATIALLY NONHOMOGENEOUS TIME-INDEPENDENT INPUTS AND SPATIO-TEMPORAL CODES FOR  $N = 2k + 1$  OSCILLATORS. NOTE THAT THE LAST ROW CAN ALSO BE CALCULATED AS THE PRODUCT OF THE TWO ROWS ABOVE IT

	$N = 2k + 1$	$N = 5$ $k = 2$	$N = 7$ $k = 3$	$N = 9$ $k = 4$
different cluster states	$\frac{N!}{k!1!k!}$	30	140	630
different input configurations	$N!$	120	5040	362880
different input configuration groups	$\binom{N}{3}$	10	35	84
different spatiotemporal codes for an input	$\binom{N-3}{k-1}$	2	6	20
different spatiotemporal codes	$\frac{\text{tr}(A^6)}{6}$	20	210	1680

which are classified into  $\binom{N}{3}$  groups distinguished by which three oscillators receive the largest stimuli  $[X, Y, Z]$  (now representing all six permutations of  $[N - 2, N - 1, N]$ ). Indeed, each group contains  $(N - 3)!3!$  input configurations. Furthermore, for a particular input configuration group, one may find  $\binom{N-3}{k-1}$  six-cycles (spatio-temporal codes) since there are  $(1/2)\binom{N-3}{k-1}$  period two oscillations (oscillators swap their phases  $y$  and  $b$  at each switch) and there are two period three oscillations (oscillators cyclically permute their phases  $b, y$ , and  $w$ ). Indeed, the total number of six-cycles  $\text{tr}(A^6)/6$  can be calculated by multiplying the number of input configuration groups with number of six-cycles for a particular input configuration. Table III summarizes these results and gives the corresponding numbers for  $N = 5, 7, 9$  oscillators.

Fig. 3 displays one of the 20 six-cycles in case of  $N = 9$  oscillators for the input configuration  $[1, 2, 3, 4, 5, 6, 7, 8, 9]$  (or rather for the input configuration group  $[a, b, c, d, e, f, X, Y, Z]$  where  $[a, b, c, d, e, f]$  represents the permutations of  $[1, 2, 3, 4, 5, 6]$  while the triplet  $[X, Y, Z]$  represents the permutations of  $[7, 8, 9]$ ). Notice that at each switch the underlined oscillators swap their phases  $y$  and  $b$  while the overlined oscillators cyclically permute their phases  $b, y$ , and  $w$ . The other 19 spatio-temporal codes appearing for the same input configuration group may be obtained by permuting the underlined oscillators and noncyclically permuting the overlined oscillators for each node.

*B. Representing Spatio-Temporal Codes: Weighted Order Parameter*

As the dynamics of (4) is high dimensional and somewhat difficult to visualize, we introduce a WOP that takes an arbitrarily weighted combination of phases from the oscillators. Using weights  $\rho_n > 0$  with  $\sum_{n=1}^N \rho_n = N$ , the WOP is defined as

$$R = \left| \frac{1}{N} \sum_{n=1}^N \rho_n e^{i\theta_n} \right|. \tag{10}$$

Note that  $0 \leq R \leq 1$  and  $R$  attains its maximum when the oscillators are in full synchrony, that is

$$R = 1 \iff \theta_n = \theta_m \tag{11}$$

for all  $n, m$ . However,  $R = 0$  does not necessarily correspond to evenly spaced phases.

The conventional choice for the weights is  $\rho_n = 1$  for all  $n$  but the resulting order parameter does not distinguish between the different symmetric copies of cluster states. For this paper, we set

$$\rho_n = N2^{-\sigma_n} \quad (12)$$

where the sequence  $[\sigma_1, \sigma_2, \dots, \sigma_{N-1}, \sigma_N]$  is a permutation of  $[1, 2, \dots, N-1, N-1]$ . The WOP is as a scalar output signal that averages the multiple signals of individual oscillators in a nontrivial way. The nonuniform weight distribution gives a different scalar output for each different cluster state. So a sufficiently long period of the WOP contains all necessary information about a spatio-temporal code.

In real neural ensembles, each neuron produces an extracellular electric field [26]. When a probe is inserted into a neural ensemble, the measured voltage represent a weighted average of extracellular potentials of individual neurons. The neurons that are closer to the probe in physical space contribute more to the signal than the ones that are further away. The low-pass filtered voltage signal is called the local field potential (LFP) [24], [25]. Using this analogy, we say that the WOP mimics an LFP for neural ensembles. However, the weight distribution (12) of the WOP is an assumption; it is not determined from the weight distribution of an LFP.

Fig. 4 displays the WOP (10) with weights (12) and exponents  $[\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5] = [4, 2, 3, 1, 4]$  for the first six six-cycles in Table II. The WOP is plotted on the left and the identified cluster states  $S$  are showed on the right. The input magnitude  $p = 10^{-4}$  and noise strength  $\eta = 5 \times 10^{-5}$  are considered. Note that the initial conditions are chosen close to one of the cluster states of the corresponding six-cycle to eliminate transient switches. Comparing panels (a) to (b), (c) to (d), and (e) to (f) they only differ in initial conditions while comparing other panels [e.g., (a) to (c)] the input configurations differ as well; see Table II. Notice that the time series of the WOP are different in each case, that is, this quantity reflects the variety of different spatio-temporal codes. In the next section, we describe a strategy for how such codes may be learned via synchronization and frequency adaptation.

#### IV. ADAPTIVE LEARNING

In this section, we propose a learning rule that allows one to transmit spatio-temporal codes between the teaching and learning systems. We consider the adaptive couplings in (1) of the form

$$\begin{aligned} u(\theta_n, \phi_n) &= u_0 \sin(\theta_n - \phi_n) \\ v(\theta_n, \phi_n) &= v_0 \sin(\theta_n - \phi_n) \end{aligned} \quad (13)$$

where  $u_0 > 0$  indicates the strength of the synchronization coupling and  $v_0 > 0$  indicates the strength of the adaptation coupling between the teaching and learning systems. As we will see below, as long as  $u_0$  is sufficiently large to ensure synchronization of the two systems, the adaptation process allows the

