

The Time Dimension for Scene Analysis

DeLiang Wang, *Fellow, IEEE*

Abstract—A fundamental issue in neural computation is the binding problem, which refers to how sensory elements in a scene organize into perceived objects, or percepts. The issue of binding is hotly debated in recent years in neuroscience and related communities. Much of the debate, however, gives little attention to computational considerations. This review intends to elucidate the computational issues that bear directly on the binding issue. The review starts with two problems considered by Rosenblatt to be the most challenging to the development of perceptron theory more than 40 years ago, and argues that the main challenge is the figure-ground separation problem, which is intrinsically related to the binding problem. The theme of the review is that the time dimension is essential for systematically attacking Rosenblatt's challenge. The temporal correlation theory as well as its special form—oscillatory correlation theory—is discussed as an adequate representation theory to address the binding problem. Recent advances in understanding oscillatory dynamics are reviewed, and these advances have overcome key computational obstacles for the development of the oscillatory correlation theory. We survey a variety of studies that address the scene analysis problem. The results of these studies have substantially advanced the capability of neural networks for figure-ground separation. A number of issues regarding oscillatory correlation are considered and clarified. Finally, the time dimension is argued to be necessary for versatile computing.

Index Terms—Binding, desynchronization, figure-ground separation, oscillatory correlation, scene analysis, segmentation, synchronization, temporal correlation, time dimension, versatility.

I. ROSENBLATT'S CHALLENGE

IN his classic book more than 40 years ago, *Principles of Neurodynamics*, Rosenblatt [107] summarized in the last chapter a list of problems facing the study of perceptron theory at the time. Two problems among the list “represent the most baffling impediments to the advance of perceptron theory” [107, p. 580]. These are the problems of figure-ground separation and the recognition of topological relations. The development of the field of neural networks in the ensuing years has largely validated the foresight of Rosenblatt. In particular, major progress has been made in the understanding of error-correction procedures for training multilayer and recurrent perceptrons [4], [11], [81], [110]. The field of neural networks has been firmly established and has enjoyed tremendous success both as a field of theoretical study for brain function and as a technology for solving pattern recognition and related problems. On the other hand, progress has been extremely limited in addressing Rosenblatt's two chief problems in the framework of perceptrons.

Manuscript received March 23, 2004; revised November 18, 2004. This work was supported in part by the National Science Foundation under Grant IIS-0081058 and by the U.S. Air Force under AFOSR Grant F49620-01-1-0027.

The author is with the Department of Computer Science and Engineering and the Center for Cognitive Science, The Ohio State University, Columbus, OH 43210-1277 USA (e-mail: dwang@cse.ohio-state.edu).

Digital Object Identifier 10.1109/TNN.2005.852235

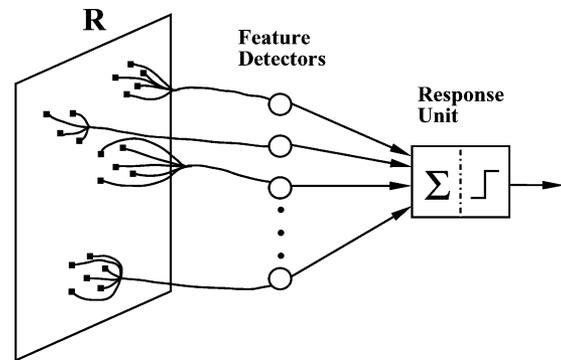


Fig. 1. Network architecture of a perceptron. The input layer is denoted by R . Each feature detector receives input from a specific area of R . The response unit computes the weighted sum of all the detectors and checks whether the sum exceeds a certain threshold.

The figure-ground separation problem concerns how to separate a figure from its background in a scene. Since natural scenes generally contain multiple objects, a closely related problem is the scene segmentation problem—how to segment a scene into its constituent objects. Literally speaking these two problems are not the same, but they are frequently treated as the same with different emphases: one on extracting target object and another on parsing the entire scene. As a result, figure-ground separation and scene segmentation are often used interchangeably; such is the case in Rosenblatt's book. The recognition of topological relations concerns how to compute various spatial relations between objects in a scene, e.g., whether object A is inside object B or whether B is to the left of object C . Solving a spatial recognition problem would require a solution to figure-ground separation first, and in this sense, the latter is a more basic problem. Both problems can be treated as major aspects of the scene analysis problem.

The deficiency of perceptrons for figure-ground separation was later hit hard in a landmark book by Minsky and Papert [85]. Through mathematical analysis, they pointed out that perceptrons are fundamentally limited in analyzing topological patterns. To better understand their mathematical results and perceptrons, let us give some details about perceptron theory. Perceptrons, introduced by Rosenblatt [106], [107], may be viewed as classification networks. Fig. 1 shows a typical perceptron that computes a predicate, which consists of a binary input layer R (symbolizing retina), a layer of binary feature-detecting units, and a response unit that represents the result of a binary classification. A feature detector senses a specific area of R , and it produces 1 if all the pixels of the area are black and 0 otherwise. The response unit is a logic threshold unit whose input is a weighted sum of all the units in the detector layer and whose output is 1 if the sum exceeds a threshold and 0 otherwise. Minsky and Papert defined the *order* of a predicate as the smallest number k

for which one can compute the predicate with feature detectors that sense no more than k pixels of R . A predicate is topologically invariant, or is a topological predicate, if it is unchanged when an input figure undergoes distortion that does not alter the connectedness or inside–outside relationship among the parts of the figure.

Minsky and Papert [85] proved that all except one topological predicate are of infinite order. That is, to compute such predicates requires feature detectors whose receptive fields are unbounded in size. One such predicate is connectedness: to tell whether an input pattern is connected or not. On a two-dimensional (2-D) grid, a binary pattern is defined as connected if any pixel of the pattern is connected to any other pixel through a path inside the pattern that traverses via only the up, down, left, and right links (i.e., four nearest neighborhood); otherwise the pattern is disconnected. The connectedness predicate serves as the cornerstone for their analysis. The order of this predicate increases at least as fast as $|R|^{1/2}$. What are the implications of this result? For any fixed R , their theorem is not about *whether* a perceptron exists to solve the problem. Indeed, with a finite, discrete R , there are a finite number of connected patterns, and one can trivially find a perceptron whose feature detectors recognize individual connected patterns and thus solve the problem. However, predicates that have unbounded order require large feature detectors relative to the size of R , and too many of them to be computationally feasible [85]. For example, on a 2×2 R , the number of connected patterns is 13, and on a 3×3 R , it is 222. The number of connected patterns grows exponentially except for one-dimensional (1-D) R ; see the Appendix for a proof. Clearly, this way of computing the connectedness predicate is computationally intractable for all but very small R s. In other words, their (negative) result is about the scalability or computational complexity of perceptrons.¹

The discovery of the backpropagation algorithm for training multilayer perceptrons [109]² led to a remarkable resurgence of interest in neural networks and has widely been heralded as having overcome the limitations of perceptrons revealed by Minsky and Papert. Its ability to train multilayer networks demonstrates clearly that the backpropagation algorithm is more powerful than the perceptron learning rule that is applicable to only simple perceptrons with one layer of trainable weights (Fig. 1 shows a simple perceptron since only the weights of the response unit are subject to training). However, although Minsky–Papert analysis is based on simple perceptrons, their conclusions expose a general problem in the perceptron framework for processing topological patterns, regardless of the learning rules employed. The problem, according to their 1988 expanded edition [86], “had nothing to do with learning at all; it had to do with the relationships between the perceptron’s architecture and the characters of the problems presented to it” [86, p. xiii]. They further stated that it is a representation problem, and “no machine can learn to recognize X unless it possesses, at least potentially, some scheme for representing X ” [86, p. xiii].

¹Their book is often credited for the downfall of neural networks in the late 1960s.

²This is the standard reference for the backpropagation algorithm, though earlier versions had been discovered (see [149]).

