Learning of Spatiotemporal Codes in a Coupled Oscillator System

Gábor Orosz, Peter Ashwin, and Stuart Townley

Abstract—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 modelled with reference to a number of partially synchronized cluster states and transitions between them. Forcing the teaching system by steady but spatially nonhomogeneous inputs produce cyclic sequences of transitions between the cluster states, that is, information about inputs is encoded via a ‘winnerless competition’ process into spatiotemporal codes. The large variety of codes can be learnt by the learning system that adapts its frequencies to those of the teaching system. We visualize the dynamics using ‘weighted order parameters’ that are analogous to ‘local field potentials’ in neural systems. Since spatiotemporal coding is a mechanism that appears in olfactory systems, the developed learning rules may help to understand information processing in these neural ensembles.

Index Terms—Coupled oscillator system, winnerless competition, heteroclinic network, spatiotemporal code, adaptive learning.

I. INTRODUCTION AND MOTIVATION

WORK on a variety of neural systems suggest that an important mechanism at play is the use of spatiotemporal codes to represent information [1], [2]. In particular, there is biological evidence that this may be an important part of the encoding of spatially non-homogenous inputs in olfactory systems such as the insect antennal lobe and mammalian olfactory bulb [3]. A plausible nonlinear dynamical mechanism that robustly produces a spatiotemporal code is a process of winnerless competition on a cycle of unstable states [4], [5]. The present paper takes these ideas further; we study how a variety of different inputs can be converted into different spatiotemporal codes. Furthermore, we derive a learning rule that allows to transmit information between two systems (that are capable of performing ‘winnerless competition’) via synchronization and frequency adaptation. Note that by spatiotemporal code we rather mean identity-temporal code since information depends on which neurons are acting rather that where they lie in physical space.

The dynamics of the neural network we consider is well-understood as 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 [6].

Consequently, the dynamics consists of a sequence of quasi-steady states near partially synchronized cluster states interspersed with fast transitions - called switches - between different cluster states. A spatiotemporal code is then a cyclic sequence of switches between cluster states, i.e., a cyclic path along the graph of the heteroclinic network. Since the 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 spatiotemporal codes. Depending on the magnitude of inputs there exist a characteristic switching time and the obtained coding scenario is very robust against noise. 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 spatiotemporal 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 idealised ‘teaching system’ that has its frequencies detuned by the input and produces the corresponding spatiotemporal 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 spatiotemporal code, i.e., about the input provided to the teaching system.

The paper is organized as follows: in Section II we introduce the coupled oscillator system consisting of two copies of the system discussed in [7] 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 spatiotemporal code. Learning is illustrated using the weighted order parameter. Finally, in Section V we outline some of the consequences and outstanding problems of this approach.

G. Orosz, P. Ashwin, and S. Townley are at the Mathematics Research Institute, School of Engineering, Computing and Mathematics, University of Exeter, Exeter EX4 4QF, United Kingdom. Email: g.orosz@exeter.ac.uk
II. MODELLING

We consider two coupled phase oscillator systems of the type studied in [2], each containing \( N \) oscillators: a teaching system with phases \( \theta = \text{col}[\theta_1, \ldots, \theta_N] \) and a learning system with phases \( \phi = \text{col}[\phi_1, \ldots, \phi_N] \) that are coupled to each other. For each \( n = 1, \ldots, N \) we have

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

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

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

(1)

where the dot denotes differentiation with respect to the 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 \( u, v \) 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 learning 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 these couplings in Sections III, IV.

III. SWITCHING DYNAMICS BETWEEN CLUSTER STATES

In this section we summarize the dynamics of the teaching system (that is not influenced by the learning system), generalizing the discussion in [2] to larger numbers of oscillators. More details about the simplest nontrivial case of \( N = 5 \) oscillators can be found in [2]. We consider the first \( N \) equations of (1)

\[
\dot{\theta}_n = \Omega_n + \frac{1}{N} \sum_{m=1}^{N} g(\theta_n - \theta_m) + \eta \xi_n, \quad \text{(2)}
\]

when the natural frequencies \( \Omega_n \) are close to the average frequency

\[
\Omega = \frac{1}{N} \sum_{n=1}^{N} \Omega_n. \quad \text{(3)}
\]