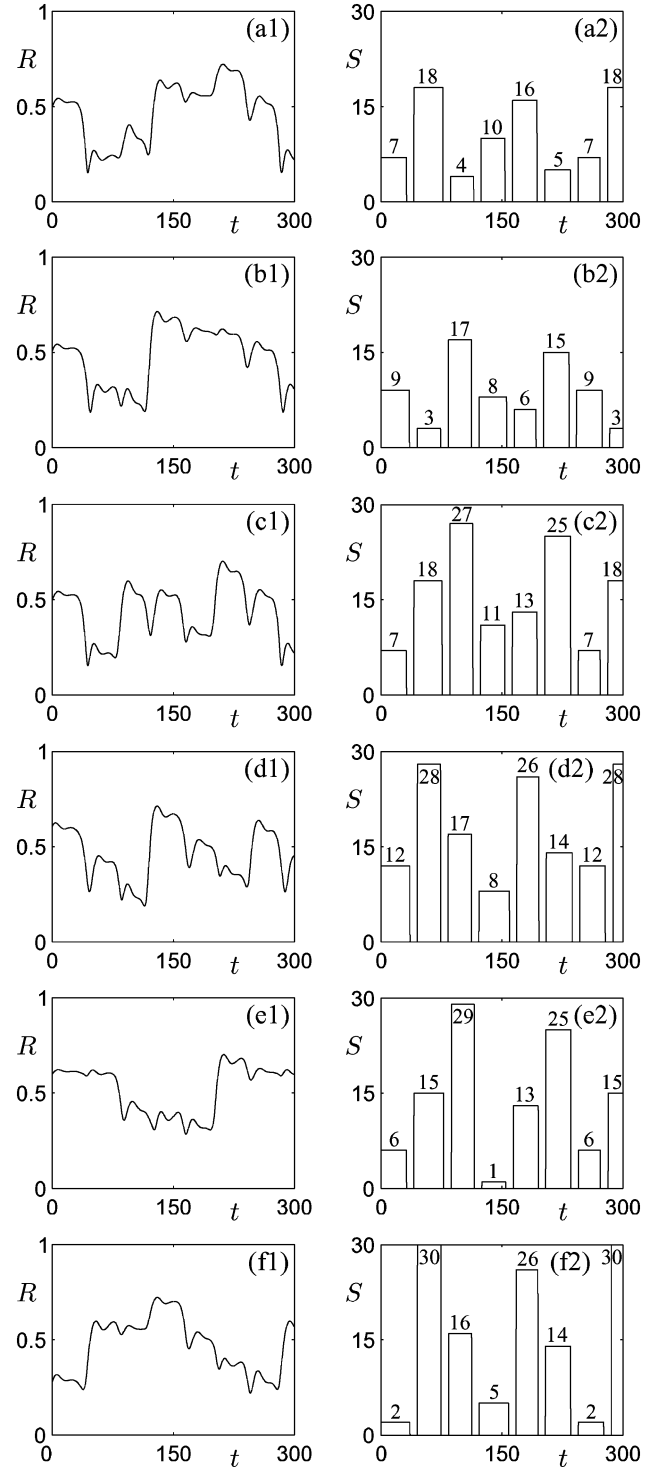


Fig. 4. Time series corresponding to the six-cycles (a)–(f) in Table II. On the left, the WOP  $R$  is shown as a function of time  $t$  where the weights are defined by (12) with exponents  $[\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5] = [4, 2, 3, 1, 4]$ . On the right, the identified cluster states  $S$  are displayed. The input magnitude  $p = 10^{-4}$  and noise strength  $\eta = 5 \times 10^{-5}$  are considered while the input configurations are in the first column of Table II.

learning system to learn the spatio-temporal code in a characteristic time scale of  $1/v_0$ .

By synchronization, we mean that the phases of teaching system  $\theta_n(t)$  and the phases of the learning system  $\phi_n(t)$  approach each other as time  $t$  increases. By adaptation, we mean



that the frequencies of the learning system  $\omega_n(t)$  approach the constant frequencies of the teaching system  $\Omega_n$  (6) as time  $t$  increases. To this end, we define the error variables

$$\varepsilon_n = \phi_n - \theta_n \quad \kappa_n = \omega_n - \Omega_n \quad (14)$$

and show that the state

$$\varepsilon_n(t) \equiv 0 \quad \kappa_n(t) \equiv 0 \quad (15)$$

is asymptotically stable for sufficiently strong synchronization coupling  $u_0$  (regardless the strength of the adaptation coupling  $v_0$ ). This means that

$$\lim_{t \rightarrow \infty} \varepsilon_n(t) = 0 \quad \lim_{t \rightarrow \infty} \kappa_n(t) = 0 \quad (16)$$

for an open set of initial conditions. That is, by definition (14), the phases synchronize and the frequencies adapt so that

$$\lim_{t \rightarrow \infty} (\phi_n(t) - \theta_n(t)) = 0 \quad \lim_{t \rightarrow \infty} \omega_n(t) = \Omega_n. \quad (17)$$

Note that according to (1) and (13) only the teaching phases  $\theta_n$  enter the learning system but not the teaching frequencies. So we assume that we know the teaching phases but have no information about the teaching frequencies; then using the learning system, we are able to extract the teaching frequencies as limit values of the learning frequencies.

#### A. Lyapunov Function Analysis

Here we construct a Lyapunov function to prove the asymptotic stability of the synchronized-adapted state, i.e., to show that synchronization and adaptation happen for any nonspecific initial condition (global convergence). For more details on Lyapunov functions, see [4]. For the following analysis, we set the noise to zero in (1), i.e.,  $\eta = 0$ . Now taking the difference of the first two sets of equations, subtracting  $\dot{\Omega}_n = 0$  from the third set of equations and using the learning rule (13) and definition (14), we obtain

$$\begin{aligned} \dot{\varepsilon}_n &= \kappa_n + \frac{1}{N} \sum_{m=1}^N (g(\theta_n - \theta_m + \varepsilon_n - \varepsilon_m) - g(\theta_n - \theta_m)) \\ &\quad - u_0 \sin \varepsilon_n \\ \dot{\kappa}_n &= -v_0 \sin \varepsilon_n. \end{aligned} \quad (18)$$

We define a Lyapunov function

$$V(\varepsilon_1, \dots, \varepsilon_N, \kappa_1, \dots, \kappa_N) = \sum_{n=1}^N \left( \frac{1}{2} f^2(\varepsilon_n) + \frac{1}{2} \kappa_n^2 \right) \quad (19)$$

where

$$\begin{aligned} f(\varepsilon_n) &= 2\sqrt{v_0} \sin \left( \frac{\varepsilon_n}{2} \right) \\ \Rightarrow \frac{1}{2} f^2(\varepsilon_n) &= v_0 (1 - \cos \varepsilon_n) \\ \Rightarrow \left( \frac{1}{2} f^2(\varepsilon_n) \right)' &= f(\varepsilon_n) f'(\varepsilon_n) = v_0 \sin \varepsilon_n. \end{aligned} \quad (20)$$

The derivative of (19) along trajectories of the system (18) gives

$$\begin{aligned} \frac{dV}{dt}(\varepsilon_1, \dots, \varepsilon_N, \kappa_1, \dots, \kappa_N) &= -v_0 \sum_{n=1}^N \left( u_0 \sin^2 \varepsilon_n - \sin \varepsilon_n \right. \\ &\quad \left. \times \frac{1}{N} \sum_{m=1}^N (g(\theta_n - \theta_m + \varepsilon_n - \varepsilon_m) - g(\theta_n - \theta_m)) \right). \end{aligned} \quad (21)$$

The function (19) is positive definite for  $\varepsilon_n \in (-\pi, \pi)$  and  $\kappa_n \in (-\infty, \infty)$  and we will show that the derivative (21) is negative definite when  $u_0$  is large enough. In fact,  $\dot{\kappa}_n$  is chosen in (18) so that  $\dot{\kappa}_n = -((1/2)f^2(\varepsilon_n))'$ , and consequently, the  $\kappa_n \sin \varepsilon_n$  terms are canceled in the derivative (21).

