SIAM J. Applied Dynamical Systems
Vol. 14, No. 1, pp. 188–201

Stability and Long Term Behavior of a Hebbian Network of Kuramoto Oscillators*

Robert W. Hölzel† and Katharina Krischer†

Abstract. We investigate the limit sets of a network of coupled Kuramoto oscillators with a coupling matrix determined by a Hebb rule. These limit sets are the output of the network if used for the recognition of a defective binary pattern out of several given patterns, with the output pattern encoded in the oscillators' phases. We show that if all pairs of given patterns have maximum Hamming distance, there exists a degenerate attractive limit set that contains the steady states corresponding to each of the given patterns. As a result, switching between output patterns occurs for arbitrarily small modifications of the dynamics (for example, due to frequency inaccuracies). Even if the maximum Hamming distance constraint is dropped, numerical results suggest that the structural instability of the vector field persists. We conclude that the unique interchangeability of output patterns in Hebbian networks of Kuramoto oscillators, while sacrificing robustness, makes these networks more flexible than similar neural networks with separated, attractive output states.

Key words. oscillatory network, neural network, Kuramoto oscillators, associative memory, Hebb rule

AMS subject classifications. 34D06, 34D30, 82C32, 92B20

DOI. 10.1137/140965168

1. Introduction.

Networks of coupled phase oscillators with a coupling matrix determined by a Hebb rule have been a popular model of neural network theorists for some time [1, 2, 3, 4, 5, 7, 8, 10, 14, 15, 16, 17, 18], in particular, because they are tractable with powerful mathematical tools originally developed for the description of synchronization phenomena. More recently, such phase oscillators have also been successfully employed in experimental neural networks [6, 12]. In spite of this considerable amount of research that has already gone into this type of network, its dynamics are still not understood as well as other, very similar neural network models (the Hopfield network is a good example [9]). We will explain specifically what we mean by "not well understood" right after introducing the equations and mode of operation of a Hebbian network of Kuramoto oscillators: The network of coupled phase oscillators is described by the dynamical equations

$$\dot{\varphi}_i = \omega_i + \frac{1}{N}\sum_{j=1}^{N} w_{ij}\sin(\varphi_j - \varphi_i), \tag{1.1}$$

where φ_i is the phase shift of an individual oscillator with phase $\vartheta(t) = \Omega t + \omega_i t + \varphi_i(t)$. Ω is the mean frequency, ω_i is a small deviation from Ω with $\sum \omega_i = 0$, and N is the size of the network. The coefficients w_{ij} are the entries of the coupling matrix. The network can be

*Received by the editors April 16, 2014; accepted for publication (in revised form) by B. Sandstede December 5, 2014; published electronically February 24, 2015.
http://www.siam.org/journals/siads/14-1/96516.html
†Physik-Department E19a, Technische Universität München, James-Franck-Strasse 1, D-85748 Garching, Germany (robert.holzel@mytum.de, krischer@ph.tum.de).

used to identify a binary pattern vector $\boldsymbol{\xi}$ ($\xi_i = \pm 1,\ i = 1, \dots N$) as one of a given set of M memorized patterns $\boldsymbol{\xi}^k$, $k = 1, \dots, M$, in the following way: First, the coupling matrix is set to $w_{ij} = \xi_i \xi_j$ for some initial pattern $\boldsymbol{\xi}$. As a result, the phase shifts φ_i will evolve towards a distribution reflecting this pattern, i.e., the overlap

$$m = \frac{1}{N} \sqrt{\left(\sum_{i=1}^{N} \xi_i \sin \varphi_i \right)^2 + \left(\sum_{i=1}^{N} \xi_i \cos \varphi_i \right)^2} \tag{1.2}$$

will become equal to 1 (or slightly smaller if not all frequencies are exactly the same). The overlap is a convenient way to check whether a given pattern is currently represented by the state of the network. Like the system (1.1), m is invariant under global rotations. After this initialization of the network, the coupling coefficients are set to $w_{ij} = \sum_k \xi_i^k \xi_j^k$, which is an application of the Hebb rule. If recognition is successful, the network now evolves towards a state close to the memorized pattern which is closest to its initial state. For example, if $\boldsymbol{\xi}$ is a slightly defective copy of $\boldsymbol{\xi}^1$, the desired final state of the network would be one where $m^1 = 1$ and $m^k = 0,\ k = 2, \dots M$.

The initialization process is well understood—it is equivalent to the synchronization of an array of coupled Kuramoto oscillators, where the overlap m is the order parameter [13]. In contrast, surprisingly little is known about the dynamics of the recognition process; especially, it is not known what exact states the network settles for and how robust these states are. With respect to these questions, previous publications concerned with Hebbian networks of Kuramoto oscillators fall in one of two categories. In publications of the first category, it is just assumed that patterns are recognized sufficiently well for practical purposes and example simulations and/or experiments are presented to corroborate this [6, 7, 10, 12]. Publications of the second category generally use a mean field approach [1, 2, 3, 4, 5, 15, 16, 17, 18], dealing with the order parameters m^k. With the mean field approach, it can be shown that in the thermodynamic limit $N \to \infty$, there is a phase transition between a so-called "glassy state" equivalent to failed pattern recognition and a regime with a "condensed" memorized pattern [2, 15]. This transition depends on the three macroscopic parameters α, $\Delta\omega$, and T, where $\alpha = M/N$ is the load rate, $\Delta\omega$ is the width of the oscillators' frequency distribution, and T is a temperature determining the magnitude of an additional white noise term. In the limit of $\alpha \to 0$ and $T \to 0$, the phase transition is equivalent to the transition in the original Kuramoto model.

One point that is not considered in the mean field approach is the relationship between states in phase space representing different condensed patterns, although it seems to be implicitly assumed that they are separate attractors. Also, little is known about their exact location. This issue is the motivation for the article at hand. We investigate the dynamics of individual oscillators given by (1.1) directly in order to characterize the limit set(s) approached by the dynamics during the pattern recognition process.

Section 2 deals with optimal sets of memorized patterns. Optimal here means that the memorized patterns are mutually orthogonal ($\boldsymbol{\xi}^k \cdot \boldsymbol{\xi}^l = 0$ for $k \neq l$). This is equivalent to all patterns having the maximum possible Hamming distance from each other (since inverted patterns are equivalent to the original pattern, the maximum Hamming distance is $N/2$, not N). It turns out that there are no separate attractors corresponding to the memorized patterns. As a consequence, pattern recognition is not structurally stable.

Section 3 is concerned with the more general case of $\boldsymbol{\xi}^k \cdot \boldsymbol{\xi}^l \neq 0$. Since it is harder to argue analytically in this case, we present some numerical results that corroborate the theory that pattern recognition in this case is no more robust than in the special case of memorized patterns with maximum Hamming distance.

In the discussion in section 4, we give an explanation why this kind of behavior was not found in previous analyses using a mean field approach.