In particular, we consider the frequency distribution

\[
\Omega_n = \Omega + p \left( I_n - \frac{N+1}{2} \right), \quad \text{(4)}
\]

where \( p \ll \Omega \) and the set \( [I_1, \ldots, I_N] \) is a permutation of \( [1, \ldots, N] \). The second term in (4) can be interpreted either as a detuning to the average frequency \( \Omega \) or as a constant external stimulus. For any input configuration \( [I_1, \ldots, I_N] \) each oscillator receives a different constant stimulus of the order of the input magnitude \( p \). In this paper we will use the abbreviation: oscillator \( n \) receives stimulus \( I_n \) meaning that oscillator \( n \) receives stimulus \( p(I_n - \frac{N+1}{2}) \). Note that there exist \( N! \) different input configurations. These configurations are special in the sense that the stimuli \( I_n \) are equidistant whilst the qualitative dynamics do not change for near-equidistant stimuli as will be demonstrated in Section IV-B. Note that the right-hand side of (2) only depends on the phase differences, so it is sufficient to examine these to determine the behavior of the system.

In this paper we consider the coupling function

\[
g(\varphi) = -\sin(\varphi + \alpha) + r \sin(2\varphi + \beta), \quad \text{(5)}
\]

and use the parameters

\[
\Omega = 1.0, \quad r = 0.2, \quad \alpha = 1.8, \quad \beta = -2.0. \quad \text{(6)}
\]

Note that the dynamics explained below are robust to sufficiently small changes in these parameters, i.e., structurally stable. To obtain such dynamics for \( N \geq 11 \), parameters have to be slightly altered compared to (6) but the dynamics still exist in an open set of parameters, that is, still structurally stable.

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

\[
[I_1(t), \ldots, I_N(t)] = \Omega t \left( [1, \ldots, 1] + \left[ \underbrace{y, \ldots, y, w, b, \ldots, b}_k \right] \right), \quad \text{(7)}
\]

with constant frequency \( \Omega \) and phases \( y, w \) and \( b \) (which stand for yellow, white and blue following the coloring of oscillators in [2]). In this paper we will use the abbreviated names \( b \) or \( w \) cluster/oscillators, \( y \) cluster/oscillators and \( w \) oscillator. One may substitute (7) into (2) and obtain (23) in Appendix A that provides the frequency \( \Omega \) and the phase differences \( b - w \) and \( y - w \) (which are unique for parameters (6)).
Furthermore, there are additional cluster states obtained by permuting the components of \( \{ y, y, w, b, y \} \) so that there are \( N!/(k!k!k!) \) cluster states. Table I lists all (thirty) states for \( N = 5 \) oscillators so that the term \( \Omega t \{ 1, \ldots, 1 \} \) is not spelled out (and this convention is adopted in the remainder of this paper).

One may linearize (2) about (7) (or about any of its symmetrical copies) and calculate the eigenvalues (24) and the eigenvectors (25-28) as shown in Appendix A. For parameters (6), there are eigenvalues on both sides of the imaginary axis, that is, the above cluster states are saddle type. The eigenvalues and eigenvectors show that the \( y \) cluster is ‘stable’, that is, perturbations desynchronizing the \( y \) cluster decay in time. While 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 \( k \) different vectors appearing in (26) show that there are \( k \) different perturbations depending on which \( y \) oscillator has its phase advanced with respect to the other \( y \) oscillators. The \( k \) vectors appearing in (27) correspond to the \( k \) different perturbations that can be applied to the \( b \) oscillators.

Following the unstable directions in (27) 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 (winnerless competition), 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 (26)) and \( k \) outgoing edges (corresponding to the \( k \) different vectors in (27)). 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 \( b \) oscillator changes to \( y \) and the phases of the \( w \) and \( y \) oscillators change to \( y \) and \( b \), respectively.