Can a multilayer perceptron with backpropagation training recognize topological patterns? A direct attempt to address figure-ground separation was made by Sejnowski and Hinton [113] based on a related learning algorithm—Boltzmann machine. They demonstrated its effectiveness on toy problems, but “its usefulness for large problems is still uncertain” [113, p. 703]. A later study using an extended version of the backpropagation algorithm to perform a segmentation task was made by Mozer *et al.* [90], but again, success on only toy problems was reported. Of course, analytical statements cannot be derived from individual attempts, but as far as the connectedness predicate is concerned, Minsky and Papert [86] explicitly claimed that multilayer networks are no more powerful [86, p. 52]. We find no report contrary to this claim since their 1988 edition. Further insight can be obtained by observing the architecture and training of a multilayer perceptron. The notion of an order may appear irrelevant for a multilayer perceptron given that hidden units typically receive input from all the input units. At issue, however, are how many hidden units are required and how long is the training process. Because of the infinite order of the predicate and the exponential increase of the number of connected patterns, for R not too small, the number of hidden units required to represent a solution would be prohibitively large. Also, a practical training process can use only a tiny fraction of all possible training samples, and one should not expect example-based learning procedures, such as backpropagation, to generalize to exponentially more unseen samples. Hence, the Minsky and Papert claim is a projection with strong justification, and it is at least premature to regard that the limitations of simple perceptrons no longer exist for their multilayer offsprings.

To solve the figure-ground separation problem in the general form, the solution must be valid regardless of the shape, position, size, orientation, etc., of each object on the retina; that is, the solution must be invariant to topological transformations. In a sense, it presupposes a solution to the connectedness problem. As we shall see later, the ability to segment connected components of an image plus some counting mechanism yields a solution to detect the connectedness predicate. So, the limitations of perceptrons in dealing with topological patterns can be treated as a special case of Rosenblatt’s challenge. In other words, besides the mathematical rigor of Minsky–Papert analysis, the major findings of the book should have been hardly surprising to Rosenblatt.

Modern neurocomputing research has made important advances in understanding generalization characteristics of multilayer networks and their statistical underpinnings, which in turn have led to more effective training algorithms and better generalization capability for certain classes of problems [4], [11]. However, mainstream neural networks are preoccupied with analyzing statistical properties of input patterns, and have all but shunned away from Rosenblatt’s challenge. But the challenge remains a considerable cloud. For example, Hinton and Sejnowski [53] acknowledged that “a major challenge for unsupervised learning is to get a system of this general type to learn appropriate representations for images” [53, p. xiv].

This paper argues that introducing the dimension of time is essential for a systematic attack on Rosenblatt’s challenge. I

will first state in Section II that Rosenblatt’s challenge is intrinsically related to the binding problem, and a key representation to resolving the binding problem lies in the temporal correlation theory. In Section III, I introduce a special form of the temporal correlation theory: oscillatory correlation, which is primarily motivated by computational/mathematical and biological considerations, and lay out the computational requirements for the representation theory. Section IV reviews several dynamical mechanisms that have successfully met the computational requirements, and as a specific example the section also describes a solution to the connectedness problem. In Section V, I review the progress on applying oscillator networks to solving a variety of scene analysis tasks including segmentation, and these results demonstrate the computational utility of the oscillatory correlation theory. Section VI briefly discusses the biological relevance of the theory. Section VII discusses and clarifies a number of issues, such as readout and internal time versus external time (or oscillator time versus physical time). Concluding remarks are given in Section VIII, where I argue that versatile neural computation requires the time dimension.

A large number of papers have been published on neural dynamics and temporal coding, and this paper reviews only computational developments that bear directly on the binding problem, rather than attempting an exhaustive literature survey. As a result, I acknowledge at the outset that many otherwise interesting papers are not included in this review.

II. BINDING PROBLEM

A fundamental attribute of perception is the ability to group elements of a perceived scene into coherent objects, which has been extensively studied in perceptual psychology under the title of perceptual organization or perceptual grouping [98]. This ability is remarkable considering the fact that the input to visual organization is a retinal image and the organization takes place rapidly and effortlessly. Despite centuries of research from multiple disciplines, how perceptual organization is accomplished in the brain remains a mystery. Observing that visual features, such as color, orientation, motion, and depth, are first extracted by feature-detecting neurons in different areas of the visual system, a related question is how these initial responses are bound together in the brain to form perceived objects, or percepts? This is called the binding problem. At the heart of the problem is the fact that the sensory input contains multiple objects, which makes it necessary to resolve the issue of which features should bind with which others. A related formulation is often illustrated using two objects of different shapes, say, a triangle and a square, and different locations, say, top and bottom, arranged in such a way that the triangle is at the top and the square is at the bottom (a layout discussed by Rosenblatt [107];³ see [136]). This layout is shown in Fig. 2. Given feature detectors for triangle, square, top, and bottom, how can the perceptual system bind locations and shapes in order to correctly perceive that the triangle is at the top (binding “top” and “triangle”) and the square is at the bottom (binding “bottom” and “square”), rather than the other

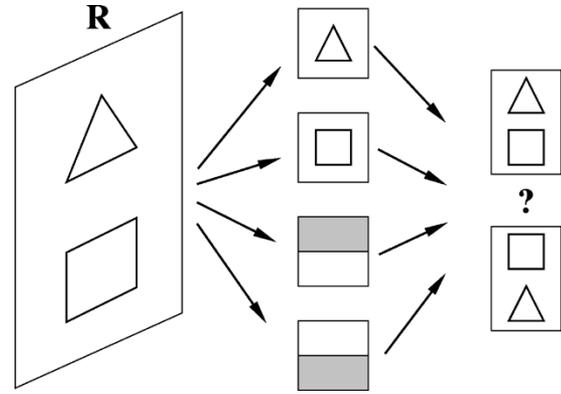


Fig. 2. Binding problem in a perceptron with four detectors for triangle, square, top, and bottom. The response unit needs to tell whether the triangle is on top (and the square at bottom) or the square is on top (and the triangle at bottom).

way around: the square is at the top and the triangle is at the bottom? What makes this formulation attractive is that human observers can make wrong feature conjunctions, producing “illusory conjunctions,” when stimuli are briefly presented in a way to prevent selective attention to individual stimuli [126], [127]. A potential confusion for this popular formulation (see [108]) is that object-level attributes such as shape and size are undefined before percepts are computed; that is, they are not defined before the more fundamental problem of figure-ground separation is solved.⁴ For this reason, I will refer to the binding of local features to form objects—a task of figure-ground separation—when talking about the binding problem, while realizing that there are multiple levels of the binding problem.

How does the brain resolve the binding problem? Concerned with the difficulty for visual shape recognition that is created by multiple, simultaneously present objects, Milner [84] suggested that different objects be separated in time, leading to synchronization of firing activity within the cells activated by the same object. Also concerned with object recognition, Uttal [132] proposed an autocorrelation model that emphasizes spatial correlation of feature detectors. Later, in one of most cited technical reports, von der Malsburg [135] argued that the time structure of neural signals provides the neural basis for his correlation theory, which directly and systematically addresses the binding (integration) problem. In a subsequent study, von der Malsburg and Schneider [137] gave a concrete demonstration of the temporal correlation theory for the task of segregating two auditory inputs based on their distinct onset times—an example of the cocktail-party problem (I will come back to this problem in Section V.) This paper marks an important turning point because it, for the first time, addresses a figure-ground separation task using neural oscillators as building blocks, and synchrony and desynchrony among neural oscillations to represent a solution to a binding problem.

I should note that the temporal correlation theory is a theory of representation, concerned with how neurons triggered by different objects are represented in a neural network, not a theory of computation. It does not by itself address *how* multiple patterns

³Note that Rosenblatt himself did not formulate this as a binding problem.

⁴Treisman in formulating her feature integration theory [125] also failed to make this distinction.

lead to multiple cell assemblies with different time structures. This is the key issue to be taken up in the next section.