2. Recognition from a set of mutually orthogonal patterns. Consider again the dynamical equations for the recognition step:

$$\dot{\varphi}_i(\varphi_1, \ldots, \varphi_N) = \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \xi_i^k \xi_j^k \sin(\varphi_j - \varphi_i). \tag{2.1}$$

To analyze this equation it is helpful to abstract from the global phase shift symmetry. Therefore, from here on, a fixed point is referred to as attractive if $N-1$ eigenvalues of the Jacobian are smaller than zero (the last one, corresponding to the eigenvector $(1, 1, \ldots, 1)^{\mathrm{T}}$ is always equal to zero, due to the global symmetry). Also, a stationary state is called nonisolated or degenerate only if the fixed point in question is not just trivially nonisolated (i.e., in any neighborhood of the fixed point stationary states that are *not* located along the $(1, 1, \ldots, 1)^{\mathrm{T}}$ eigenvector do exist).

For each pattern $\boldsymbol{\xi}^k$, there is a corresponding one-dimensional manifold of fixed points of (2.1). Let $\boldsymbol{\varphi}^{*k}$ denote any point on this manifold in phase space. In the case of the $M = 1$ pattern (i.e., for an initialization to $\boldsymbol{\xi}^1$), $\boldsymbol{\varphi}^{*1}$ is a global attractor, namely, the synchronized state of a set of coupled Kuramoto oscillators. In the terminology we introduced above, this global attractor is isolated and attractive. For more than one memorized pattern, the $\boldsymbol{\varphi}^{*k}$ are generally isolated and hyperbolically unstable [1]. However, in the special case of mutually orthogonal patterns (i.e., $\boldsymbol{\xi}^l \cdot \boldsymbol{\xi}^m = N\delta^{lm}$, where δ is the Kronecker delta), things are different. For this case, we prove the following three statements.

Theorem 2.1. *The $\boldsymbol{\varphi}^{*k}$ are fixed points with an eigenvalue spectrum of $(N - M) \times -1$ and $M \times 0$.*

Proof. Consider the Jacobian $\boldsymbol{J}$ of (2.1):

$$J_{ij} = \frac{\partial \dot{\varphi}_i}{\partial \varphi_j} = \frac{1}{N} \sum_{k=1}^{M} \left(\xi_i^k \xi_j^k \cos(\varphi_j - \varphi_i) - \delta_{ij} \sum_{p=1}^{N} \xi_i^k \xi_p^k \cos(\varphi_p - \varphi_i) \right). \tag{2.2}$$

At the state $\boldsymbol{\varphi}^{*\boldsymbol{l}}$ corresponding to pattern $\boldsymbol{\xi}^l$ this becomes

$$J_{ij}|_{\boldsymbol{\varphi}^{*\boldsymbol{l}}} = \frac{1}{N} \sum_{k=1}^{M} \left(\xi_i^k \xi_j^k \xi_j^l \xi_i^l - \delta_{ij} \sum_{p=1}^{N} \xi_i^k \xi_p^k \xi_p^l \xi_i^l \right) = \frac{1}{N} \sum_{k=1}^{M} \xi_i^k \xi_j^k \xi_j^l \xi_i^l - \delta_{ij} \delta^{kl} \xi_i^k \xi_i^l. \tag{2.3}$$

Therefore, $\boldsymbol{J}$ can be written as

$$\boldsymbol{J}|_{\boldsymbol{\varphi}^{*\boldsymbol{l}}} = \sum_{k=1}^{M} \boldsymbol{A}^{kl} - \mathbb{I} \tag{2.4}$$

with $A_{ij}^{kl} = \xi_i^k \xi_j^k \xi_i^l \xi_j^l / \mathrm{N}$.

$\boldsymbol{A}^{kl}$ has a single nonzero eigenvalue for the eigenvector $\boldsymbol{\chi}^{kl} = \boldsymbol{\xi}^k \circ \boldsymbol{\xi}^l$, where "$\circ$" denotes componentwise multiplication. The corresponding eigenvalue is 1. This can be seen by calculating $\boldsymbol{A}^{kl} \cdot \boldsymbol{\chi}$ with an arbitrary vector $\boldsymbol{\chi}$:

$$\left(\boldsymbol{A}^{kl} \cdot \boldsymbol{\chi}\right)_i = \frac{1}{N} \sum_{j=1}^{N} \xi_i^k \xi_j^k \xi_i^l \xi_j^l \chi_j = \frac{1}{N} \xi_i^k \xi_i^l \sum_{j=1}^{N} \xi_j^k \xi_j^l \chi_j = \frac{1}{N} \chi_i^{kl} \boldsymbol{\chi}^{kl} \cdot \boldsymbol{\chi}. \tag{2.5}$$

Therefore all vectors orthogonal to $\boldsymbol{\chi}^{kl}$ are eigenvectors of $\boldsymbol{A}^{kl}$ with eigenvalue 0, while $\boldsymbol{\chi}^{kl}$ is an eigenvector with eigenvalue 1.

In conjunction with (2.4), this result shows that $\boldsymbol{J}|_{\boldsymbol{\varphi}^{*l}}$ has M eigenvalues equal to 0 with mutually orthogonal eigenvectors $\boldsymbol{\chi}^{kl}$, while the rest of the eigenvalues are equal to -1. Without loss of generality, this holds for all $\boldsymbol{J}|_{\boldsymbol{\varphi}^{*k}}$, where $k = 1, \ldots, M$. ■

Theorem 2.2. *The $\boldsymbol{\varphi}^{*k}$ are nonisolated; they are part of a single, connected set of degenerate stationary states which comprises all straight lines connecting any pair $\boldsymbol{\varphi}^{*k}$, $\boldsymbol{\varphi}^{*l}$ in phase space defined by*

$$\dot{\boldsymbol{\varphi}}|_{\boldsymbol{\varphi}^{*k} + (\boldsymbol{\varphi}^{*l} - \boldsymbol{\varphi}^{*k})u} = 0 \tag{2.6}$$

with the parameter $u \in \mathbb{R}$ and $k, l \in \{1, \ldots, M\}$.

Proof. For the second proposition, we compute the Taylor expansion of the dynamics along the eigenvectors with eigenvalue zero of $\boldsymbol{J}|_{\boldsymbol{\varphi}^{*k}}$. We show that the flow is zero for all points in phase space along these vectors. We then show that the eigenvectors with eigenvalue zero connect pairs of memorized patterns in phase space.