<table>
<thead>
<tr>
<th>( y, y, w, b, y )</th>
<th>( s_1 )</th>
<th>( y, w, b, y, w )</th>
<th>( s_16 )</th>
<th>( w, b, y, w, b )</th>
<th>( s_20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b, b, y, y, w )</td>
<td>( s_2 )</td>
<td>( y, w, b, y, y )</td>
<td>( s_21 )</td>
<td>( b, y, w, y, b )</td>
<td>( s_22 )</td>
</tr>
<tr>
<td>( b, y, b, y, w )</td>
<td>( s_3 )</td>
<td>( w, b, y, b, y )</td>
<td>( s_23 )</td>
<td>( b, y, b, w, y )</td>
<td>( s_24 )</td>
</tr>
<tr>
<td>( y, b, b, y, w )</td>
<td>( s_4 )</td>
<td>( y, w, b, b, y )</td>
<td>( s_25 )</td>
<td>( b, y, b, y, w )</td>
<td>( s_26 )</td>
</tr>
<tr>
<td>( y, b, y, b, w )</td>
<td>( s_5 )</td>
<td>( w, y, b, b, y )</td>
<td>( s_27 )</td>
<td>( b, y, b, w, y )</td>
<td>( s_28 )</td>
</tr>
<tr>
<td>( b, y, b, y, w )</td>
<td>( s_6 )</td>
<td>( y, w, b, y, b )</td>
<td>( s_29 )</td>
<td>( b, y, w, b, y )</td>
<td>( s_30 )</td>
</tr>
</tbody>
</table>

Table I

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

<table>
<thead>
<tr>
<th>( y, y, w, b, y )</th>
<th>( s_8 )</th>
<th>( y, w, b, y, y )</th>
<th>( s_28 )</th>
<th>( w, b, y, b, y )</th>
<th>( s_32 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b, b, y, y, w )</td>
<td>( s_9 )</td>
<td>( b, w, y, b, y )</td>
<td>( s_33 )</td>
<td>( y, b, w, b, y )</td>
<td>( s_34 )</td>
</tr>
<tr>
<td>( b, y, b, y, w )</td>
<td>( s_{10} )</td>
<td>( b, y, b, w, y )</td>
<td>( s_35 )</td>
<td>( y, b, w, b, y )</td>
<td>( s_36 )</td>
</tr>
<tr>
<td>( y, b, b, y, w )</td>
<td>( s_{11} )</td>
<td>( y, b, w, b, y )</td>
<td>( s_37 )</td>
<td>( y, b, y, b, w )</td>
<td>( s_38 )</td>
</tr>
<tr>
<td>( y, b, y, b, w )</td>
<td>( s_{12} )</td>
<td>( y, w, b, y, b )</td>
<td>( s_39 )</td>
<td>( y, b, y, w, b )</td>
<td>( s_40 )</td>
</tr>
</tbody>
</table>

Fig. 2. Two 6-cycles (spatiotemporal 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 largest stimuli changes to \( y \), 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 \( w \).

<table>
<thead>
<tr>
<th>( y, b, y, b, y, b, b, y )</th>
<th>( s_{248} )</th>
<th>( y, b, b, y, b, y, w )</th>
<th>( s_48 )</th>
<th>( y, b, y, b, y, w, b )</th>
<th>( s_40 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y, b, b, y, y, b, b, y )</td>
<td>( s_200 )</td>
<td>( y, b, y, b, y, b, w )</td>
<td>( s_20 )</td>
<td>( y, b, y, y, b, w, b )</td>
<td>( s_28 )</td>
</tr>
<tr>
<td>( y, b, b, y, b, y, b )</td>
<td>( s_{15} )</td>
<td>( y, b, y, b, y, b )</td>
<td>( s_8 )</td>
<td>( y, b, b, y, b, w )</td>
<td>( s_6 )</td>
</tr>
</tbody>
</table>

Fig. 3. One of the twenty 6-cycles (spatiotemporal 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 stimuli changes to \( z \), the phase of the other \( b \) oscillators change to \( w \) and the phases of the \( w \) and \( y \) oscillators change to \( y \) and \( b \), respectively. Thus, the underlined oscillators (receiving stimuli \( \{ 1, 2, 3, 4, 5 \} \)) swap their phases \( y \) and \( b \), while the overlined oscillators (receiving stimuli \( \{ 7, 8, 9 \} \)) cyclically permute their phases \( b, y \) and \( w \).