The main alternative to the temporal correlation theory is hierarchical coding. The central idea is to rely on individual neurons arranged in some cortical hierarchy to integrate information so that cells higher in the hierarchy respond to larger and more specialized parts of an object. Eventually, this leads to the scenario that individual objects are represented by individual neurons, and for this reason hierarchical coding is also known as the grandmother or cardinal cell representation [7].⁵

In the following, I will limit my discussion to computational considerations; see Gray [49] for biological evidence for and against the hierarchical representation. There are several computational problems with this representation. First, to be able to bind features that belong to the same object, object representation must already exist in the brain and image analysis is thus limited to recognition of familiar objects in the image. In addition to the issue of how individual representations are created in the first place, hierarchical coding would not allow perceiving novel objects, an ability the perceptual system undoubtedly possesses [126]. Second, perceiving an object, with all its vivid details such as location, shape, color distribution, orientation, and size, is different from simply identifying, say, that the object is a dog [60], [126]. This creates the following dilemma. If the representation explicitly encodes and prestores all such details in order to deal with all possible scenarios, a vast majority of which would never occur to an observer in a lifetime, it would require prohibitively many cells. If the representation stays above these details, the binding problem recurs when facing an image with multiple objects: how to make sure that only a subset of image elements that belong to the object are fed to a recognizer? One possible way out of this dilemma is perform a top-down search starting from stored templates, one by one. This could be computationally feasible if the number of patterns and their possible variations is limited, but it would severely limit the scope of image analysis.

The binding problem in the visual domain is a subject of intense debate that has captured the interest of researchers from many disciplines. For example, the journal *Neuron* published a special issue in 1999 discussing the binding problem. It featured articles by leading researchers both for and against the two binding hypotheses discussed above; see also Robertson [105].

The above discussions should make it clear that, from the standpoint of the temporal correlation theory, the figure-ground separation problem is basically the same as the binding problem and the theory is primarily motivated by the need to address the problem. On the other hand, we note that the hierarchical coding mechanism is not much different from the perceptron framework as discussed by Rosenblatt [107], where a variety of architectures, including multilayer and recurrent ones, are studied. Hence, the challenge facing Rosenblatt, discussed in the previous section, also underscores the limitations of hierarchical coding.

⁵The term "grandmother cell" dramatizes the claim implied by hierarchical coding that cells must exist to code all individual objects including, hence, one's grandmother.

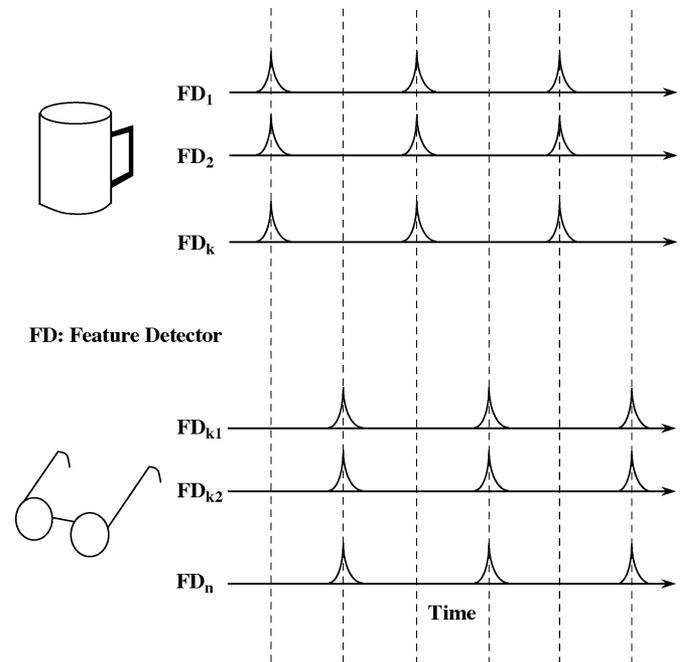


Fig. 3. Oscillatory correlation representation. Each object (a cup or a pair of glasses) is represented by an assembly of feature detectors that are oscillators. Oscillators within an assembly synchronize their activity and different assemblies desynchronize.

III. OSCILLATORY CORRELATION THEORY AND COMPUTATIONAL REQUIREMENTS

To make progress one must study concrete cases. We focus on a special form of temporal correlation, called *oscillatory correlation* [124], whereby feature detectors are represented by oscillators and binding is represented by synchrony within an assembly of oscillators and desynchrony between different assemblies. This representation is illustrated in Fig. 3. The oscillatory correlation proposal is motivated by three considerations. First, the activity of an oscillator describes well that of a neuron or a local group of neurons. Second, oscillatory correlation is consistent with coherent oscillations in the brain. Third, the use of oscillators facilitates a good deal the computation of synchrony and desynchrony. Two remarks about the oscillatory correlation theory are in order. First, temporal correlation need not be oscillatory correlation and there can be other ways to encode temporal structure. In practice, however, few other alternatives have been seriously pursued. Second, oscillators need not always produce periodic activity—they do so only when they reach steady behavior.⁶ Thus, if the external input to an oscillator changes with time or the oscillator response is within a transient period, the oscillator activity may not exhibit periodic behavior although it is still an oscillator mathematically.

As I noted earlier, the oscillatory correlation theory is a representation theory, not a computational mechanism. How to compute required synchrony and desynchrony when facing an input scene is an issue largely separate from the representation, and it is obviously a critical one in order to solve the binding problem.

⁶This description of oscillations excludes so-called strange oscillators, which are really chaotic units.

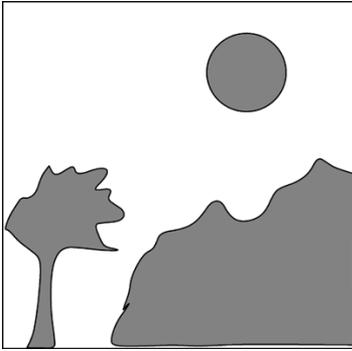


Fig. 4. A visual scene with three caricature objects.

The fact that the correlation theory is not a computational mechanism for addressing the binding problem has been repeatedly criticized by its opponents [48], [114]. However, the theory has already motivated major progress in addressing this very issue, which is the topic of this review.

The discovery of coherent oscillations in the cat visual cortex in the late 1980s [38], [50] immediately triggered a lot of computational research aimed at either modeling the biological phenomena or attacking the binding problem. Most of the early models use long-range connections to achieve synchronization in an assembly of oscillators. However, long-range connections lead to the problem of indiscriminant segmentation: oscillators would synchronize no matter whether they are activated by the same object or different ones [121], [138]. This is illustrated in Fig. 4, where three objects comprise an input scene. It is immediately clear from the figure that the three objects are separated on the basis of connectedness, but globally connected networks cannot encode topology and thus fail to separate the three patterns in the figure. To accomplish even this elementary task, one needs to use locally connected networks.

In general, to provide a computational mechanism for the oscillatory correlation theory, three key functions must be achieved: 1) the mechanism must be capable of synchronizing a locally coupled assembly of oscillators; 2) it must be capable of desynchronizing different assemblies of oscillators that are activated by different objects; and 3) both synchrony and desynchrony must occur rapidly. The last function is dictated by the fact that perceptual organization and object recognition take place efficiently in order to deal with the external environment that can change rapidly [10]; also experimental evidence shows that neural synchrony occurs quickly. These requirements became major stumbling blocks for many computational attempts subsequent to the experimental discovery of synchronous oscillations. A major reason is that harmonic oscillators cannot synchronize with local connections (see [124] for more discussion), even though such oscillators, due to their simplicity, were widely used for achieving oscillatory correlation. This also explains why such models require all-to-all connectivity to achieve synchrony. Fortunately, different types of oscillators yield qualitatively different behaviors at the network level. The next section reviews several mechanisms that directly address the computational requirements.

IV. DYNAMICAL FOUNDATIONS

In this section, I survey several prominent types of oscillator model and networks of such oscillators, which successfully address the computational requirements discussed in the previous section. Specifically, I discuss Wilson–Cowan, relaxation, and spike oscillators. Toward the end of the section, I also make an attempt to relate different types of oscillator model.

A. Wilson–Cowan Oscillators

The oscillator model that underlays the first successful computational attempts is that proposed by Wilson and Cowan [152]. The model was originally proposed to understand dynamical properties of two interacting neuron populations: an excitatory subpopulation whose activity is described by x and an inhibitory subpopulation whose activity is described by y . A single oscillator i is defined as

$$\dot{x}_i = -x_i + S_x(ax_i - by_i + I_i) \quad (1a)$$

$$\dot{y}_i = -y_i + S_y(cx_i - dy_i) \quad (1b)$$

where a, b, c, d are parameters. I_i denotes external stimulation to the oscillator and

$$S_r(u) = \frac{1}{1 + \exp[-\lambda(u - \theta_r)]}, \quad r \in \{x, y\} \quad (2)$$

is a sigmoidal nonlinearity with parameters λ, θ_x , and θ_y . The original equations [152] have more terms and parameters, including multiplicative terms before the sigmoids and different time constants for (1a) and (1b). As in the following, our description does not aim at mathematical generality, which can be achieved, for example, by substituting parameters for constants; rather, it aims at elucidating basic properties. Fig. 5(a) illustrates the interactions between x and y .