*Lemma 1:* If

$$u_0 > 2(1 + 2r)(N - 1)/N \quad \text{and} \quad v_0 > 0 \quad (22)$$

then

$$\frac{dV}{dt} \leq 0. \quad (23)$$

The proof of this lemma is presented in part B of the Appendix and it allows us to state the following theorem.

*Theorem 1:* If (22) holds, then solution (15) is asymptotically stable, that is, (16) and (17) hold for an open set of initial conditions.

*Proof:* Lemma 1 implies that the derivative of the Lyapunov function (21) is negative semidefinite

$$\frac{dV}{dt} = 0 \Rightarrow (\varepsilon_1, \dots, \varepsilon_N, \kappa_1, \dots, \kappa_N) = (0, \dots, 0, k_1, \dots, k_N) \quad (24)$$

where  $k_n \in \mathbb{R}$ . However,  $\varepsilon_n(t) \equiv 0$  implies  $\dot{\varepsilon}_n(t) \equiv 0$  and the dynamics of the system, i.e., the first row of (18), shows that in this case,  $\kappa_n(t) \equiv 0$ . Consequently, the largest invariant set inside  $(0, \dots, 0, k_1, \dots, k_N)$  is the origin  $(0, \dots, 0, 0, \dots, 0)$ . According to the Krasovskii–LaSalle invariance principle [4], one can conclude that the origin is asymptotically stable. ■

Theorem 1 states that if the synchronization coupling is strong enough ( $u_0$  is large enough) then the phases of the learning system and those of the teaching system synchronize to each other and the frequencies of the learning system adapt to those of the teaching system. Note that Theorem 1 provides us with a sufficient condition for  $u_0$ , that is, one may experience synchronization and adaptation for smaller values of  $u_0$ . Formula (21) shows that the synchronization is followed by adaptation for any value of  $v_0 > 0$ .

#### B. Numerical Simulations

We provide numerical evidence to illustrate the above analytical results. In Figs. 5–8, we illustrate the ability of the learning system to robustly and repeatably learn a wide range of spatio-temporal codes. In each figure, we plot the WOP  $R$

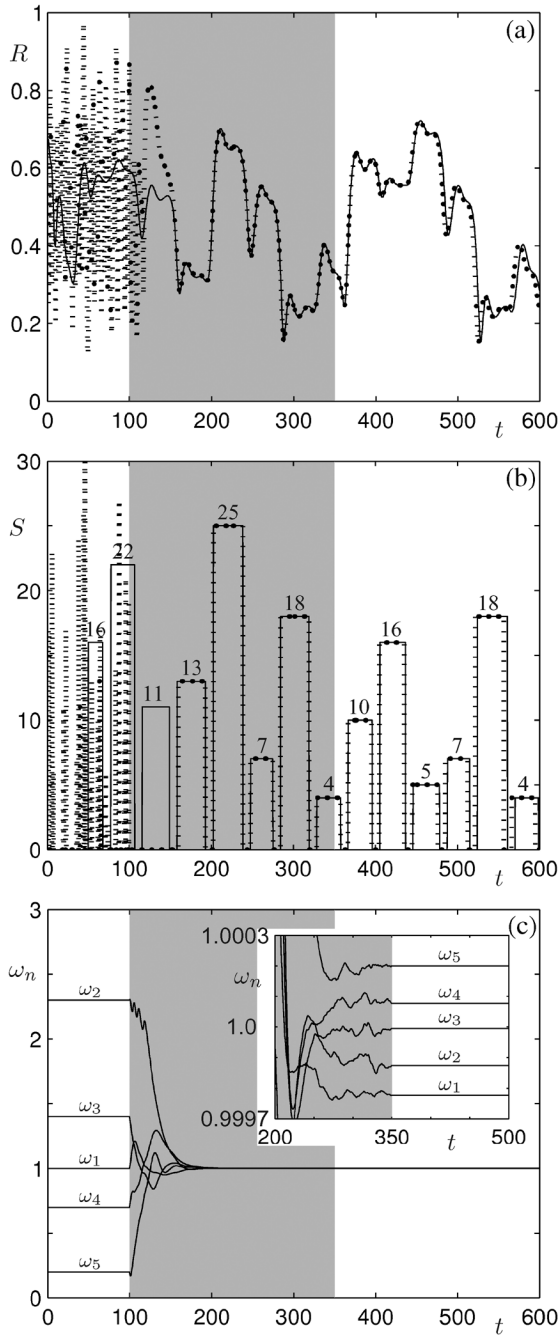


Fig. 5. Illustrations showing the process of adaptive learning of a spatio-temporal code for  $N = 5$  oscillators in each system. During the gray shaded region, the adaptive couplings are turned on:  $u_0 = 0.5$  and  $v_0 = 0.05$ ; during the remainder of the period, these are set to zero. (a) WOP  $R$  as a function of time  $t$  where the weights are defined by (12) with exponents  $[\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5] = [4, 2, 3, 1, 4]$ . (b) Identified cluster states  $S$ , for both the teaching system (solid line) and the learning system (dotted line). Observe also that the learning system continues to shadow the teaching system even after the adaptive couplings are turned off. (c) Time evolution of frequencies of the learning system  $\omega_n$  are displayed. Observe that the ordering of the frequencies is interchanged during the learning process. The input configuration  $[I_1, I_2, I_3, I_4, I_5] = [1, 2, 3, 4, 5]$ , the input magnitude  $p = 10^{-4}$ , and noise strength  $\eta = 5 \times 10^{-5}$  are considered.

given in (10) and the identified states  $S$  as a function of time  $t$ , for both the teaching system (solid line) and the learning system (dotted line). If the two systems are synchronized, then the solid and dotted lines stay close. We also display the frequencies of