A. 6-cycles as spatiotemporal codes

For \( p = 0, \eta = 0 \), the time interval between switches increases exponentially with exponent \( \ln(\lambda_2/|\lambda_3|) \) (oscillations slow down). For \( p = 0, \eta > 0 \), 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 time \( T_s \approx -(\ln \eta)/\lambda_3 \). For \( p > 0, \eta = 0 \), the input configuration \( \{ I_1, \ldots, I_N \} \) 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 6 switches (6-cycles) that are approached after a finite number of transient switches. These cyclic paths correspond to stable limit cycles in state space, which 6-cycle is approached, and via which transient path, depends on the initial condition (the transients are not discussed in detail in this article). The characteristic switching time is now \( T_s \approx -(\ln p)/\lambda_3 \). Finally for \( p > \eta > 0 \), the input dominates the dynamics, i.e., 6-cycles persist, as will be demonstrated in Section IV-B.

We remark that along the cyclic paths each switch has a slightly different switching time. One may incorporate these time differences into the spatiotemporal code to give a one-to-
one correspondence between the inputs and codes [9]. In this paper we consider relatively small signal-to-noise ratio in the input signal so these small differences in switching time are not easily decodeable. On the other hand the coding scenario explained here is very robust against noise.

The information about the input configuration \([I_1, \ldots, I_N]\) is encoded into the appearing 6-cycles which are the spatiotemporal codes. First we discuss the simplest nontrivial case of \(N = 5\) oscillators. Fig. 2 shows the two 6-cycles (spatiotemporal 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 6. (There are two different 6-cycles since there are two different period three oscillations with cyclically permuted phases.)

In fact, the same spatiotemporal 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 \(2! \cdot 3! = 12\) 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 spatiotemporal 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 spatiotemporal codes as shown in Table II.

In this way all (twenty) possible 6-cycles appear as spatiotemporal codes. Note that the total number of 6-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 node \(s_i\) to \(s_j\) and \(A_{ij} = 0\) otherwise. The number of cyclic paths of length \(l\) are 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 2-, 3-, and 4-cycles) the 6-cycles are non-repetitive.

The above coding strategy persists for larger number of oscillators. In general there are \(N!\) input configurations \([I_1, \ldots, I_N]\) which are classified into \(^N\mathcal{S}_3\) 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 \(^{(N - 3)!}\mathcal{S}_3\) input configurations. Furthermore, for a particular input configuration group one may find \(^{(N - 3)!}\mathcal{S}_3\) 6-cycles (spatiotemporal codes) since there are \(\frac{1}{2}^\frac{(N - 3)!}{2!}\) 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 6-cycles \(\text{tr}(A^6)/6\) can be calculated by multiplying the number of input configuration groups with number of 6-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 twenty 6-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 any of the six permutations of \([7, 8, 9]\)). Notice

\[
\begin{array}{|c|c|}
\hline
\text{input configuration group} & \text{spatiotemporal codes} \\
\hline
[a, b, X, Y, Z] & \begin{array}{c}
\begin{aligned}
(a) & \quad 818 & 827 & 829 & 830 & 831 & 817 \\
(b) & \quad 87 & 810 & 809 & 808 & 806 & 816
\end{aligned}
\end{array} \\
\hline
[a, X, b, Y, Z] & \begin{array}{c}
\begin{aligned}
(c) & \quad 818 & 827 & 829 & 830 & 831 & 817 \\
(d) & \quad 87 & 810 & 809 & 808 & 806 & 816
\end{aligned}
\end{array} \\
\hline
[a, X, Y, Z, b] & \begin{array}{c}
\begin{aligned}
(e) & \quad 815 & 820 & 830 & 829 & 828 & 827 \\
(f) & \quad 85 & 813 & 811 & 812 & 813 & 826
\end{aligned}
\end{array} \\
\hline
[a, X, b, Y, Z] & \begin{array}{c}
\begin{aligned}
(X, a, b, Y, Z) & \quad 810 & 819 & 820 & 821 & 822 & 818 \\
(Y, a, b, Z) & \quad 817 & 819 & 820 & 821 & 822 & 818 \\
(Z, a, b, X, Y) & \quad 823 & 824 & 825 & 826 & 827 & 828 \\
(X, Y, a, b, Z) & \quad 823 & 824 & 825 & 826 & 827 & 828 \\
(Y, Z, a, b, X) & \quad 823 & 824 & 825 & 826 & 827 & 828 \\
(Z, Y, X, a, b) & \quad 823 & 824 & 825 & 826 & 827 & 828
\end{aligned}
\end{array} \\
\hline
\end{array}
\]
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 nineteen spatiotemporal codes appearing for the same input configuration group may be obtained by permuting the underlined oscillators and non-cyclically permuting the overlined oscillators for each node.