Both the x -nullcline (i.e., $\dot{x} = 0$) and the y -nullcline are sigmoidal functions, as illustrated in Fig. 5(b). In certain parameter ranges, the two nullclines intersect at a single, unstable fixed point and the system produces a stable limit cycle, or oscillation [see Fig. 5(b)]. A typical trace of oscillation is shown in Fig. 5(c), where the x activity is plotted.

König and Schillen [68] described the first network of Wilson–Cowan oscillators that reaches synchrony with only local coupling. They introduced time delays for connections between the x unit and the y unit of a single oscillator [see Fig. 5(a)]. In addition, they introduced time delays into local connections among different oscillators. Specifically, the x unit of oscillator i connects to the y unit of i 's neighboring oscillators, and these connections have the time delay close to that of the connection from x to y within oscillator i . With such connections, one would expect that phase synchrony is a steady behavior of a locally coupled network, and their simulations suggest that this is indeed the case. In a subsequent paper [112], they also studied the problem of desynchronization by extending their network to include longer range connections among the x units of coupled oscillators. Fig. 6 shows the architecture of their overall model that includes both synchronizing connections between nearest neighbors and desynchronizing connections between second nearest neighbors. Again, these

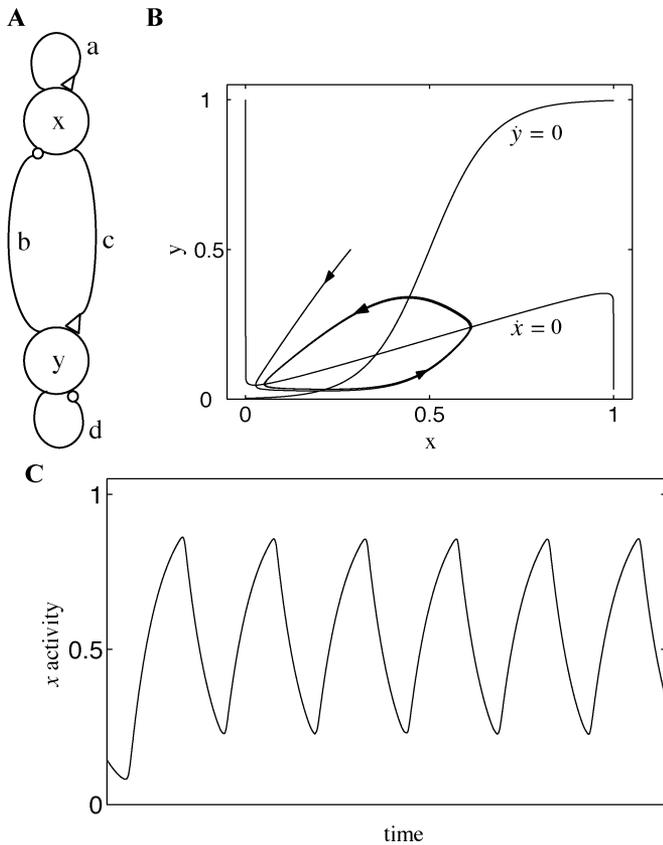


Fig. 5. Single Wilson-Cowan oscillator. (a) Diagram of the oscillator, composed of an interacting pair of an excitatory and an inhibitory unit. Small triangles indicate excitatory connections and small circles inhibitory connections. (b) Nullclines and oscillatory trajectory as shown in the phase plane. The arrows indicate the direction of the trajectory. (c) The x activity of the oscillator with respect to time.

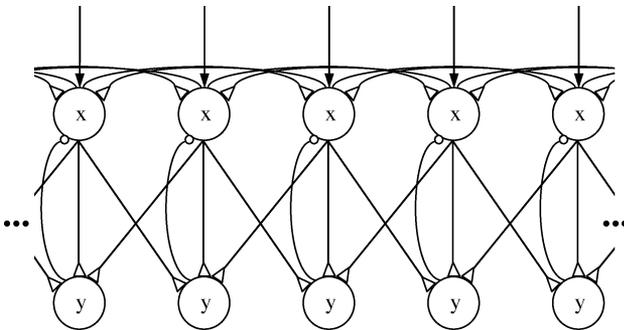


Fig. 6. Architecture of the Schillen and König model (adapted from [112]). Delay coupling between adjacent oscillators serves as synchronizing connections, and delay coupling between second nearest neighbors serves the purpose of desynchronization. As in Fig. 5, small triangles indicate excitatory connections and small circles inhibitory connections.

desynchronizing connections need to have proper time delays in relation to the intrinsic delays within a single oscillator. Their simulation results show that, under such conditions, two assemblies of Wilson-Cowan oscillators can be desynchronized. It is doubtful, however, that more than two assemblies can be separated this way.

Although time delays are naturally present in synaptic transmission and nerve conduction, reliance on delicate time

delays to achieve synchrony and desynchrony tends to limit the flexibility and robustness of the model. Wang proposed a different network of locally coupled Wilson-Cowan oscillators that achieves synchronization without using time delays [139]. In this network, oscillators are locally coupled through excitatory connections between the x units. In addition, the paper introduced a dynamic normalization scheme that quickly modifies the dynamic weights between neighboring oscillators so that the dynamical weights of all oscillators are equalized. The concept of fast-changing connection weights (or dynamic links) was first introduced by von der Malsburg [135]. With the aid of dynamic weight normalization, Wang demonstrated that a locally coupled network reaches robust synchronization. Subsequently, Campbell and Wang constructed a network where local connections exist between both the excitatory units and the inhibitory units [17]. Using a piecewise linear approximation of the sigmoid function in (1), they proved that for a chain of Wilson-Cowan oscillators with nearest neighbor coupling, the synchronous solution is asymptotically stable. A further analysis of the nullclines and the coupling strengths shows that fast synchronization can be achieved by carefully choosing system parameters. Moreover, they demonstrated that the introduction of a global unit, called the global separator, leads to desynchronization among synchronized oscillator assemblies. Their simulations show that up to five assemblies can be clearly separated in phase space.

B. Relaxation Oscillators

Unlike other oscillator models, a relaxation oscillator exhibits two time scales. Relaxation oscillation was first studied by van der Pol [133] to model properties of a triode circuit, and has since been widely studied in mechanics, biology, chemistry, and engineering.

A relaxation oscillator i is typically defined as a reciprocally connected pair of excitatory unit x_i and inhibitory unit y_i [124]

$$\dot{x}_i = 3x_i - x_i^3 + 2 - y_i + I_i \quad (3a)$$

$$\dot{y}_i = \varepsilon \left(\alpha \left(1 + \tanh \left(\frac{x_i}{\beta} \right) \right) - y_i \right). \quad (3b)$$

Here, I_i denotes external stimulation to the oscillator and α and β are parameters. The parameter ε is a small positive number, and it induces a slow time scale for the evolution of y . The x -nullcline is a cubic function and the y -nullcline is a sigmoid function. If $I > 0$, the two nullclines intersect only at a point along the middle branch of the cubic, and in this case the oscillator produces a stable limit cycle, illustrated in Fig. 7(a). The oscillator is called *enabled*. The limit cycle alternates between a phase of relatively high x values and a phase of relatively low x values, called the *active* and *silent* phase, respectively. Within each of the two phases, the oscillator exhibits near steady-state behavior, and its trajectory in the silent phase corresponds to the left branch (LB) of the cubic and in the active phase the right branch (RB). In contrast, the transition between the two phases takes place rapidly, and it is referred to as jumping. α determines the relative durations during which the limit cycle spends in the two phases—a larger α produces a relatively shorter active phase. If $I < 0$, the two nullclines intersect at a stable fixed