The vector field at points along the vector $\boldsymbol{\chi}^{kl}$, starting at $\boldsymbol{\varphi}^{*k}$, and its first derivative are given by

$$\dot{\varphi}_i(u) = \frac{1}{N} \sum_{j=1}^{N} w_{ij} \sin(\varphi_j^{*k} - \varphi_i^{*k} + (\chi_j^{kl} - \chi_i^{kl})u), \tag{2.7}$$

$$\partial_u \dot{\varphi}_i(u) = \frac{1}{N} \sum_{j=1}^{N} w_{ij} (\chi_j^{kl} - \chi_i^{kl}) \cos(\varphi_j^{*k} - \varphi_i^{*k} + (\chi_j^{kl} - \chi_i^{kl})u). \tag{2.8}$$

Here, u is a real parameter and both $\dot{\varphi}_i(0) = 0$ and $\partial_u \dot{\varphi}_i(0) = 0$ hold. The higher order derivatives with respect to u are given by

$$\partial_u^{(2n)} \dot{\varphi}_i(0) = \frac{1}{N} \sum_{j=1}^{N} w_{ij} (-1)^n (\chi_j^{kl} - \chi_i^{kl})^{2n} \sin(\varphi_j^* - \varphi_i^*) = 0 \tag{2.9}$$

and

$$\partial_u^{(2n+1)} \dot{\varphi}_i(0) = \frac{1}{N} \sum_{j=1}^{N} w_{ij} (-1)^n (\chi_j^{kl} - \chi_i^{kl})^{2n+1} \cos(\varphi_j^* - \varphi_i^*), \tag{2.10}$$

where $n \in \mathbb{N}$. Since all entries of $\boldsymbol{\chi}^{kl}$ have the same absolute value, namely, $|\chi_i^{kl}| = 1$, the last equation reduces to

$$\partial_u^{(2n+1)} \dot{\varphi}_i(0) = (-1)^n 2^{2n} \partial_u \dot{\varphi}_i(0) = 0. \tag{2.11}$$

This means, all derivatives with respect to u vanish for $u = 0$. Since $\dot{\varphi}_i(u)$ is analytical, it must be identical to zero for all u. As a result, a whole line of stationary states extends from $\boldsymbol{\varphi}^{*k}$ in phase space along each of the eigenvectors $\boldsymbol{\chi}^{kl}$. Since $\boldsymbol{\chi}^{kk} = (1, 1, \ldots, 1)^T$ corresponds to the global phase shift symmetry, there remain $M - 1$ zero eigenvectors in the subspace orthogonal to the $(1, 1, \ldots, 1)^T$ direction. These vectors connect pairs of patterns in phase space, which can be verified by using the fact that

$$\chi_j^{kl} - \chi_i^{kl} \propto \varphi_j^{*l} - \varphi_i^{*l} - \varphi_j^{*k} - \varphi_i^{*k} \tag{2.12}$$

to substitute for $\boldsymbol{\chi}^{kl}$ in (2.7), resulting in the left-hand side of (2.6). ■

Theorem 2.3. *The set defined by* (2.6) *as a whole is attractive, which means that the* $\boldsymbol{\varphi}^{*k}$ *are neutrally stable, if*

$$\forall m, l \neq m : \operatorname{span}(\boldsymbol{\chi}^{km},\ k \neq l, m) \cap \operatorname{span}(\boldsymbol{\chi}^{kl},\ k \neq l, m) = 0. \tag{2.13}$$

Proof. Consider the Jacobian $\boldsymbol{J}$ along the vector joining $\boldsymbol{\varphi}^{*l}$ and $\boldsymbol{\varphi}^{*m}$, $l \neq m$, $l, m \in \{1, \ldots, M\}$, in phase space. Points along this vector are given by $\boldsymbol{\varphi}(u) = \boldsymbol{\varphi}^{*l} + (\boldsymbol{\varphi}^{*m} - \boldsymbol{\varphi}^{*l})u$, $u \in \mathbb{R}$. The individual summands constituting the entries of the Jacobian given in (2.2) evaluate to

$$\begin{aligned}
\xi_i^k \xi_j^k \cos(\varphi_j(u) - \varphi_i(u)) &= \xi_i^k \xi_j^k \cos(\varphi_j^{*l} - \varphi_i^{*l} + (\varphi_j^{*m} - \varphi_j^{*l})u - (\varphi_i^{*m} - \varphi_i^{*l})u) \\
&= \xi_i^k \xi_j^k \cos\left((\xi_i^l \xi_j^l - 1)\frac{\pi}{2} + (\xi_i^m \xi_j^m - \xi_i^l \xi_j^l)\frac{\pi}{2}u\right) \\
&= \xi_i^k \xi_j^k \xi_i^l \xi_j^l \cos\left((\xi_i^m \xi_j^m - \xi_i^l \xi_j^l)\frac{\pi}{2}u\right) \\
&= \xi_i^k \xi_j^k \left(\xi_i^l \xi_j^l (1 - \gamma) + \xi_i^m \xi_j^m \gamma\right),
\end{aligned}$$

where $\gamma = (1 - \cos(u\pi))/2$. $\boldsymbol{J}$ can now be written as

$$\boldsymbol{J} = (1 - \gamma) \sum_{k=1}^{M} \boldsymbol{A}^{kl} + \gamma \sum_{k=1}^{M} \boldsymbol{A}^{km} - \mathbb{I}. \tag{2.14}$$

Here, $\boldsymbol{A}^{xy}$ are the matrices introduced in (2.4). The eigenvalue spectra of the three terms on the right-hand side of (2.14) are

- $M \times (1 - \gamma)$ and $(N - M) \times 0$ for the first term,
- $M \times \gamma$ and $(N - M) \times 0$ for the second term, and
- $N \times -1$ for the third term.

This means that $\lambda \leq 0$ for all eigenvalues λ of $\boldsymbol{J}$, since the largest eigenvalue of a sum of hermitian matrices cannot surpass the sum of the largest eigenvalues of the summands (see, for example, [11]). $\boldsymbol{J}$ has two eigenvalues that are always equal to zero, one belonging to the eigenvector $(1, 1, \ldots, 1)^T$ because of the global phase shift invariance, the other belonging to the eigenvector $\boldsymbol{\chi}^{lm}$, which is the direction of the invariant connection between $\boldsymbol{\varphi}^{*l}$ and $\boldsymbol{\varphi}^{*m}$ in phase space.

For more zero eigenvalues, there must be one or more vectors $\boldsymbol{\chi} \in \operatorname{span}(\boldsymbol{\chi}^{kl},\ k \neq l, m)$ which simultaneously fulfill $\boldsymbol{\chi} = \operatorname{span}(\boldsymbol{\chi}^{km},\ k \neq l, m)$. By enforcing (2.13), additional zero