**B. Representing spatiotemporal codes - weighted order parameter**

As the dynamics of (2) is high dimensional and somewhat difficult to visualize, we introduce a weighted order parameter (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

\[
R = \left| \frac{1}{N} \sum_{n=1}^{N} \rho_n e^{i \theta_n} \right|. \tag{8}
\]

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

\[
R = 1 \quad \Leftrightarrow \quad \theta_n = \theta_m, \tag{9}
\]

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

\[
\rho_n = N 2^{-\sigma_n}, \tag{10}
\]

where the sequence \([\sigma_1, \sigma_2, \ldots, \sigma_{N-1}, \sigma_N]\) is a permutation of \([1, 2, \ldots, N-1, N-1]\). Such a weight distribution mimics the case when a probe is inserted into a neural ensemble to measure a local field potential (LFP). The neurons that are close to the probe in physical space contribute more to the LFP than the ones that are further away; for biophysical background see [2].

Fig. 4 displays the WOP with exponents \([\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5] = [4, 2, 3, 1, 4]\) for the first six 6-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 while the input configurations are in the first column of Table II.

Note that the initial conditions are chosen close to one of the cluster states of the corresponding 6-cycle to eliminate transient switches. Comparing panels (a) to (b), (c) to (d) and (e) to (f) they only differ in initial conditions.

---

**TABLE III**

THE NUMBER OF SPATIALLY NONHOMOGENEOUS TIME INDEPENDENT INPUTS AND SPATIOTEMPORAL CODES FOR \( N = 2k + 1 \) OSCILLATORS. NOTE THAT THE LAST ROW CAN ALSO BE CALCULATED AS THE PRODUCT OF THE TWO ROWS ABOVE IT.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( N = 2k + 1 )</th>
<th>( N = 5 )</th>
<th>( N = 7 )</th>
<th>( N = 9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>different cluster states</td>
<td>( \frac{N!}{k!} )</td>
<td>30</td>
<td>140</td>
<td>630</td>
</tr>
<tr>
<td>different input configurations</td>
<td>( N! )</td>
<td>120</td>
<td>5040</td>
<td>362880</td>
</tr>
<tr>
<td>different input configuration groups</td>
<td>( \binom{N}{k} )</td>
<td>10</td>
<td>35</td>
<td>84</td>
</tr>
<tr>
<td>different spatiotemporal codes for an input</td>
<td>( \binom{N-3}{k-1} )</td>
<td>2</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>different spatiotemporal codes</td>
<td>( \frac{\text{tr}(A^\theta)}{6} )</td>
<td>20</td>
<td>210</td>
<td>1680</td>
</tr>
</tbody>
</table>
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 spatiotemporal codes.

In the next section we describe a strategy for how such codes may be learnt via synchronization and frequency adaptation.

**IV. ADAPTIVE LEARNING**

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

\[
\begin{align*}
  u(\theta_n, \phi_n) &= u_0 \sin(\theta_n - \phi_n), \\
  v(\theta_n, \phi_n) &= v_0 \sin(\theta_n - \phi_n),
\end{align*}
\]

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 learning system to learn the spatiotemporal 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 \( \Omega_n \) as time \( t \) increases. To this end, we define the error variables

\[
\varepsilon_n = \phi_n - \theta_n, \quad \kappa_n = \omega_n - \Omega_n,
\]

and show that the state

\[
\varepsilon_n(t) \equiv 0, \quad \kappa_n(t) \equiv 0,
\]

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

\[
\lim_{t \to \infty} \varepsilon_n(t) = 0, \quad \lim_{t \to \infty} \kappa_n(t) = 0,
\]

for an open set of initial conditions. That is by definition (12) the phases synchronize and the frequencies adapt so that

\[
\lim_{t \to \infty} (\phi_n(t) - \theta_n(t)) = 0, \quad \lim_{t \to \infty} \omega_n(t) = \Omega_n.
\]