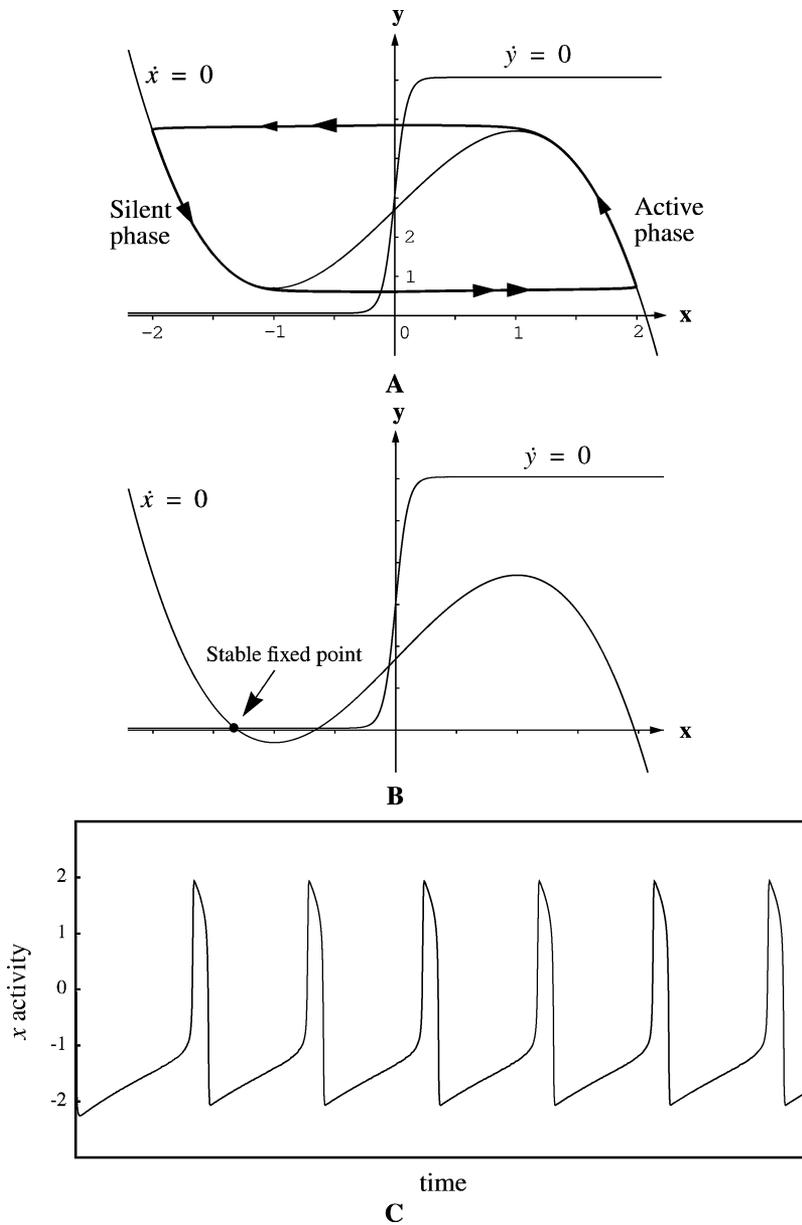


Fig. 7. Single relaxation oscillator. (a) The oscillator is enabled. In this case it produces a limit cycle, shown as the bold curve. The arrows indicate the direction of trajectory, and double arrows indicate jumping. (b) The oscillator is excitable. In this case, it approaches the stable fixed point. (c) The x activity of the oscillator with respect to time.

point on LB of the cubic [see Fig. 7(b)]. In this case no oscillation occurs, and the oscillator is called *excitable*. Whether an oscillator is enabled or excitable depends on external stimulation. Hence, oscillations in (3) are stimulus-dependent.

The relaxation oscillator may be interpreted either as a model of action potential generation, where x represents the membrane potential of a neuron and y represents the level of activation of ion channels, or oscillating bursts of neuronal spikes, where x represents the envelope of the bursts. Fig. 7(c) shows a typical trace of the x activity, akin to a spike train. In fact, standard neuronal models, including the FitzHugh–Nagumo model [42], [91] and Morris–Lecar model [89], are of the relaxation type.

When analyzing synchronization properties of a pair of relaxation oscillators, Somers and Kopell [119] found the fast threshold modulation mechanism. Subsequently, Terman and Wang [124] found the selective gating mechanism for desyn-

chronizing oscillator assemblies. Due to their importance for subsequent developments, these two mechanisms are explained in some detail below.

To explain the fast threshold modulation mechanism, consider a pair of identical relaxation oscillators coupled through excitatory connections between their x units. Furthermore, assume that the excitatory coupling can be viewed as binary (or Heaviside), resembling chemical synapses; that is, the two oscillators are uncoupled unless one of them is in the active phase. When one oscillator is in the active phase, the effect of its excitatory coupling on the other is to raise the latter's cubic by an amount equal to the connection strength. Specifically, denote the two oscillators by $o_1 = (x_1, y_1)$ and $o_2 = (x_2, y_2)$, as defined in (3). Let both start in the silent phase, not far apart. This is shown in Fig. 8(a) with o_1 leading the way. The figure illustrates the solution of the oscillator system in the singular limit $\varepsilon \rightarrow 0$.

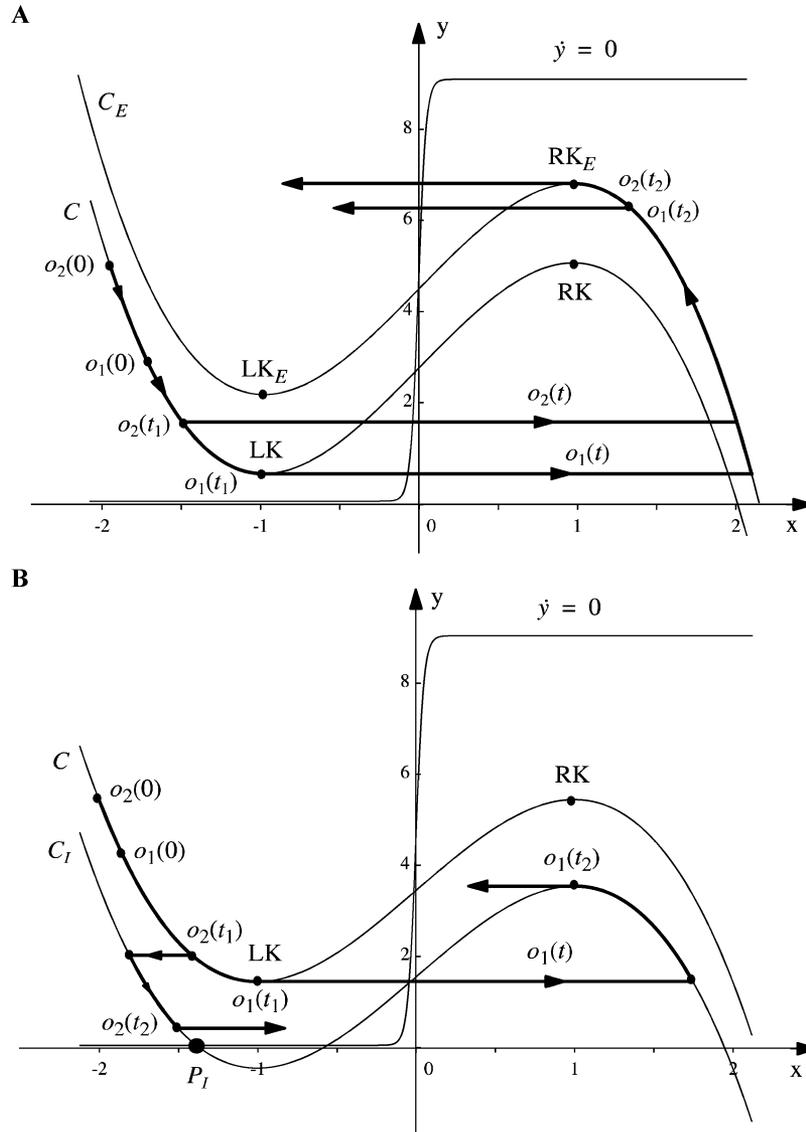


Fig. 8. Analysis of coupled relaxation oscillators. (a) Fast threshold modulation. C and C_E indicate the uncoupled and the excited cubic, respectively. LK and RK indicate the left knee and the right knee of C , and LK_E and RK_E the left knee and right knee of C_E , respectively. (b) Selective gating. C and C_I indicate the uncoupled and the inhibited cubic, respectively. P_I denotes a stable fixed point at an intersection point between C_I and the y -nullcline.