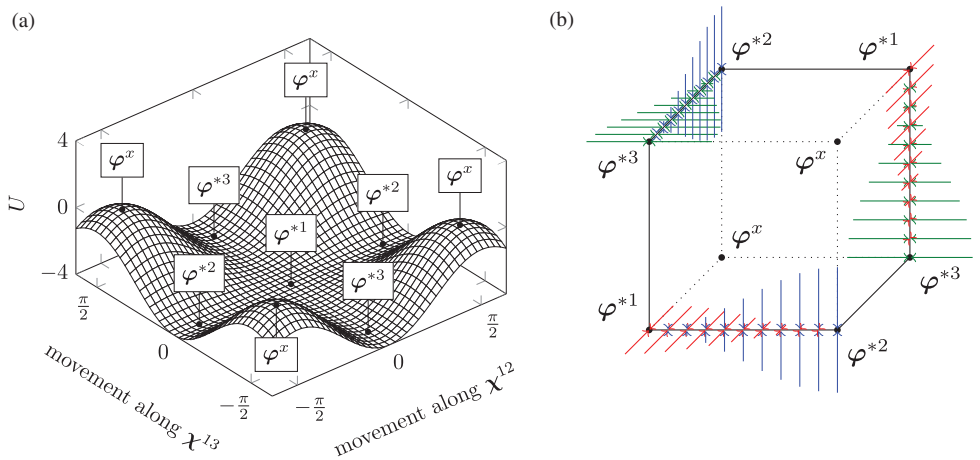

Figure 1. (a) *Surface plot of the potential U of the network dynamics in recognition mode (given by (2.15)) on a two-dimensional plane in $\boldsymbol{\varphi}$-space spanned by $\boldsymbol{\chi}^{12} = \boldsymbol{\xi}^1 \circ \boldsymbol{\xi}^2$ and $\boldsymbol{\chi}^{13} = \boldsymbol{\xi}^1 \circ \boldsymbol{\xi}^3$. $\boldsymbol{\varphi}^{*1}$ is the origin. The location of special patterns in the plane is marked by labels.* (b) *Three-dimensional section of $\boldsymbol{\varphi}$-space spanned by $\boldsymbol{\chi}^{12}$, $\boldsymbol{\chi}^{13}$, and $\boldsymbol{\chi}^{23}$ (so the front face of the cube is part of the plane shown in* (a)*. $\boldsymbol{\varphi}^{*1}$ is the origin. Solid lines mark stationary states. The direction of the eigenvectors with negative eigenvalues for these states is indicated by the colored arrows (red: $\boldsymbol{\chi}^{23}$, green: $\boldsymbol{\chi}^{12}$, blue: $\boldsymbol{\chi}^{13}$). The absolute value of the eigenvalue is indicated by the length of the line, ranging from 0 to 1. Note that all eigenvalues along directions other than $(1,1,1,1,1,1,1,1)^T$, $\boldsymbol{\chi}^{12}$, $\boldsymbol{\chi}^{13}$, and $\boldsymbol{\chi}^{23}$ are equal to -1 for these states. Also note that the global invariant direction is orthogonal to the cube.*

eigenvalues are avoided and therefore the degenerate set of states as a whole is attracting. Because of this, it is also established that the $\boldsymbol{\varphi}^{*k}$ are neutrally stable. ∎

We include Theorem 2.3 to corroborate that, *in general*, the set described in Theorem 2.2 is a maximally connected component of the set of equilibria. If (2.13) is violated, the set may still be maximally connected, but we are unable to prove or disprove it. Unfortunately, it is not particularly intuitive to grasp what condition (2.13) means, especially for a larger set of patterns. All we can say is that with decreasing load rate $\alpha = M/N$ it becomes less and less likely that (2.13) is violated if patterns are picked randomly. Another way to put it is that in a sufficiently large network, it is always possible to choose patterns such that (2.13) holds. In our numerical simulations with small numbers of oscillators and patterns, violations of (2.13) never occurred. Constructed examples where (2.13) *is* violated seem to go hand in hand with a serious overload of the network (for example, you could take $M = N$ mutually orthogonal patterns and end up with $\boldsymbol{J} = 0$, i.e., the degenerate set of fixed points is the entire phase space). With these considerations in mind, we assume that violations of (2.13) will not play an important role when dealing with the neural network at hand.

For a better impression of the location of the degenerate state in phase space, consider an example with $N = 8$ oscillators and $M = 3$ memorized mutually orthogonal patterns $\boldsymbol{\xi}^1 = (1,1,1,1,1,1,1,1)^{\mathrm{T}}$, $\boldsymbol{\xi}^2 = (1,1,1,1,-1,-1,-1,-1)^{\mathrm{T}}$, and $\boldsymbol{\xi}^3 = (1,1,-1,-1,1,1,-1,-1)^{\mathrm{T}}$. Figure 1(a) visualizes the potential function of (2.1) (i.e., $-\partial U/\partial\varphi_i = \dot{\varphi}_i$), given by

$$U = -\frac{1}{16}\sum_{i=1}^{8}\sum_{j=1}^{8}\sum_{k=1}^{3}\xi_i^k\xi_j^k\cos(\varphi_j - \varphi_i) \tag{2.15}$$

in the vicinity of $\boldsymbol{\varphi}^{*1} = 0$ on a two-dimensional cross section of phase space spanned by the vectors $\boldsymbol{\chi}^{12}$ and $\boldsymbol{\chi}^{13}$, which is the center manifold of $\boldsymbol{\varphi}^{*1}$. $\boldsymbol{\varphi}^{*1}$ is a representation of pattern $\boldsymbol{\xi}^1$. By moving a distance of $\pi/2$ from $\boldsymbol{\varphi}^{*1}$ along one of the two axes, one arrives at a representation of patterns $\boldsymbol{\xi}^2$ and $\boldsymbol{\xi}^3$, respectively,

$$\boldsymbol{\varphi}^{*2} = (\pi/2, \pi/2, \pi/2, \pi/2, -\pi/2, -\pi/2, -\pi/2, -\pi/2)^{\mathrm{T}},$$
$$\boldsymbol{\varphi}^{*3} = (\pi/2, \pi/2, -\pi/2, -\pi/2, \pi/2, \pi/2, -\pi/2, -\pi/2)^{\mathrm{T}}.$$

By moving a distance of $\pi/2$ along both $\boldsymbol{\chi}^{12}$ and $\boldsymbol{\chi}^{13}$ from the origin, one arrives at another pattern orthogonal to the other three:

$$\boldsymbol{\varphi}^{x} = (\pi/2, \pi/2, -\pi/2, -\pi/2, -\pi/2, -\pi/2, \pi/2, \pi/2)^{\mathrm{T}}.$$

Since this pattern does not enter the coupling function, the potential function behaves differently here, showing a maximum.

If the dynamics were confined to the $\boldsymbol{\chi}^{12}$-$\boldsymbol{\chi}^{13}$–plane, the system could be expected to settle for any state in the potential valleys between $\boldsymbol{\varphi}^{*1}$ and $\boldsymbol{\varphi}^{*2}$ or $\boldsymbol{\varphi}^{*1}$ and $\boldsymbol{\varphi}^{*3}$, depending on the initial conditions (if there are no perturbations; for pattern recognition, the final state will be closer to one of the patterns, because the initial pattern is closer to this pattern as well). These values would then constitute the attractive limit set. However, since the potential landscape for $\boldsymbol{\varphi}^{*2}$ or $\boldsymbol{\varphi}^{*3}$ looks analogous to Figure 1(a) in the $\boldsymbol{\chi}^{12} - \boldsymbol{\chi}^{23}$ and $\boldsymbol{\chi}^{13} - \boldsymbol{\chi}^{23}$ subspaces, respectively, the minimal subspace containing the attractive limit set is three-dimensional.