**A. Lyapounov function analysis**

For the following analysis we set the noise to zero, i.e., \( \eta = 0 \). Now taking the difference of the first two sets of equations in (1), subtracting \( \Omega_n = 0 \) from the third sets equation in (1) and using the learning rule (11) and definition (12) we obtain

\[
\dot{\varepsilon}_n = \kappa_n + \frac{1}{N} \sum_{m=1}^{N} (g(\theta_n - \theta_m + \varepsilon_n - \varepsilon_m) - g(\theta_n - \theta_m))
\]

\[
- u_0 \sin \varepsilon_n,
\]

\[
\dot{\kappa}_n = -v_0 \sin \varepsilon_n.
\]

We define a Lyapounov function

\[
V(\varepsilon_1, \ldots, \varepsilon_N, \kappa_1, \ldots, \kappa_N) = \sum_{n=1}^{N} \left( \frac{1}{2} f^2(\varepsilon_n) + \frac{1}{2} \kappa_n^2 \right)
\]

where

\[
f(\varepsilon_n) = 2\sqrt{v_0} \sin \left( \frac{\varepsilon_n}{2} \right) \Rightarrow \frac{1}{2} f^2(\varepsilon_n) = v_0 (1 - \cos \varepsilon_n)
\]

\[
\Rightarrow \left( \frac{1}{2} f^2(\varepsilon_n) \right)' = f(\varepsilon_n) f'(\varepsilon_n) = v_0 \sin \varepsilon_n \quad (18)
\]

This function is positive definite for \( \varepsilon_n \in (-\pi, \pi), \kappa_n \in (-\infty, \infty). \) In fact, in (1) \( \kappa_n \) is chosen so that \( \kappa_n = (\frac{1}{2} f^2(\varepsilon_n))' \) in a spirit of adaptive control.

The derivative of (17) along trajectories of the system (16) gives

\[
\frac{dV}{dt}(\varepsilon_1, \ldots, \varepsilon_N, \kappa_1, \ldots, \kappa_N) = -v_0 \sum_{n=1}^{N} \left( u_0 \sin^2 \varepsilon_n - \sin \varepsilon_n \sum_{m=1}^{N} (g(\theta_n - \theta_m + \varepsilon_n - \varepsilon_m) - g(\theta_n - \theta_m)) \right).
\]

(19)

**Lemma 1:** If

\[
u_0 > 2(1 + 2r)(N - 1)/N \quad \text{and} \quad v_0 > 0,
\]

then

\[
\frac{dV}{dt} \leq 0.
\]

(21)

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

**Theorem 1:** If (20) holds then solution (13) is asymptotically stable, that is, (14,15) hold for an open set of initial conditions.

**Proof:** Lemma 1 implies that the derivative of the Lyapounov function (19) is negative semi-definite:

\[
\frac{dV}{dt} = 0 \Rightarrow (\varepsilon_1, \ldots, \varepsilon_N, \kappa_1, \ldots, \kappa_N) = (0, \ldots, 0, k_1, \ldots, k_N),
\]

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 (16), show that in this case \( \kappa_n(t) \equiv 0 \). Consequently, the largest invariant set inside \((0, \ldots, 0, k_1, \ldots, k_N)\) is the origin \((0, \ldots, 0, 0, \ldots, 0)\). According to the Krasovskii-LaSalle invariance principle [2] 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 (19) shows that the synchronization is followed by adaptation for any value of \( v_0 > 0 \).
We provide numerical evidence to illustrate the above analytical results. In Figures 6, 7 we illustrate the ability of the learning system to robustly and repeatably learn a wide range of spatiotemporal codes. In each figure we plot the weighted order parameter (WOP) $R$ given in (8) 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 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 $\Omega_n$ (4). During the grey shaded region the adaptive couplings (11) are non-zero: $u_0 = 0.5$ 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 spatiotemporal 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 as well. Regarding the frequencies of the teaching system $\Omega_n$ (4) 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 6-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 6-cycle in Fig. 2(a).

In the time interval $t \in (0, 100)$ the adaptive couplings are turned off. Fig. 5(a,b) shows that the teaching system (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 the learning system are coupled and their phases rapidly synchronize to each other as shown in Fig. 5(a,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). Notice 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. One may observe in Fig. 5(a,b) 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). Notice that the characteristic time of synchronization and adaptation is shorter than the time period associated with the 6-cycle.