Designate the cubic corresponding to the uncoupled case as C . The solution consists of several pieces. The first piece is when both oscillators move along LB of C . This piece ends when o_1 reaches the left knee of C , LK, at $t = t_1$. The second piece starts when o_1 jumps to RB. Now, o_1 excites o_2 and raises its cubic from C to C_E [see Fig. 8(a)]. Denote the left and right knee of C_E by LK_E and RK_E , respectively. If $|y_1 - y_2|$ is relatively small compared to the coupling strength, o_2 lies below LK_E and jumps up as a result of the excitatory coupling. Since these interactions occur on the fast time scale, o_1 and o_2 are in effect synchronized in jumping up. Once o_2 jumps up, o_1 's cubic is also raised to C_E . A key observation is that, even though the jumping does not alter the vertical distance between the two oscillators [see Fig. 8(a)], their separation in time (i.e., phase) is substantially reduced because of the different speeds of motion along the y axis before and after the jumping. Such a speed is determined by an oscillator's vertical distance to the y nullcline, and it is slower before the jumping. The third piece starts when

both oscillators lie on RB. They will then evolve in slow time. Note that, now, o_2 leads the way, reversing the order in the first piece. The third piece lasts until o_2 reaches RK_E at $t = t_2$. The fourth piece starts when o_2 jumps to LB. With o_2 on LB, o_1 's cubic is lowered to C and o_1 jumps down as well. Once both oscillators are in the silent phase, the above process repeats.

Somers and Kopell proved a theorem that the synchronous solution in the oscillator pair has a domain of attraction in which the approach to synchrony has a geometric (or exponential) rate [119]. Their numerical simulations suggest that fast synchronization appear also in 1-D networks. Subsequently Terman and Wang [124] proved that the Somers–Kopell theorem for a pair of oscillators extends to an arbitrary network of locally coupled relaxation oscillators. Their proof clarifies the following relationship between a Euclidean metric and a time metric. When oscillators evolve either in the silent phase or the active phase, their distances in the Euclidean metric along the y axis change but their distances in the time metric remain constant. On the

other hand, when oscillators jump together, their y distances remain unchanged while their time distances change (compress).

In addition to extending the fast threshold modulation mechanism to a locally coupled network, Terman and Wang [124] proposed and analyzed LEGION⁷ architecture for achieving both synchrony and desynchrony. For a simplest 2-D LEGION network, an oscillator is excitatorily coupled with its four nearest neighbors, and a global inhibitor receives input from all the oscillators and then inhibits them. Specifically, an additive coupling term S_i is introduced into the right-hand side of (3a), given by

$$S_i = \sum_{k \in N(i)} W_{ik} H(x_k - \theta_x) - W_z H(z - \theta_z) \quad (4)$$

where H stands for the Heaviside step function, W_{ik} is the connection weight from oscillator k to i , and $N(i)$ is the set of four neighbors of i . Both θ_x and θ_z are thresholds, and θ_x is chosen between LB and RB along the x axis. W_z in (4) is the weight of inhibition from the global inhibitor z , defined as

$$\dot{z} = \phi(\sigma_\infty - z). \quad (5)$$

Here, ϕ is a parameter, and $\sigma_\infty = 1$ if $x \geq \theta_x$ for at least one oscillator and $\sigma_\infty = 0$ otherwise. If σ_∞ equals one, $z \rightarrow 1$.

To illustrate how the selective gating mechanism desynchronizes oscillator assemblies on a LEGION network, we consider an example with two oscillators that are coupled only through the global inhibitor. Each oscillator is meant to correspond to a synchronized assembly. Let both oscillators start in the silent phase and not far apart with $y_1 < y_2$. Fig. 8(b) shows the situation. Similar to Fig. 8(a), the singular solution of the system has several pieces. The first piece lasts until o_1 reaches LK at $t = t_1$. Since o_1 and o_2 are on LB during the first piece, $z = 0$. The second piece starts when o_1 jumps up. When o_1 crosses θ_z during the jumping, σ_∞ switches from zero to one and z approaches one on the fast time scale. When z crosses θ_z , both o_1 's and o_2 's cubic are lowered from C to C_I , the inhibited cubic. One can readily choose the system parameters so that C_I intersects with the sigmoid at a stable fixed point P_I along LB; this is shown in Fig. 8(b). The third piece corresponds to the duration when o_1 is in the active phase, while o_2 is in the silent phase. During this piece o_2 approaches P_I , and it cannot jump up as long as o_1 is on RB. The third piece lasts until o_1 reaches the right knee of C_I at $t = t_2$. The fourth piece starts when o_1 jumps to LB. When o_1 crosses θ_z during the process of jumping down, $z \rightarrow 0$. When z crosses θ_z , both o_1 's and o_2 's cubic return to C . There are now two cases to consider. In the first case, o_2 lies below LK, as shown in Fig. 8(b), and it jumps up immediately. In the second case, both o_1 and o_2 lie on LB, with o_2 leading the way. For both situations, the same analysis repeats. This analysis shows that global inhibition causes desynchronization of the two oscillators: o_1 and o_2 are never in the active phase simultaneously.

In general, the selective gating mechanism works as follows. When an oscillator, say, o_i , jumps to the active phase, it quickly activates the global inhibitor, which prevents the oscillators of

different assemblies from jumping up but does not affect o_i 's ability to recruit the oscillators of the same assembly thanks to local excitation. Combining selective gating and fast threshold modulation, Terman and Wang [124] proved the following theorem: a LEGION network achieves both synchronization within each oscillator assembly and desynchronization between different assemblies in no greater than m cycles of oscillations, where m is the number of the assemblies. In other words, both synchronization and desynchronization are achieved rapidly.

As a concrete application of the dynamics described above, I discuss a general solution to the connectedness problem [142], which, as explained in Section I, is a classic problem in neural computation. Before explaining how to compute the predicate, I show the response of a two-dimensional LEGION network to two binary images: one connected and one disconnected. The connected image is a cup figure shown in Fig. 9(a), while the disconnected one is the word "CUP" shown in Fig. 9(d). To ensure that the network has no binding preference at the beginning, the phase of each oscillator is randomized, as illustrated in Fig. 9(b). In the figure, the diameter of a circle corresponds to the x activity of the respective oscillator. Fig. 9(c) shows a snapshot of network activity shortly after the beginning. All the oscillators corresponding to the cup are synchronized, while the remaining ones are excitable. Fig. 9(e)–(g) shows the network response to "CUP." Fig. 9(h) depicts the temporal activity of all the stimulated oscillators for the connected cup image, where unstimulated oscillators are omitted since they do not oscillate. The oscillator activity corresponding to each connected pattern is combined in the display. The upper panel of Fig. 9(h) shows the activity of the assembly representing the cup, and the middle one shows that of the global inhibitor. Synchrony occurs in the first oscillation period. The situation for the disconnected "CUP" is depicted in Fig. 9(i), where the upper three traces show the assembly activities corresponding to the three patterns. Fig. 9(i) shows that synchrony within each assembly and desynchrony between different ones are both achieved in the first two periods.

As illustrated in the above simulations, when an assembly jumps to the active phase, the global inhibitor is triggered, and this happens as many times within an oscillation period as the number of patterns in the input image. Thus, how many patterns are in the input image can be revealed by the ratio of the frequency of the global inhibitor to the oscillation frequency of an enabled oscillator [142]. A ratio of one indicates that the input image contains one pattern, and thus the figure is connected. A ratio greater than one indicates that the input image contains more than one pattern and therefore the figure is disconnected. The connectedness predicate is then given by whether the ratio exceeds a threshold θ , chosen in the range $2 > \theta > 1$. The ratios for Fig. 9(h) and (i) are given in the bottom traces, where $\theta = 1.6$ is used. The figure shows that, beyond a short beginning duration that corresponds to the synchronization and desynchronization process, the connectedness predicate is correctly computed.

C. Spike Oscillators

The fact that a neuron generates a train of action potentials—the fundamental signal communicated between neurons—has motivated considerable investigation that explicitly

⁷LEGION stands for locally excitatory globally inhibitory oscillator networks [147].