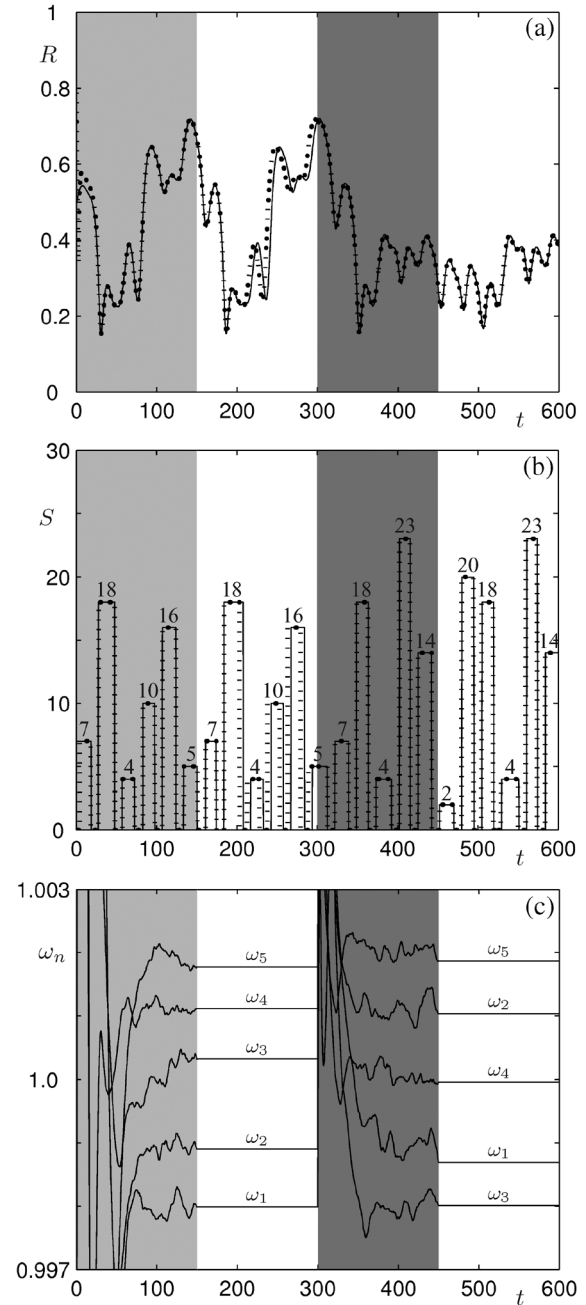


Fig. 6. Illustration as in Fig. 5 except there are two learning phases and the teaching systems is changed between the two phases. At time  $t = 300$ , the input configuration  $[I_1, I_2, I_3, I_4, I_5]$  is suddenly changed from  $[1, 2, 3, 4, 5]$  to  $[3, 1, 4, 2, 5]$ . Both spatio-temporal codes are rapidly learned. The input magnitude  $p = 10^{-3}$  and noise strength  $\eta = 5 \times 10^{-4}$  are considered.

the learning system  $\omega_n$  as a function of time  $t$ . The frequencies of the learning system are considered to be adapted when they stay close to the time-independent frequencies of the teaching system  $\Omega_n$  (6). During the gray shaded region, the adaptive couplings (13) are nonzero:  $u_0 = 0.5$  and  $v_0 = 0.05$ . During the white regions, these couplings are turned off:  $u_0 = 0$  and  $v_0 = 0$ .

Fig. 5 illustrates the adaptive learning of a spatio-temporal code for  $N = 5$  oscillators. General initial conditions are considered in the phases of the teaching system, in the phases of the learning system, and in the frequencies of the learning system



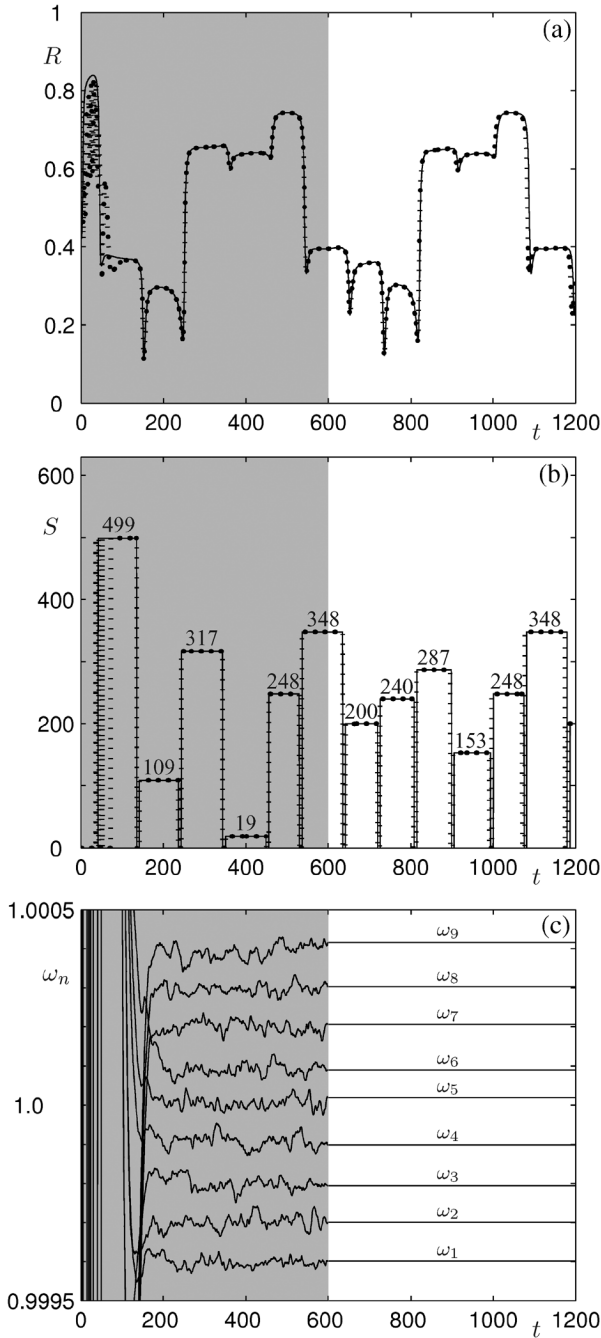


Fig. 7. Illustration as in Fig. 5 where each subsystem contains  $N = 9$  oscillators. The weights used in the order parameter  $R$  are defined by (12) with exponents  $[\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8, \sigma_9] = [6, 3, 1, 5, 7, 8, 8, 4, 2, \cdot]$ . The input configuration is  $[I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9] = [1, 2, 3, 4, 5, 6, 7, 8, 9]$ , the input magnitude is  $p = 10^{-4}$ , and the noise strength is  $\eta = 5 \times 10^{-5}$ .

as well. Regarding the frequencies of the teaching system  $\Omega_n$  (6), we consider the input configuration  $[I_1, I_2, I_3, I_4, I_5] = [1, 2, 3, 4, 5]$  and the input magnitude  $p = 10^{-4}$  while the noise strength is set to  $\eta = 5 \times 10^{-5}$ . Consequently, one of the two six-cycles in Fig. 2 (also shown in the first row of Table II) is expected to appear. In fact, after some transient steps, the teaching system approaches the six-cycle in Fig. 2(a).