In [2], [3] 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

Fig. 5. Illustrations showing the process of adaptive learning a spatiotemporal code for $N = 5$ oscillators in each system. During the grey 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. Panel (a) shows the weighted order parameter $R$ as a function of time $t$ where the weights are defined by (10) with exponents $[\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5] = [4, 2, 3, 1, 4]$ and panel (b) shows the 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. In panel (c) the 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.
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 6-cycles listed in Table II can be learned. By changing the input configuration one can force both systems to ‘switch’ from one 6-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. Note that the teaching system approaches the 6-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 non-specific. The phases of the two systems rapidly synchronize as shown by the solid and dotted lines in Fig. 6(a,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 6-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 and in the following time interval \( t \in (450, 600) \) (when the adaptive couplings are turned off again) shadowing happens. Note the cluster states \( s_{18} \) and \( s_4 \) appear in both spatiotemporal 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 are increased. However, the number of different input configurations and the number of different spatiotemporal codes increase; see Table III. Indeed, spatiotemporal codes can be learned for larger numbers of oscillators as demonstrated 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 6-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.

These numerical results demonstrate that the constructed learning rules for frequency adaptation allow us to transmit information about the spatiotemporal codes from the teaching system to the learning system in a robust way.

V. CONCLUSIONS AND DISCUSSION

We studied a coupled oscillator system that can exhibit winnerless competition between cluster states and where information about spatially nonhomogeneous time-independent inputs are encoded into spatiotemporal codes. These 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.

In future we wish extend our research in two directions. One is to find spatiotemporal codes in oscillators system where the coupling is not all-to-all but some connections are missing. The other is to be able to transmit information between systems using lower dimensional coupling (e.g., a scalar variable). It seems plausible to use a weighted order parameter (WOP) for this purpose since all information about the spatiotemporal code is contained within a sufficiently long period of the WOP.

In olfactory neurosystems (one or two synapses away from the receptors) odor information is encoded into spatiotemporal codes \([?], [?] \). In these systems (e.g., antennal lobes of insects or olfactory bulbs of mammals), it is suggested that 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 spatiotemporal 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. To pursue this direction one will need to consider higher fidelity neuromodels (e.g., Hodgkin-Huxley model \([?]) \) and possibly incorporate axonal and synaptic delays \([?]) \) as well.

ACKNOWLEDGMENTS

The authors would like to acknowledge conversations with Roman Borisnyuk, Steve Coombs and John Wordsworth.

APPENDIX

A. Determining cluster states and their stability

Substituting the cluster state (7) (or any of its symmetrical copies) into (2) and defining the phase differences \( \tilde{y} = y - w \) and \( \tilde{b} = b - w \) one may obtain

\[
\tilde{\Omega} = \Omega + \frac{1}{2} \left( k g(0) + g(\tilde{y}) + k g(\tilde{y} - \tilde{b}) \right),
\]

\[
\tilde{\Omega} = \Omega + \frac{1}{2} \left( k g(0) + g(\tilde{b}) + k g(\tilde{b} - \tilde{y}) \right),
\]

\[
\tilde{\Omega} = \Omega + \frac{1}{4} \left( g(0) + k g(-\tilde{y}) + k g(-\tilde{b}) \right),
\]

which determine \( \tilde{y}, \tilde{b}, \) and \( \tilde{\Omega} \). For example, for \( N = 5, k = 2 \) using parameters (6) we obtain \( \tilde{y} \approx -1.82, \tilde{b} \approx 1.10 \) and \( \tilde{\Omega} \approx 0.85 \).