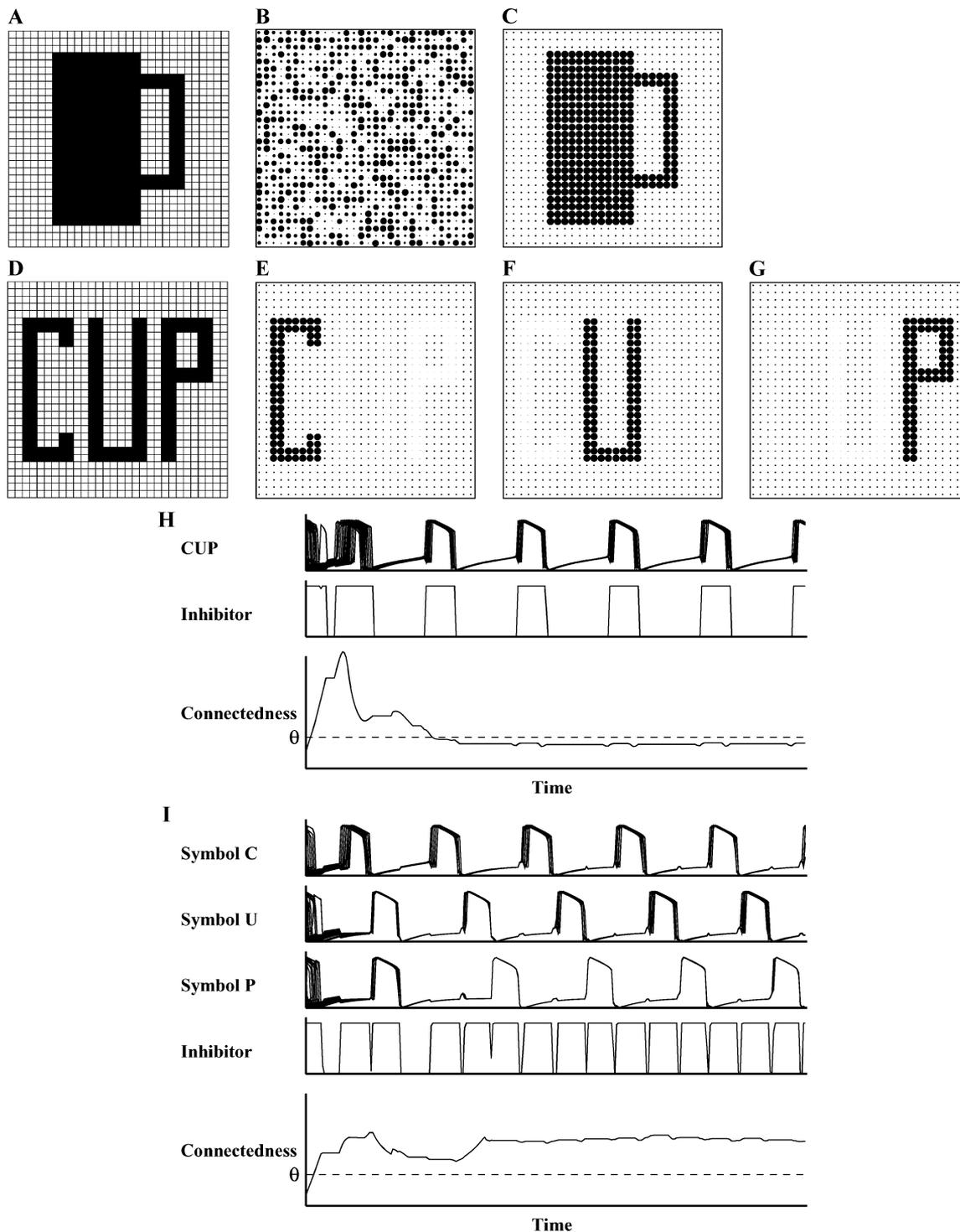


Fig. 9. Solution to the connectedness problem (from [142]). (a) A connected cup image presented to a 30×30 LEGION network. (b) A snapshot at the beginning of system evolution. (c) A subsequent snapshot taken shortly afterwards. (d) A disconnected image with three patterns forming the word “CUP” presented to the same network. (e)–(g) Subsequent snapshots taken shortly after the system starts. (h) Temporal activity of every enabled oscillator for the connected cup image. (i) Temporal activity for the disconnected “CUP” image. In both (h) and (i), the upper traces show the combined x activities of the oscillator assemblies indicated by their respective labels, the next-to-bottom trace the activity of the global inhibitor, and the bottom one the ratio of the frequency of the global inhibitor to that of an enabled oscillator together with θ .

models action potentials. An action potential corresponds to a short pulse of the membrane potential, which can be modeled by a spike oscillator. Many neuronal models produce a train of spikes, including the classic Hodgkin–Huxley model [54], the integrate-and-fire model [87], [101], and the model of Eckhorn

et al. [37]. Since the outputs of these neuronal models are phenomenologically very similar, our following discussion is based on the integrate-and-fire oscillator model, which is simple and has intrinsic connections with other, more complex models (see [47] and [58]).

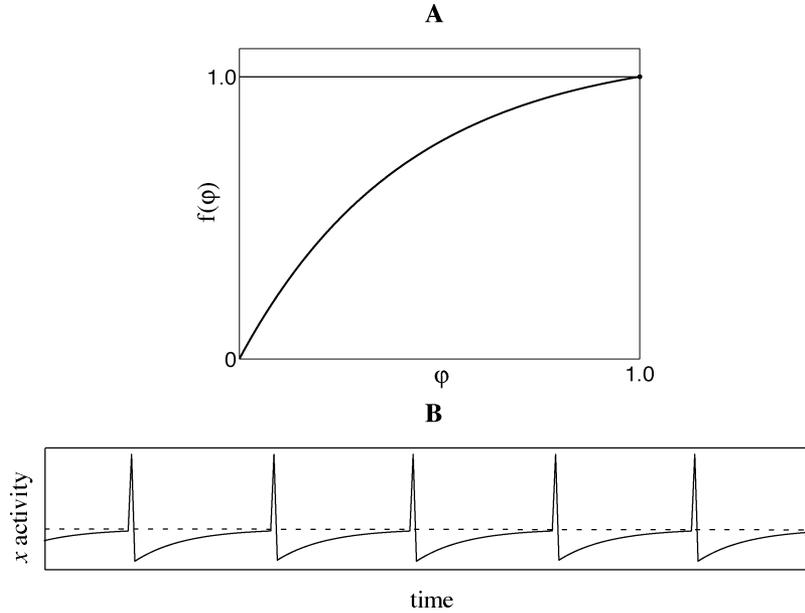


Fig. 10. Single spike oscillator. (a) The solution of the oscillator variable, x , plotted with respect to its phase, φ . The thin solid line indicates the firing threshold. (b) The x activity of the oscillator with respect to time. The firing threshold is indicated by the dash line, and the spikes show the times when the oscillator reaches the threshold.

A spike oscillator i can be described by one variable corresponding to the membrane potential

$$\dot{x}_i = -x_i + I_i \quad (6)$$

where I_i indicates external input, which controls the oscillation period. When the membrane potential reaches a threshold (set to one) at time t , the oscillator fires or produces a spike and its potential is instantly set to zero

$$x_i(t^+) = x_i(t) - 1. \quad (7)$$

Note that, unlike a Wilson–Cowan or relaxation oscillator, the generation of a spike does not arise from the dynamics; rather it results from an algorithmic description. The conductance-based Hodgkin–Huxley model does not need this algorithmic component.

The above oscillator equations can be solved analytically [87]. Indicate the phase of the oscillator by φ , $\varphi \in [0, 1]$, and we have $x(\varphi) = f(\varphi) = I[1 - \exp(-\tau\varphi)]$, where $\tau = \log[I/(I-1)]$ is the period of the oscillation. The function $f(\varphi)$ increases monotonically (i.e., $f'(\varphi) > 0$) and is concave down (i.e., $f''(\varphi) < 0$). Fig. 10(a) shows $f(\varphi)$ and Fig. 10(b) shows an oscillation trace.

To define a network of spike oscillators, an additive coupling term S_i is included in the right-hand side of (6), given by

$$S_i = \sum_{k \in N(i)} W_{ik} \sum_m \delta(t - t_k^m) \quad (8)$$

where δ is the Dirac delta function and t_k^m indicates the m th firing time of oscillator k . When k spikes, it increases x_i by the weight W_{ik} . $N(i)$ indicates the set of oscillators that connects to i .