In the time interval  $t \in (0, 100)$ , the adaptive couplings are turned off. Fig. 5(a) and (b) shows that the teaching system

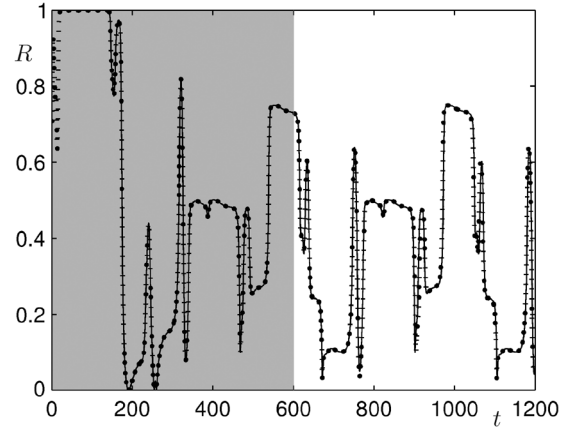


Fig. 8. Illustration as in Fig. 7(a) where each subsystem contains  $N = 51$  oscillators. The weights used in the order parameter  $R$  are defined by (12) with exponents  $\sigma_1 = N - 1$  and  $\sigma_n = N - n + 1, n = 2, \dots, N$ . The input configuration is given by  $I_n = n, n = 1, \dots, N$ , the input magnitude is  $p = 10^{-7}$ , and the noise strength is  $\eta = 5 \times 10^{-8}$ . The phase shifts  $\alpha = 1.6$  and  $\beta = -4.45$  are changed compared to (3) to obtain sequential switching.

(solid line) quickly approaches a transient switching path that terminates at the cyclic path discussed above. The learning system (dotted line), however, exhibits very different dynamics due to the initial frequencies being far from  $\Omega = 1$  as displayed in Fig. 5(c). In the time interval  $t \in (100, 350)$ , the teaching and learning systems are coupled and their phases rapidly synchronize to each other as shown in Fig. 5(a) and (b). During the synchronization process, the frequencies of the learning system rapidly approach a neighborhood of  $\Omega = 1$ . However, it takes a longer time until they adapt to the detuning pattern of the teaching system (i.e., to the input configuration) as shown by the inset in Fig. 5(c). Observe that the ordering of the frequencies is interchanged during the learning phase. Notice that the system adapts in spite of the relatively small signal-to-noise ratio  $p/\eta = 2$ . In the time interval  $t \in (350, 600)$ , the adaptive couplings are turned off again. Fig. 5(a) and (b) shows that the trajectory of the learning system shadows the trajectory of the teaching system even though the two system are uncoupled. The small deviations between the trajectories are due to small deviations between the learned frequencies and the prescribed equidistant teaching frequencies in Fig. 5(c). The characteristic time of synchronization and adaptation is shorter than the time period associated with the six-cycle.

In [30] and [32], it was demonstrated that by adapting the coupling constants in a simple network of oscillators, it is able to learn a periodic signal (that corresponds to a periodic orbit of the adapted system). One of the interesting features of system (1) is that many different periodic signals can be transmitted from the teaching system to the learning system for a wide range of periods. In fact, any of the six-cycles listed in Table II can be learned. By changing the input configuration, one can force both systems to “switch” from one six-cycle to another as demonstrated in Fig. 6.

In the time interval  $t \in (0, 150)$ , the adaptive couplings are turned on and the input configuration  $[I_1, I_2, I_3, I_4, I_5] = [1, 2, 3, 4, 5]$  is applied. The teaching system approaches the six-cycle shown on the left in the first row of Table II. The

initial phases of the teaching system are set close to state  $s_7$  to avoid transient switches but the phases and frequencies of the learning system are nonspecific. The phases of the two systems rapidly synchronize as shown by the solid and dotted lines in Fig. 6(a) and (b), and after synchronization, the learning frequencies adapt to the teaching ones as plotted in Fig. 6(c). In the time interval  $t \in (150, 300)$ , the adaptive couplings are turned off and learning trajectory shadows the teaching one. In the time interval  $t \in (300, 450)$ , the input configuration is changed to  $[I_1, I_2, I_3, I_4, I_5] = [3, 1, 4, 2, 5]$  and the teaching system “switches” to the six-cycle shown on the right in the sixth row of Table II. Again, a rapid synchronization of the phases is followed by adaptation of the frequencies. In the following time interval  $t \in (450, 600)$  (when the adaptive couplings are turned off again), shadowing happens. Note that the cluster states  $s_{18}$  and  $s_4$  appear in both spatio-temporal codes and that is why no transient switches happen when the system “switches” from one code to the other. In Fig. 6, the input magnitude  $p = 10^{-3}$  and the noise strength  $\eta = 5 \times 10^{-4}$  are used (such that the signal-to-noise ratio is still  $p/\eta = 2$ ). Observe that the larger input results in faster switching comparing to Fig. 5.

As it was described in Section III, the switching dynamics remain very similar when the number of oscillators is increased. However, the number of different input configurations and the number of different spatio-temporal codes increase; see Table III. Indeed, spatio-temporal codes can be learned for larger numbers of oscillators as demonstrated in Fig. 7 for  $N = 9$  where the adaptive couplings are turned on until  $t = 600$ . Here we consider the input configuration  $[I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9] = [1, 2, 3, 4, 5, 6, 7, 8, 9]$ , the input magnitude  $p = 10^{-4}$ , the noise strength  $\eta = 5 \times 10^{-5}$ , and general initial conditions. The six-cycle displayed in Fig. 3 is approached after a couple of transient switches. The dynamics are clearly comparable to the case of smaller oscillators although there are many more cluster states.

Finally, to move toward more realistic situations, we show that sequential switching exists and can be learned for much larger number of oscillators. Fig. 8 shows learning for  $N = 51$  oscillators for the input configuration  $I_n = n$ ,  $n = 1, \dots, N$ , input magnitude  $p = 10^{-7}$ , and noise strength  $\eta = 5 \times 10^{-8}$ . The parameters  $\alpha = 1.6$  and  $\beta = -4.45$  are tuned compared to (3) to find one of the eigenvalue configurations (8) or (9). The latter case, shown in Fig. 8, appears more robustly but the dynamics is structurally stable in both cases. Here the number of spatio-temporal codes is very large  $\approx 10^{18}$ ; see Table III. As  $N = 2k + 1$  is increased, the parameter domain associated with the sequential switching becomes smaller. In this paper, we only studied cluster states consisting of three clusters that contain  $k$ ,  $k$ , and 1 oscillators, respectively. One may find other (even larger) parameter domains with sequential switching between different cluster configurations.

These numerical results demonstrate that the constructed learning rules for frequency adaptation allow us to transmit information about the spatio-temporal codes from the teaching system to the learning system in a robust way. Note that one may alter the coupling strengths in the teaching or/and the learning systems to make the two systems nonidentical. Nu-

merical simulations show that learning is still possible if such alterations are smaller than the input magnitude.