Linearizing (2) about the cluster state (7) (or about any of its symmetrical copies) results in the eigenvalues

\[
\lambda_1 = 0,
\]

\[
\lambda_2 = \frac{1}{4} \left( k g'(0) + g'(\tilde{y}) + g'(\tilde{y} - \tilde{b}) \right),
\]

\[
\lambda_3 = \frac{1}{4} \left( k g'(0) + g'(\tilde{b}) + g'(\tilde{b} - \tilde{y}) \right),
\]

\[
\lambda_4 = \mu + i \sqrt{\nu},
\]

\[
\lambda_5 = \mu - i \sqrt{\nu},
\]

where \( \lambda_2 \) and \( \lambda_3 \) has multiplicity \( k - 1 \) and \( \mu, \nu \in \mathbb{R} \) are complicated expressions containing \( g'(\cdot) \) at \( \tilde{y}, -\tilde{y}, \tilde{b}, -\tilde{b}, \tilde{b} - \tilde{y}, \) and \( -\tilde{y} - \tilde{b} \); see \([?]) \). For \( N = 5 \) and parameters (6) we have the eigenvalues \( \lambda_2 = -0.2834, \lambda_3 = 0.1703, \lambda_{4,5} = -0.1012 \pm i 0.2848. \)
For the cluster state (7) the corresponding $N = 2k + 1$-dimensional eigenvectors are

$$v_1 = c_1 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

(25)

$$v_2 = c_{2,1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \chi + c_{2,2} \begin{bmatrix} \chi \\ \vdots \\ \chi \end{bmatrix} \chi + \ldots + c_{2,k} \begin{bmatrix} \chi \\ \vdots \\ \chi \end{bmatrix} \chi,$$

(26)

$$v_3 = c_{3,1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \chi + c_{3,2} \begin{bmatrix} \chi \\ \vdots \\ \chi \end{bmatrix} \chi + \ldots + c_{3,k} \begin{bmatrix} \chi \\ \vdots \\ \chi \end{bmatrix} \chi,$$

(27)

$$v_4 = c_4 \begin{bmatrix} p_y + i \sqrt{q_y} \\ \vdots \\ p_y + i \sqrt{q_y} \end{bmatrix}, \quad v_5 = c_5 \begin{bmatrix} p_y - i \sqrt{q_y} \\ \vdots \\ p_y - i \sqrt{q_y} \end{bmatrix},$$

(28)

where $\chi = -1/(k - 1)$, the scalars $c_1, c_4, c_5$ are arbitrary while one in the set of scalars $\{c_{2,1}, \ldots, c_{2,k}\}$ is zero and the others are arbitrary and the same holds for the set of scalars $\{c_{3,1}, \ldots, c_{3,k}\}$. The expressions of $p_y, q_y, p_b, q_b \in \mathbb{R}$ contain $g'(.)$ at $\vec{y}, -\vec{y}, \vec{b}, -\vec{b}, \vec{y} - \vec{b}$, and $\vec{b} - \vec{y}$; see [7]. For symmetric copies of (7) the eigenvectors can be determined by permuting components of (26-28) appropriately. Note that $\nu$ in $\lambda_{4,5}$ and, consequently, $q_y, q_b$ in $v_{4,5}$ are not necessary positive.

**B. Proof of Lemma 1**

**Proof:** Since $v_0 > 0$ is assumed, to prove that (21) holds one needs to prove that

$$U(\varepsilon_1, \ldots, \varepsilon_N, \kappa_1, \ldots, \kappa_N) := \sum_{n=1}^{N} u_0 \sin^2 \varepsilon_n - \sin \varepsilon_n \sum_{m=1}^{N} (g(\theta_n - \theta_m + \varepsilon_n - \varepsilon_m) - g(\theta_n - \theta_m)) \geq 0.$$  

(29)

Using (5) one can obtain that

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

(30)

and notice that

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

(32)

Let us define the new variables

$$p_n = \sin \frac{\varepsilon_n}{2}, \quad q_n = \cos \frac{\varepsilon_n}{2}.$$  

(33)

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 (30,31) into (29) and using (33) and one can obtain

$$U = \sum_{n=1}^{N} \left( 4u_0 p_n^2 q_n^2 - \sum_{m=1}^{N} A(p_n^2 q_m q_n - p_n p_m q_n^2) \right) \geq \sum_{n=1}^{N} \left( 4u_0 p_n^2 q_n^2 - \frac{2M}{N} p_n^2 \sum_{m=1}^{N} q_m q_n - \frac{2M}{N} q_n^2 \sum_{m=1}^{N} |p_n p_m| \right),$$

(34)

where we used (32) 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 (34) is non-negative if $u_0 > M(N - 1)/N = 2(1 + 4r)(N - 1)/N.$

\[ \blacksquare \]