Figure 1(b) shows the limit set in this subspace. The four patterns lie on the corners of a cube in phase space. Points on the edges that are not adjacent to $\boldsymbol{\varphi}^{x}$ are fixed points with eigenvalues $\lambda \leq 0$. The eigenvectors belonging to negative eigenvalues are embedded in the faces of the cube (the absolute value of λ is visualized by the colored arrows along the edges). As a consequence, the degenerate state as a whole is an attractor. To visualize the flow on the two-dimensional stable manifold surrounding the steady states on the lines connecting the $\boldsymbol{\varphi}^{*k}$, you can use the colored arrows in Figure 1(b): Starting out at $\boldsymbol{\varphi}^{*1}$ and going right towards $\boldsymbol{\varphi}^{*2}$, attraction is first strong in the $\boldsymbol{\chi}^{23}$-direction and very weak in the $\boldsymbol{\chi}^{13}$-direction. In the middle between $\boldsymbol{\varphi}^{1}$ and $\boldsymbol{\varphi}^{2}$, attraction is the same in both directions. Going even further towards $\boldsymbol{\varphi}^{*2}$, the two directions have reversed roles, with attraction now being strong in the $\boldsymbol{\chi}^{13}$-direction and weak in the $\boldsymbol{\chi}^{23}$-direction. At $\boldsymbol{\varphi}^{*1}$ and $\boldsymbol{\varphi}^{*2}$, the attraction in one of the two directions vanishes completely, and this direction becomes part of the two-dimensional center manifold of the steady state in question.

Since the vector field is not structurally stable, a recognized pattern can be destroyed by arbitrarily small deviations of the network from ideal behavior. For example, if the frequencies of the oscillators are not exactly the same, an additional small constant term appears in the dynamics: $\dot{\varphi}_i \to \dot{\varphi}_i + \omega_i$, $\omega_i \ll 1$. To illustrate the effects of such inaccuracies, a numerical simulation of

$$\dot{\varphi}_i = \omega_i + \frac{1}{8} \sum_{j=1}^{8} \sum_{k=1}^{3} \xi_i^k \xi_j^k \sin(\varphi_j - \varphi_i) \tag{2.16}$$

was performed. To obtain the ω_i, first 8 values ω_i' were chosen randomly from the interval $[0\,\mathrm{s}^{-1}, \Delta\omega = 0.02\,\mathrm{s}^{-1}]$ with a uniform probability distribution. Then, the average value $\overline{\omega'}$

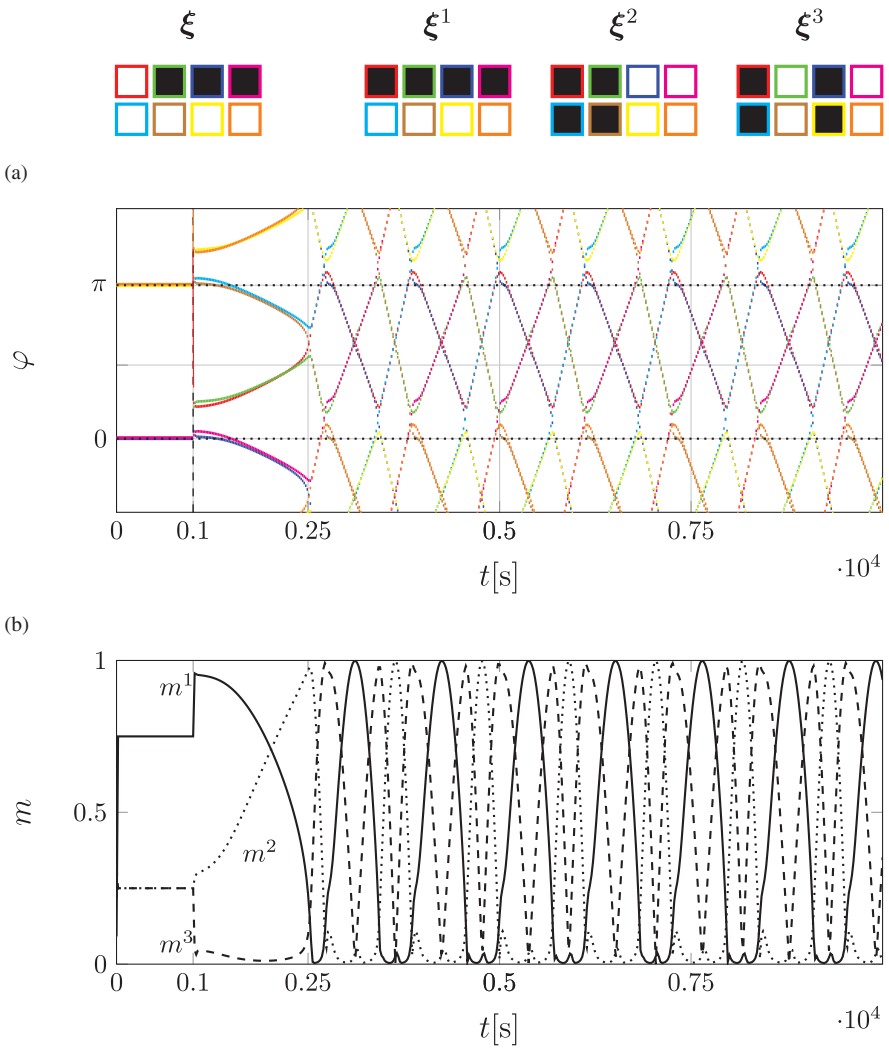

Figure 2. (a) *Numerical time integration of the phase shifts in* (2.16) *with random initial conditions,* $N = 8$ *oscillators,* $M = 3$ *memorized patterns, and one erroneous bit in the initial pattern. The initial pattern* $\boldsymbol{\xi}$ *and the memorized patterns* $\boldsymbol{\xi}^k$ *are depicted above. Black squares correspond to* $\xi_i = -1$*, white squares correspond to* $\xi_i = 1$*. The color of the border around each square corresponds to the color of the phase shift curve in the plot. The coupling was switched to recognition mode at* $t = 1000\,\mathrm{s}$*. The choice of the frequency deviations* ω_i *in* (2.16) *is described in the text.* (b) *Time evolution of the overlaps* m^1*,* m^2*, and* m^3 *of the pattern represented by the network with the memorized patterns.*

was subtracted from each ω_i' to obtain ω_i. The latter was done to avoid a global change in frequency leading to uniformly rotating solutions.

Adding the small terms ω_i to the dynamics in (2.1) is equivalent to introducing a small random tilt in the potential landscape of Figure 1(a). As a consequence, the network does not settle for a steady state anymore during recognition (see Figure 2(a)). In the example at hand, after the initial fast recognition (i.e., quickly evolving towards a state close to $\boldsymbol{\varphi}^{*1}$), the system begins to "roll" towards $\boldsymbol{\varphi}^{*2}$. At this crossroad, the system leaves the $\boldsymbol{\chi}^{12}$-$\boldsymbol{\chi}^{13}$

plane and moves towards $\boldsymbol{\varphi}^{*3}$; from there on the system cycles through the patterns $\boldsymbol{\varphi}^{*3}$-$\boldsymbol{\varphi}^{*1}$-$\boldsymbol{\varphi}^{*3}$-$\boldsymbol{\varphi}^{*2}$. This can be seen especially well in the time evolution of the overlaps m^1, m^2, and m^3 (Figure 2(b)). Another way to imagine the motion in phase space is to look at the cube in Figure 1(b). The solid lines (and their extensions in the cyclical phase space), form the attractive limit set of the network. At the beginning of the recognition process the system starts out close to the $\boldsymbol{\varphi}^{*1}$-corner and gets rapidly attracted to the $\boldsymbol{\varphi}^{*1}$-$\boldsymbol{\varphi}^{*2}$–edge, moving very slowly towards $\boldsymbol{\varphi}^{*2}$ due to the effect of the perturbation terms. Arriving at $\boldsymbol{\varphi}^{*2}$, the system speeds up and continues along the $\boldsymbol{\varphi}^{*2}$-$\boldsymbol{\varphi}^{*3}$–edge, and so on. Depending on the distribution of the ω_i, different movement patterns occur. In particular, the originally recognized patterns may not be visited again at all during the traversal of the limit set.