## V. CONCLUSION AND DISCUSSION

We studied a coupled oscillator system that can exhibit winnerless competition/sequential switching between cluster states and where information about spatially nonhomogeneous time-independent inputs is encoded into spatio-temporal codes. These codes are cyclic sequences of switches between cluster states. We found that one can robustly and repeatably transmit information between such coupled oscillator systems via synchronization and adaptation of frequencies provided that the synchronization coupling is sufficiently strong and independent of the strength of adaptation coupling. Note that this high degree of “trainability” was achieved without modulating the couplings inside the learning system, i.e., via a non-Hebbian learning procedure.

In olfactory neurosystems (one or two synapses away from the receptors) odor information is encoded into spatio-temporal codes [22], [23]. In these systems (e.g., antennal lobes of insects or olfactory bulbs of mammals), both neural identity and timing are used for coding, allowing the system to differentiate between a great number of different chemical mixtures. Our work shows dynamics that can robustly give an immense variety of spatio-temporal codes and gives a learning rule to transmit information about the codes effectively and rapidly. This may lay the mathematical foundation of new approaches for reading out information from these neural ensembles by using artificial neural networks that are able to learn spatio-temporal codes. We remark that the learning rule also allows transmission of information about transient switching sequences that may also be significant in olfactory systems [27].

In the future, we wish to extend our research in two directions. On the one hand, neural systems are usually not all-to-all coupled and the coupling characteristics are usually not the same for each connection. We would like to find spatio-temporal codes in oscillator systems where some connections are missing and the existing couplings are not identical. On the other hand, the current learning algorithm would require voltage recording of each neuron of the biological system, which is not biologically feasible. The algorithm also requires an artificial network that is a very close replica of the biological network, which is difficult to engineer. We would like to be able to transmit information between two nonidentical systems using lower dimensional coupling (e.g., a scalar variable). It seems plausible to use a WOP for this purpose since all information about a spatio-temporal code is contained within a sufficiently long period of this quantity. The corresponding learning rule may be applicable to “read out” encoded states from biological systems through a local field potential.

## APPENDIX

### A. Determining Cluster States and Their Stability

Substituting the cluster state (7) (or any of its symmetrical copies) into (4) when  $\eta = 0$ , and defining the phase differences

$\tilde{y} = y - w$  and  $\tilde{b} = b - w$ , one may obtain

$$\begin{aligned}\tilde{\Omega} &= \Omega + \frac{1}{N} \left( kg(0) + g(\tilde{y}) + kg(\tilde{y} - \tilde{b}) \right) \\ \tilde{\Omega} &= \Omega + \frac{1}{N} \left( kg(0) + g(\tilde{b}) + kg(\tilde{b} - \tilde{y}) \right) \\ \tilde{\Omega} &= \Omega + \frac{1}{N} \left( g(0) + kg(-\tilde{y}) + kg(-\tilde{b}) \right)\end{aligned}\quad (25)$$

which determine  $\tilde{y}$ ,  $\tilde{b}$ , and  $\tilde{\Omega}$ .

Linearizing (4) about the cluster state (7) (or about any of its symmetrical copies) results in the eigenvalues

$$\begin{aligned}\lambda_1 &= 0 \\ \lambda_2 &= \frac{1}{N} \left( kg'(0) + g'(\tilde{y}) + kg'(\tilde{y} - \tilde{b}) \right) \\ \lambda_3 &= \frac{1}{N} \left( kg'(0) + g'(\tilde{b}) + kg'(\tilde{b} - \tilde{y}) \right) \\ \lambda_4 &= \frac{1}{2}(\mu + i\sqrt{\nu - \mu^2}) \\ \lambda_5 &= \frac{1}{2}(\mu - i\sqrt{\nu - \mu^2})\end{aligned}\quad (26)$$

where

$$\begin{aligned}\mu &= \frac{1}{N} \left( k \left[ g'(-\tilde{y}) + g'(\tilde{b}) + g'(-\tilde{b}) + g'(\tilde{y} - \tilde{b}) \right] \right. \\ &\quad \left. + g'(\tilde{y}) + g'(\tilde{b} - \tilde{y}) \right) \\ \nu &= \frac{4}{N^2} \left( k^2 \left[ g'(-\tilde{y}) + g'(-\tilde{b}) \right] \left[ g'(\tilde{y} - \tilde{b}) + g'(\tilde{b} - \tilde{y}) \right] \right. \\ &\quad \left. + k \left[ g'(\tilde{y})g'(-\tilde{b}) + g'(-\tilde{y})g'(\tilde{b}) + g'(\tilde{y})g'(\tilde{b} - \tilde{y}) \right. \right. \\ &\quad \left. \left. + g'(\tilde{b})g'(\tilde{y} - \tilde{b}) \right] + g'(\tilde{y})g'(\tilde{b}) \right).\end{aligned}\quad (27)$$

The eigenvalues  $\lambda_2$  and  $\lambda_3$  have multiplicity  $k - 1$ .

For the cluster state (7), the corresponding  $N = 2k + 1$ -dimensional eigenvectors are

$$v_1 = c_1 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\quad (28)$$

$$v_2 = c_{2,1} \begin{bmatrix} 1 \\ \chi \\ \vdots \\ \chi \\ 0 \\ \vdots \\ 0 \end{bmatrix} + c_{2,2} \begin{bmatrix} \chi \\ 1 \\ \vdots \\ \chi \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + c_{2,k} \begin{bmatrix} \chi \\ \vdots \\ \chi \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\quad (29)$$

$$v_3 = c_{3,1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \chi \\ \vdots \\ \chi \end{bmatrix} + c_{3,2} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \chi \\ 1 \\ \vdots \\ \chi \end{bmatrix} + \cdots + c_{3,k} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \chi \\ \vdots \\ \chi \\ 1 \end{bmatrix}\quad (30)$$

$$v_4 = c_4 \begin{bmatrix} p_y + i\sqrt{q_y} \\ \vdots \\ p_y + i\sqrt{q_y} \\ 1 \\ p_b + i\sqrt{q_b} \\ \vdots \\ p_b + i\sqrt{q_b} \end{bmatrix} \begin{matrix} \left. \vphantom{\begin{matrix} p_y + i\sqrt{q_y} \\ \vdots \\ p_y + i\sqrt{q_y} \\ 1 \\ p_b + i\sqrt{q_b} \\ \vdots \\ p_b + i\sqrt{q_b} \end{matrix}} \right\} k \\ \left. \vphantom{\begin{matrix} p_y + i\sqrt{q_y} \\ \vdots \\ p_y + i\sqrt{q_y} \\ 1 \\ p_b + i\sqrt{q_b} \\ \vdots \\ p_b + i\sqrt{q_b} \end{matrix}} \right\} k \end{matrix}$$