3. Recognition from a general pattern set. Many derivations in the section above depend on the mutual orthogonality of the memorized patterns. Therefore, it is much harder to make analytically founded statements about the general case of arbitrary memorized patterns. However, it is reasonable to assume that a remnant of the stable limit set remains for pattern vectors that are near orthogonal—if not as a degenerate stationary state, then at least as a network of slow manifolds, interspersed with some combination of stable and unstable, isolated fixed points, where some of the eigenvalues might be very close to, but not equal to, zero. In particular, attractors in the vicinity of the patterns are expected.

With this working hypothesis in mind, a series of numerical integrations of

$$\dot{\varphi}_i = \omega_i + \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\xi_i^k\xi_j^k\sin(\varphi_j - \varphi_i) \tag{3.1}$$

in a network of $N = 100$ oscillators with $M = 3$ memorized patterns was performed. This time, patterns were random, with the constraint that no mutually orthogonal patterns were allowed, and the initial patterns differed from $\boldsymbol{\xi}^1$ in 10 entries. The random deviations ω_i were selected as described in the last section, with a spread of $\Delta\omega = 0.02\,\mathrm{s}^{-1}$. As the values for N, M, and the number of defective bits are well within the storage and retrieval capacity of the network, the short term recognition of pattern $\boldsymbol{\xi}^1$ worked each time.

Figure 3(a) shows the time evolution of m^1, m^2, and m^3 for one particular simulation. It looks very similar to Figure 2(b). After the initial recognition of pattern $\boldsymbol{\xi}^1$, a cyclical oscillation between all three patterns sets in, indicating that the "remnant of the limit set" mentioned above indeed exists.

Figure 3(b) shows data of another simulation. In this case, the system settles for a stationary state which represents the correctly recognized pattern *even in the presence of noise*. This is never observed for the case of mutually orthogonal patterns and corroborates the assumption that there is an isolated attractor.

Surprisingly, as Figure 3(c) shows, it also happens that after some excursion in phase space, the system settles for an attractor that is close to an *incorrectly* recognized pattern. In this particular case, the system approaches a state representing pattern $\boldsymbol{\xi}^3$ for very long times. This does not necessarily contradict the idea that there is an attractor in the vicinity of $\boldsymbol{\xi}^1$. Its attraction could be too weak to balance the particular choice of ω_i.

To get an idea how small the deviations must become to stop pattern switching (as in Figure 3(a)) entirely, the spread of deviations $\Delta\omega$ was varied over several orders of magnitude,

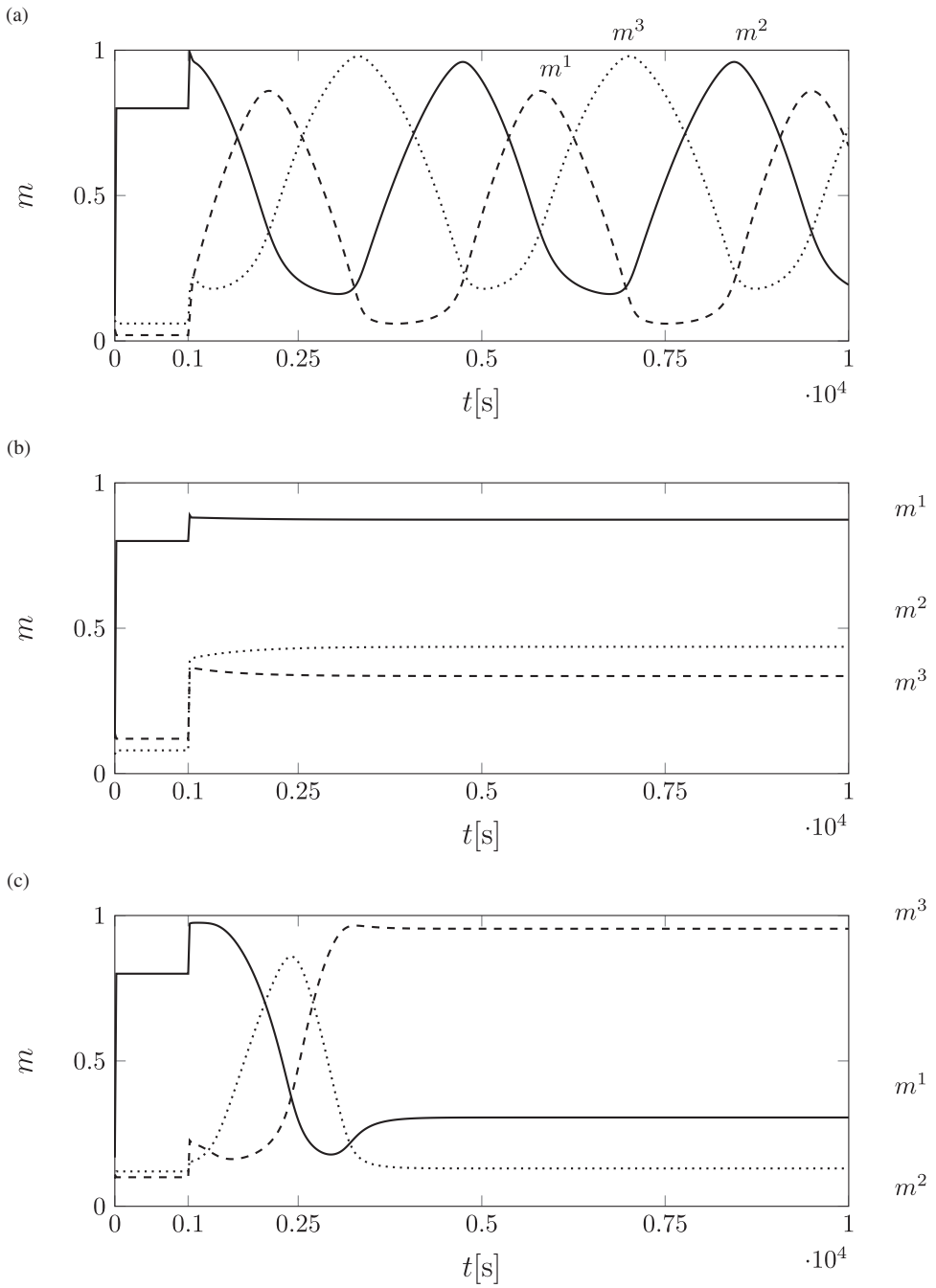

Figure 3. *Time evolution of the overlaps m^1, m^2, and m^3 for the numerical integration of (2.16) with random initial conditions, $N = 100$ oscillators, $M = 3$ memorized random patterns, and ten erroneous bits (compared to the correct pattern $\boldsymbol{\xi}^1$) in the initial pattern. The coupling was switched to recognition mode at $t = 1000\,\mathrm{s}$. The choice of the frequency deviations ω_i in (2.16) is described in the text.* (a) *pattern switching,* (b) $\boldsymbol{\xi}^1$ *is the long term solution,* (c) $\boldsymbol{\xi}^3$ *is the long term solution.*