$$v_5 = c_5 \begin{bmatrix} p_y - i\sqrt{q_y} \\ \vdots \\ p_y - i\sqrt{q_y} \\ 1 \\ p_b - i\sqrt{q_b} \\ \vdots \\ p_b - i\sqrt{q_b} \end{bmatrix} \begin{matrix} \left. \vphantom{\begin{matrix} p_y - i\sqrt{q_y} \\ \vdots \\ p_y - i\sqrt{q_y} \\ 1 \\ p_b - i\sqrt{q_b} \\ \vdots \\ p_b - i\sqrt{q_b} \end{matrix}} \right\} k \\ \left. \vphantom{\begin{matrix} p_y - i\sqrt{q_y} \\ \vdots \\ p_y - i\sqrt{q_y} \\ 1 \\ p_b - i\sqrt{q_b} \\ \vdots \\ p_b - i\sqrt{q_b} \end{matrix}} \right\} k \end{matrix}\quad (31)$$

where  $\chi = -1/(k - 1)$ , the scalars  $c_1$ ,  $c_4$ , and  $c_5$  are arbitrary while one in the set of scalars  $\{c_{2,1}, \dots, c_{2,k}\}$  is zero and the others are arbitrary and the same holds for the set of scalars  $\{c_{3,1}, \dots, c_{3,k}\}$ . The complicated expressions of  $p_y, q_y, p_b, q_b \in \mathbb{R}$  contain  $g'(\cdot)$  at  $\tilde{y}, -\tilde{y}, \tilde{b}, -\tilde{b}, \tilde{y} - \tilde{b}$ , and  $\tilde{b} - \tilde{y}$ . For symmetric copies of (7), the eigenvectors can be determined by permuting components of (29)–(31) appropriately. Note that  $(\nu - \mu^2)$  in  $\lambda_4, \lambda_5$  and, consequently,  $q_y, q_b$  in  $v_4, v_5$  are not necessarily positive.

### B. Proof of Lemma 1

*Proof:* Since  $v_0 > 0$  is assumed, to prove that (23) holds, one needs to prove that

$$\begin{aligned}U(\varepsilon_1, \dots, \varepsilon_N, \kappa_1, \dots, \kappa_N) &:= \sum_{n=1}^N \left( u_0 \sin^2 \varepsilon_n - \sin \varepsilon_n \right. \\ &\quad \left. \times \frac{1}{N} \sum_{m=1}^N \left( g(\theta_n - \theta_m + \varepsilon_n - \varepsilon_m) - g(\theta_n - \theta_m) \right) \right) \\ &\geq 0.\end{aligned}\quad (32)$$

Using (2), one can obtain that

$$g(\theta_n - \theta_m + \varepsilon_n - \varepsilon_m) - g(\theta_n - \theta_m) = A \sin \left( \frac{\varepsilon_n - \varepsilon_m}{2} \right)\quad (33)$$

where

$$\begin{aligned}A(\alpha, \beta, r, \theta_n - \theta_m, \varepsilon_n - \varepsilon_m) &= -2 \cos \left( \alpha + \theta_n - \theta_m + \frac{\varepsilon_n - \varepsilon_m}{2} \right) \\ &\quad + 4r \cos(\beta + 2(\theta_n - \theta_m) + \varepsilon_n - \varepsilon_m) \cos \left( \frac{\varepsilon_n - \varepsilon_m}{2} \right)\end{aligned}\quad (34)$$

and notice that

$$-M \leq A \leq M, \quad \text{where } M = 2(1 + 2r).\quad (35)$$

Let us define the new variables

$$p_n = \sin \frac{\varepsilon_n}{2} \quad q_n = \cos \frac{\varepsilon_n}{2}. \quad (36)$$

Note that  $p_n^2 + q_n^2 = 1$ , that is,  $p_n$  and  $q_n$  are not independent variables. Furthermore,  $p_n \in [0, 1) \Leftrightarrow \varepsilon_n \in [0, \pi)$  and  $p_n \in (-1, 0] \Leftrightarrow \varepsilon_n \in (-\pi, 0]$ , while  $q_n \in [0, 1) \Leftrightarrow \varepsilon_n \in (-\pi, \pi)$ . Substituting (33) and (34) into (32) and using (36), one can obtain

$$\begin{aligned} U &= \sum_{n=1}^N \left( 4u_0 p_n^2 q_n^2 - \frac{2}{N} \sum_{m=1}^N A (p_n^2 q_n q_m - p_n p_m q_n^2) \right) \\ &\geq \sum_{n=1}^N \left( 4u_0 p_n^2 q_n^2 - \frac{2M}{N} p_n^2 \sum_{\substack{m=1 \\ m \neq n}}^N q_n q_m - \frac{2M}{N} q_n^2 \sum_{\substack{m=1 \\ m \neq n}}^N |p_n p_m| \right) \end{aligned} \quad (37)$$

where we used (35) to derive the last inequality. By using the inequalities  $p_n p_m \leq (p_n^2 + p_m^2)/2$  and  $q_n q_m \leq (q_n^2 + q_m^2)/2$ , one may prove that the last line of (37) is nonnegative if  $u_0 > M(N-1)/N = 2(1+4r)(N-1)/N$ . ■

#### ACKNOWLEDGMENT

The authors would like to thank R. Borisyyuk, S. Coombes, and J. Wordsworth for fruitful conversations and the reviewers for detailed and useful comments.

#### REFERENCES

- [1] V. S. Afraimovich, M. I. Rabinovich, and P. Varona, "Heteroclinic contours in neural ensembles and the winnerless competition principle," *Int. J. Bifurcat. Chaos*, vol. 14, no. 4, pp. 1195–1208, 2004.
- [2] P. Ashwin, G. Orosz, J. Wordsworth, and S. Townley, "Dynamics on networks of cluster states for globally coupled phase oscillators," *SIAM J. Appl. Dyn. Syst.*, vol. 6, no. 4, pp. 728–758, 2007.
- [3] P. Ashwin and M. Timme, "When instability makes sense," *Nature*, vol. 436, no. 7047, pp. 36–37, 2005.
- [4] K. J. Åström and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton, NJ: Princeton Univ. Press, 2008.
- [5] J. R. Dias, R. F. Oliveira, and O. Kinouchi, "Chaotic itinerancy, temporal segmentation and spatio-temporal combinatorial codes," *Physica D*, vol. 237, no. 1, pp. 1–5, 2007.
- [6] A. A. Fingelkurts and A. A. Fingelkurts, "Making complexity simpler: Multivariability and metastability in the brain," *Int. J. Neurosci.*, vol. 114, no. 7, pp. 843–862, 2004.
- [7] W. J. Freeman, *Neurodynamics: An Exploration in Mesoscopic Brain Dynamics*. New York: Springer-Verlag, 2000.
- [8] D. Hansel, G. Mato, and C. Meunier, "Phase dynamics for weakly coupled Hodgkin-Huxley neurons," *Europhys. Lett.*, vol. 23, no. 5, pp. 367–372, 1993.
- [9] B. Hasselblatt and A. Katok, "The development of dynamics in the 20th century and the contribution of Jürgen Moser," *Ergodic Theory Dyn. Syst.*, vol. 22, pp. 1343–1364, 2002.
- [10] E. M. Izhikevich, "Polychronization: Computation with spikes," *Neural Comput.*, vol. 18, no. 2, pp. 245–282, 2006.
- [11] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Cambridge, MA: MIT Press, 2007.
- [12] W. Just, S. Ahn, and D. Terman, "Minimal attractors in digraph system models of neuronal networks," *Physica D*, vol. 237, no. 24, pp. 3186–3196, 2008.
- [13] K. Kaneko and I. Tsuda, "Chaotic itinerancy," *Chaos*, vol. 13, no. 3, pp. 926–936, 2003.
- [14] J. A. S. Kelso, *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press, 1995.

- [15] C. Kirst and M. Timme, "From networks of unstable attractors to heteroclinic switching," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 78, no. 6, 2008, 065201(R).
- [16] I. Z. Kiss, C. G. Rusin, H. Kori, and J. L. Hudson, "Engineering complex dynamical structures: Sequential patterns and desynchronization," *Science*, vol. 316, no. 5833, pp. 1886–1889, 2007.
- [17] H. Kori, C. G. Rusin, I. Z. Kiss, and J. L. Hudson, "Synchronization engineering: Theoretical framework and application to dynamical clustering," *Chaos*, vol. 18, no. 2, p. 026111, 2008.
- [18] R. Kozma and W. J. Freeman, "Encoding and recall of noisy data as chaotic spatio-temporal memory patterns in the style of the brains," in *Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw.*, 2000, vol. 5, pp. 33–38.
- [19] R. Kozma, W. J. Freeman, and P. Erdi, "The KIV model—Nonlinear spatio-temporal dynamics of the primordial vertebrate forebrain," *Neurocomputing*, vol. 52–54, pp. 819–826, 2003.
- [20] R. Kozma, M. Puljic, P. Balister, B. Bollobás, and W. J. Freeman, "Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions," *Biol. Cybern.*, vol. 92, no. 6, pp. 367–379, 2005.
- [21] J. Kurths and A. Pikovsky, "Symmetry breaking in distributed systems and modulational spatio-temporal intermittency," *Chaos Solitons Fractals*, vol. 5, no. 10, pp. 1893–1899, 1995.
- [22] G. Laurent, "A systems perspective on early olfactory coding," *Science*, vol. 284, no. 5440, pp. 723–728, 1999.
- [23] G. Laurent, "Olfactory network dynamics and the coding of multidimensional signals," *Nature Rev. Neurosci.*, vol. 3, no. 11, pp. 884–895, 2002.
- [24] G. Laurent and H. Davidowitz, "Encoding of olfactory information with oscillating neural assemblies," *Science*, vol. 265, no. 5180, pp. 1872–1875, 1994.
- [25] A. D. Legatt, J. Arezzo, and H. G. Vaughan, Jr., "Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: Effects of volume-conducted potentials," *J. Neurosci. Methods*, vol. 2, no. 2, pp. 203–217, 1980.
- [26] J. N. Milstein and C. Koch, "Dynamic moment analysis of the extracellular electric field of a biologically realistic spiking neuron," *Neural Comput.*, vol. 20, no. 1, pp. 1–15, 2008.
- [27] M. Rabinovich, R. Huerta, and G. Laurent, "Transient dynamics for neural processing," *Science*, vol. 321, no. 5885, pp. 48–50, 2008.
- [28] M. I. Rabinovich, R. Huerta, P. Varona, and V. S. Afraimovich, "Generation and reshaping of sequences in neural systems," *Biol. Cybern.*, vol. 95, no. 6, pp. 519–536, 2006.
- [29] M. I. Rabinovich, P. Varona, A. I. Selverston, and H. D. I. Abarbanel, "Dynamical principles in neuroscience," *Rev. Modern Phys.*, vol. 78, no. 4, pp. 1213–1265, 2006.
- [30] A. Ruiz, D. H. Owens, and S. Townley, "Existence, learning, and replication of periodic motions in recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 9, no. 4, pp. 651–661, Jul. 1998.
- [31] D. Terman, S. Ahn, X. Wang, and W. Just, "Reducing neuronal networks to discrete dynamics," *Physica D*, vol. 237, no. 3, pp. 324–338, 2008.
- [32] S. Townley, A. Ilchmann, M. G. Weiß, W. McClements, A. C. Ruiz, D. H. Owens, and D. Prätzl-Wolters, "Existence and learning of oscillations in recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 11, no. 1, pp. 205–214, Jan. 2000.
- [33] I. Tsuda, "Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems," *Behav. Brain Sci.*, vol. 24, no. 5, pp. 793–847, 2001.
- [34] A. T. Winfree, *The Geometry of Biological Time*, 2nd ed. New York: Springer-Verlag, 2001.
- [35] J. Wordsworth and P. Ashwin, "Spatiotemporal coding of inputs for a system of globally coupled phase oscillators," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 78, no. 6, p. 066203, 2008.



**Gábor Orosz** received the M.Sc. degree in applied physics from the Budapest University of Technology and Economics, Budapest, Hungary, in 2002 and the Ph.D. degree in applied mathematics from the University of Bristol, Bristol, U.K., in 2006.

He held a postdoctoral position at the University of Exeter, Exeter, U.K., before joining the University of California, Santa Barbara. His research interests are in the nonlinear dynamics of complex networks and related control algorithms with an eye to the roles of information delays. His current projects include

neural networks, transcriptional regulatory networks, highway traffic, and territoriality problems.



**Peter Ashwin** received the BA and CAS degrees in mathematics from the University of Cambridge, Cambridge, U.K. and the Ph.D. degree in mathematics from the University of Warwick, Warwick, U.K.

After having held postdoctoral and lecturer positions at the Universities of Warwick, U.K., Marburg, Germany, and Surrey, U.K., and the Institut Non-Linéaire de Nice, France, he became a Professor of Mathematics at the University of Exeter, Exeter, U.K. His research interests range from theoretical to applied topics in nonlinear dynamics and bifurcation theory, especially with symmetry. He is particularly interested in the dynamics of coupled networks and spatio-temporal pattern formation. His work on applications includes collaborations looking at coupled neuron models, fluid dynamical instabilities, and microstructure formation.



**Stuart Townley** received the B.Sc. degree in mathematics and the Ph.D. degree in engineering from the University of Warwick, Warwick, U.K., in 1983 and 1987, respectively.

He followed then with postdoctoral positions at the University of Warwick and the University of Bath, U.K. He joined the University of Exeter, Exeter, U.K. in 1990 as a Lecturer and he became a Professor of Applied Mathematics in 1999. He has research interests spanning from infinite-dimensional control theory, through robust and adaptive control, to applications of dynamical systems and control in neurodynamics, population biology, and renewable energy. He is the author of over 100 journal and refereed conference papers.