Table 1

The table shows, for different values of the frequency detuning $\Delta\omega$, how often each possible type of long term behavior occurred in a series of 100 numerical simulations of (3.1) with $N = 100$ oscillators, $M = 3$ randomly selected memorized patterns (with equal probability for both states), and 10 erroneous bits in the initial pattern. $\boldsymbol{\xi}^1$ was always the correct pattern. The relation between $\Delta\omega$ and the frequency deviations ω_i in (3.1) is explained in the text. The long term behavior of the system was evaluated at $t = 1000/\Delta\omega$, the coupling was switched to recognition mode at $t = 1000\,\mathrm{s}$. Note that the short term pattern recognition was always successful in these simulations.

$\Delta\omega$	Long term behavior				
	$\boldsymbol{\xi}^1$	$\boldsymbol{\xi}^2$	$\boldsymbol{\xi}^3$	Switching	Transient
$1 \times 10^{-5}\,\mathrm{s}^{-1}$	15	19	32	21	13
$3 \times 10^{-5}\,\mathrm{s}^{-1}$	27	28	23	8	14
$1 \times 10^{-4}\,\mathrm{s}^{-1}$	19	26	26	12	17
$3 \times 10^{-4}\,\mathrm{s}^{-1}$	25	19	28	9	19
$1 \times 10^{-3}\,\mathrm{s}^{-1}$	23	21	27	14	15
$3 \times 10^{-3}\,\mathrm{s}^{-1}$	21	21	26	11	21
$1 \times 10^{-2}\,\mathrm{s}^{-1}$	26	18	25	15	16
$3 \times 10^{-2}\,\mathrm{s}^{-1}$	19	15	19	27	20
$1 \times 10^{-1}\,\mathrm{s}^{-1}$	11	29	20	15	25

from $\Delta\omega = 1\,\mathrm{s}^{-1}$ down to $\Delta\omega = 1 \times 10^{-5}\,\mathrm{s}^{-1}$. For each value of $\Delta\omega$, a hundred simulations were run for a time of $1000/\Delta\omega$. After this time, the outcome was determined as one of five possible results:

1. The system settled for $\boldsymbol{\xi}^1$ in the long run.
2. The system settled for $\boldsymbol{\xi}^2$ in the long run.
3. The system settled for $\boldsymbol{\xi}^3$ in the long run.
4. The system exhibited pattern switching.
5. The system was still in a transient.

Intuitively, one would expect an increase in correct pattern recognitions and a decrease in faulty recognitions and pattern switching towards smaller values of $\Delta\omega$, due to the fact that even weak attractors become stronger than the effects of perturbation terms. The actual result is given in Table 1. Remarkably, there is no recognizable trend towards better long term pattern recognition at low perturbation strength at all. Instead, all patterns get chosen with roughly the same frequency. Also, the number of simulations exhibiting pattern switching does not diminish in any way with reduced values of $\Delta\omega$. All this speaks against the presence of weak attractors responsible for the long term behavior shown in Figures 3(b) and (c), at least not in the *unperturbed* dynamics.

Looking at Table 1, one might wonder what happens at $\Delta\omega = 0$, and whether this type of random selection of patterns persists even then. The answer is no. At $\Delta\omega = 0$, we always observe perfect long term pattern recognition, because in this limit, the transients become infinitely long, i.e., there is no switching between patterns. Thus, the behavior of the network at $\Delta\omega = 0$ is qualitatively different from the behavior in the presence of frequency deviations, which is to be expected if the vector field at $\Delta\omega = 0$ is not structurally stable. A possible, yet very speculative scenario that could explain the numerical results is as follows:

In the case of nonorthogonal memorized patterns, there exists a set of (up to symmetries) M special solutions $\boldsymbol{\varphi}^{k*'}$ that are close to the memorized patterns and have M zero eigenvalues.

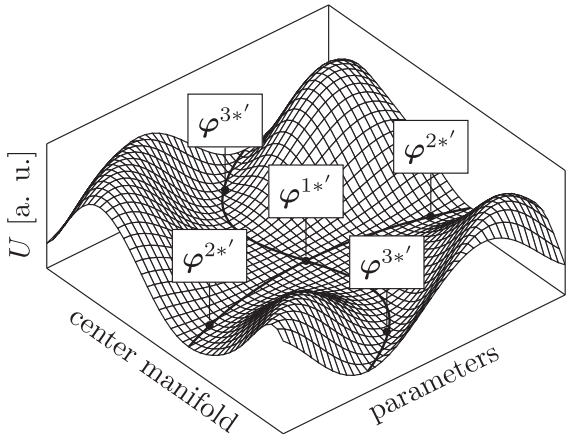

Figure 4. *Schematic plot of a conceivable potential function U of the network dynamics in recognition mode on a two-dimensional slice of the three-dimensional center manifold of $\boldsymbol{\varphi}^{1*'}$ in $\boldsymbol{\varphi}$-space for a system with $M = 3$ nonorthogonal memorized patterns. Fixed points with M zero eigenvalues that represent imperfect memorized patterns are marked with labels. The thick black lines are equipotential curves with minimal potential that connect $\boldsymbol{\varphi}^{1*'}$ to the other memorized patterns.*

These solutions are connected by a network of degenerate steady state solutions much like the perfect patterns in the orthogonal case. However, the degenerate state is not a network of straight lines now, due to the lesser symmetry of the system. Figure 4 shows a stylized potential landscape with the curved invariant manifolds. While in the ideal system no single point in phase space is attractive, an arbitrarily small perturbation can stabilize points close to the invariant manifold, if the effect "tilts" the potential landscape in the right way. Note that this stabilizing mechanism does not work in the case of orthogonal patterns, where the invariant curves connecting the $\boldsymbol{\varphi}^{k*}$ in phase space are straight lines. As already stated, this explanation is highly speculative. Whether true or not however—the fact remains that the correctly recognized patterns are not robust on a very long time scale.

4. Discussion and conclusion. As already mentioned in the introduction, up to now the robustness of pattern recognition using (1.1) has only been studied with a mean field approach before, resulting in equations for the overlaps m^k. This macroscopic view of the network establishes that there is a surface of critical values in the three-dimensional $(\alpha, \Delta\omega, T)$ parameter space that separates the regime in which the system is capable of pattern recognition from the glassy regime. It fails, however, to adequately address the question whether and how the different condensed patterns are separated from each other within the "memorized" state. The reason for this failure is the assumption that during pattern recognition, the network initially approaches a state characterized by $m^1 = \mathcal{O}(1)$ and $m^k = \mathcal{O}(N^{-1/2})$, $k = 2, \ldots, M$, which has been the basis of any mean field analysis so far [1, 2, 3, 4, 5, 15, 16, 17, 18]. *The assumption for the m^k is not necessarily true just because the assumption for m^1 holds,* because, as discussed above, the limit set of the network is actually restricted to a low-dimensional subspace containing all possible outcome patterns. For example, a state on the line connecting $\boldsymbol{\varphi}^{*1}$ and $\boldsymbol{\varphi}^{*2}$ in Figure 1(b) with an overlap of $m^1 = 0.9$ with the correctly

recognized pattern would still have an overlap of $m^2 \approx 0.44$ with an incorrect pattern ($(m^1)^2 + (m^2)^2 = 1$), even if N were increased to arbitrarily large values.

To alleviate the shortcomings of the mean field approach, for this article we have investigated the long term robustness of the recognized patterns by looking at the dynamics of the individual oscillators. The main result is that *the network represented by* (2.1) *does not "pick" one of the memorized patterns* $\boldsymbol{\xi}^i$ *by settling for a state corresponding to the order parameter* $m^i \approx 1$. *The system may settle not only for states representing the patterns themselves, but also any intermediate state between two patterns. Once such an intermediate state is reached, the system does not move towards the closer of the two patterns any more. For the case of mutually orthogonal patterns, we presented a proof that this happens, because the stationary states representing the patterns in phase space are neutrally stable and part of a single, attractive limit set. For the general case, we presented some numerical evidence that the situation is analogous.* As a consequence of this structure in phase space, recognized patterns are not robust. In the presence of arbitrarily small frequency inaccuracies $\Delta\omega$, switching between any subset of the memorized patterns can occur—there is no critical value for $\Delta\omega$ below which the phenomenon disappears.

The fact that the recognized pattern states are not separate attractors is extremely relevant for any application, because it means that pattern recognition results obtained with nonideal networks are only temporary. We would like to emphasize, however, that the lack of robustness of the recognized patterns is not *exclusively* a bad thing. First, as the simulations show, for frequency inaccuracies that are small enough there is a transient period in which the pattern is correctly recognized. This transient is long compared to the time it takes the network to perform the initial recognition. Hence, for all practical purposes, there is enough time to read out and process the recognized patterns. Second, think of the network as a model, however crude, of a neural processor dealing with a set of different instances of a common concept (like letters, for example). Each of the patterns $\boldsymbol{\xi}^k$ then embodies one instance of this concept (i.e., A, B, etc.). It seems natural that exchanging one of these instances for another should require qualitatively less effort than switching to an instance of another concept (e.g., one of a new set of memorized patterns representing the numbers 1, 2, etc.). The unique dynamics of Hebbian Kuramoto networks provide exactly this interchangeability of instances, because it takes an arbitrarily small external modification of the dynamics to transform the representation of one memorized pattern into the representation of another. For some applications, this could be an advantage over traditional networks (like the Hopfield model [9]), where memorized patterns are attractors and therefore switching patterns from the outside is a dissipative process.

As a final remark we would like to raise an interesting question. During our research on the topic at hand, we were not able to construct a network of symmetrically coupled Kuramoto oscillators with two or more separate attractive limit sets, regardless of whether a Hebb rule was used or not. There seems to be no literature proving or disproving the possibility. Certainty in this matter would establish whether neural networks of coupled Kuramoto oscillators with robust long term pattern recognition can be constructed at all.

Acknowledgment. Financial support through the cluster of excellence "Nanosystems Initiative Munich" (NIM) is gratefully acknowledged.

REFERENCES

[1] T. Aonishi, *Phase transitions of an oscillator neural network with a standard Hebb learning rule*, Phys. Rev. E (3), 58 (1998), pp. 4865–4871.

[2] T. Aonishi, K. Kurata, and M. Okada, *Statistical mechanics of an oscillator associative memory with scattered natural frequencies*, Phys. Rev. Lett., 82 (1999), pp. 2800–2803.

[3] T. Aoyagi and K. Kitano, *Effect of random synaptic dilution in oscillator neural networks*, Phys. Rev. E (3), 55 (1997), pp. 7424–7428.

[4] T. Aoyagi and K. Kitano, *Retrieval dynamics in oscillator neural networks*, Neural Comput., 10 (1998), pp. 1527–1546.

[5] A. Arenas and C. J. Pérez-Vicente, *Phase locking in a network of neural oscillators*, Europhys. Lett., 26 (1994), pp. 79–84.

[6] R. W. Hölzel and K. Krischer, *Pattern recognition with simple oscillating circuits*, New J. Phys., 13 (2011), 073031.

[7] R. W. Hölzel and K. Krischer, *Pattern recognition minimizes entropy production in a neural network of electrical oscillators*, Phys. Lett. A, 377 (2013), pp. 2766–2770.

[8] H. Hong, T. I. Um, Y. Shim, and M. Y. Choi, *Temporal association in a network of neuronal oscillators*, J. Phys. A, 34 (2001), pp. 5021–5032.

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

[10] F. C. Hoppensteadt and E. M. Izhikevich, *Oscillatory neurocomputers with dynamic connectivity*, Phys. Rev. Lett., 82 (1999), pp. 2983–2986.

[11] A. Knutson and T. Tao, *Honeycombs and sums of Hermitian matrices*, Notices Amer. Math. Soc., 48 (2001), pp. 175–186.

[12] K. Kostorz, R. W. Hölzel, and K. Krischer, *A novel type of a weakly coupled oscillatory network with associative properties*, New J. Phys., 15 (2013), 08310.

[13] Y. Kuramoto, *Self-entrainment of a population of coupled non-linear oscillators*, in International Symposium on Mathematical Problems in Theoretical Physics, Lecture Notes in Phys. 39, Springer, Berlin, 1975, pp. 420–422.

[14] T. Nishikawa, F. C. Hoppensteadt, and Y.-C. Lai, *Oscillatory associative memory with perfect retrieval*, Phys. D, 197 (2004), pp. 134–148.

[15] K. Park and M. Y. Choi, *Synchronization in a network of neuronal oscillators with finite storage capacity*, Phys. Rev. E (3), 52 (1995), pp. 2907–2911.

[16] C. J. Pérez-Vicente, A. Arenas, and L. L. Bonilla, *On the short-time dynamics of networks of Hebbian coupled oscillators*, J. Phys. A Mathematical, Nuclear and General, 29 (1996), pp. 9–16.

[17] M. Yamana, M. Shiino, and M. Yoshioka, *Oscillator neural network model with distributed native frequencies*, J. Phys. A, 32 (1999), pp. 3525–3534.

[18] M. Yoshioka and M. Shiino, *Associative memory storing an extensive number of patterns based on a network of oscillators with distributed natural frequencies in the presence of external white noise*, Phys. Rev. E (3), 61 (2000), pp. 4732–4744.