Mirollo and Strogatz analyzed a fully connected network of spike oscillators with excitatory coupling [87]. In particular, they showed that the network synchronizes when $f'(\varphi) > 0$ and $f''(\varphi) < 0$, the situation illustrated in Fig. 10(a). Furthermore, they noted in their numerical studies that synchrony could also be achieved with local coupling.

Subsequently, Hopfield and Herz carried out a systematic study on the effects of local excitatory coupling in networks of spike oscillators [55]. Their network also uses weight normalization (see Section IV-A). Hopfield and Herz reported that a 2-D locally connected network synchronizes only slowly (in about 100 periods). On the other hand, they observed that local synchronization occurs rapidly. In local synchrony, a small cluster of neighboring oscillators fires at the same time, and the whole network contains many such clusters that spike at different times. Due to the slow synchronization rate, they considered global synchrony too slow to be relevant for biological information processing. Instead, they advocated the use of local synchrony to perform useful computation. Fig. 11 illustrates the use of local synchrony to perform a useful task—image denoising. Fig. 11(a) shows a noisy image, which is generated by adding noise to a clean image with nine uniform-intensity regions on a background. In the figure, a pixel intensity is indicated by its height. To denoise, a corresponding network of spike oscillators is used that has four nearest-neighbor coupling, and the initial phases of the oscillators are set to the noisy intensities. Fig. 11(b) shows the phases of the network after local synchrony is reached. It is clear from the figure that the output resembles the original, noiseless image.

Upon further examination of the above study, Campbell *et al.* [18] found that global synchrony in the same network can be achieved quickly (in two to three periods) through appropriate choices of parameters. Their extensive numerical simulations

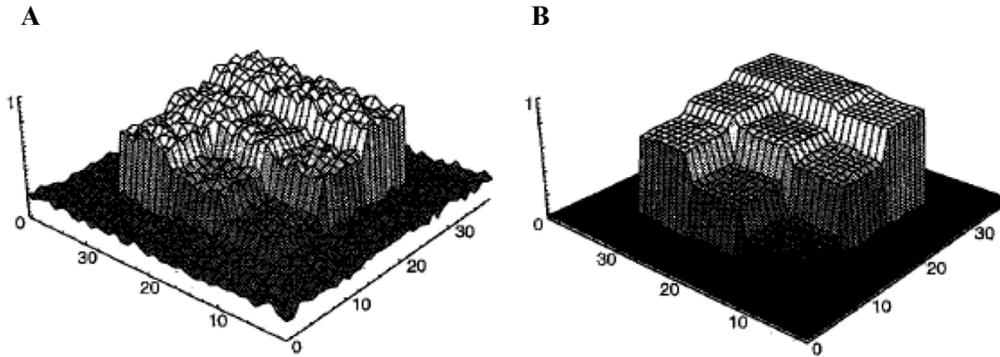


Fig. 11. Denoising with local synchrony (from [55]). (a) An image with 40×40 pixels, produced by adding uniformly distributed noise to a clean image with nine uniform-intensity regions (plateaus) on a background. (b) The phases of the corresponding 40×40 locally coupled network of phase oscillators after local synchrony is reached. The initial phases are set to the intensities in (a).

show that the time a 1-D or 2-D system takes to reach synchrony increases only as the logarithm of the system size. In addition, they studied a network of spike oscillators with local excitation and global inhibition and found that the network behavior is similar to that observed with relaxation oscillators; that is, the network exhibits both rapid synchronization within a locally coupled assembly and desynchronization among different assemblies. A major difference lies in the capacity of segmentation, or the number of different assemblies that can be desynchronized. Unlike a spike oscillator, the active phase of a relaxation oscillator is not instantaneous. As a result, a relaxation oscillator network can desynchronize only a limited number of assemblies. The number, called *segmentation capacity* [148], corresponds to the integer ratio of the oscillation period to the duration of the active phase. With an instantaneous spike, the segmentation capacity of a spike oscillator network is essentially infinite. Thus, the concept of capacity limitation does not apply to spike oscillators.

The above exposition portrays three types of oscillator model. Although these models are phenomenologically quite different, they can also be related to each other. For example, by increasing the parameter α in (3) one can shorten the active phase of a relaxation oscillation, making it approach a spike oscillator. Campbell [16], [19] showed that, by varying ε in (3), the same equation can exhibit different types of oscillation: From harmonic (sinusoidal) oscillation at one end of the spectrum to relaxation oscillation at the other end. In the middle of the spectrum one can see oscillation traces resembling those of a Wilson-Cowan oscillator. Campbell also compared networks of relaxation oscillators and those of spike oscillators in terms of rate of synchronization. His numerical results show that, for 1-D networks, the average time to synchrony with respect to network size exhibits a power scaling relationship for relaxation oscillators, whereas the corresponding relationship for spike oscillators is a logarithmic one as mentioned before. On the other hand, relaxation oscillator networks synchronize faster than spike oscillator networks in a large range of different network sizes tested. For 2-D networks, the scaling relationship for relaxation oscillators is bounded above by a logarithmic one, and again relaxation oscillator networks take less time to reach synchrony. He also pointed out that the form of coupling,

not just the basic oscillator model, is important for the rate of synchronization.

Related phenomena have been observed in other types of dynamical units. For example, Zhao and Macau found that networks of chaotic maps show similar behavior of synchrony and desynchrony as those of limit-cycle oscillators [159]. Also, Olufsen *et al.* [96] recently observed phase separation between assemblies of conductance-based neuronal models, and provided empirical evidence for phase separation from hippocampal slices.

We may conclude that all of the above types of oscillator have met the computational requirements pointed out in Section III. Of these models, analytical understanding is strongest for networks of relaxation oscillators. For the other two types of networks, the most interesting observations come from numerical studies. These advances in understanding nonlinear oscillatory dynamics have provided a firm theoretical ground for later developments that address the scene analysis problem.

V. TOWARDS THE SCENE ANALYSIS PROBLEM

The oscillatory correlation theory and related dynamics provide a general framework for addressing the scene analysis problem, which remains one of the most challenging problems in machine perception ([36, p. 10]). To deal with real-world scenes, an oscillator network needs to encode some measure of similarity between the corresponding scene elements. What determines the similarity between local sensory elements? In the visual domain, this has been systematically studied in Gestalt psychology (or Gestalt grouping, see [67] and [151]). The following is a list of major grouping principles that include both classical and new ones [98].

- *Proximity*: Elements that lie close in space tend to group.
- *Similarity*: Elements that have similar attributes, such as luminance, color, depth, or texture, tend to group.
- *Common fate*: Elements that move coherently (common motion) tend to group. We note that this may be regarded as an instance of similarity, and it is listed separately to emphasize visual dynamics.
- *Good continuity*: A set of elements that form smooth continuations of each other tend to group.

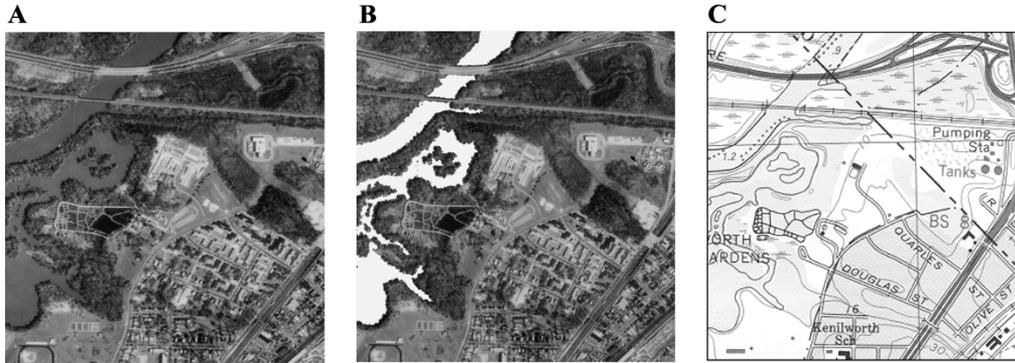


Fig. 12. Extraction of hydrographic objects (from [23]). (a) A satellite image with 640×606 pixels. (b) The corresponding extraction result, where extracted objects are marked as white. (c) The corresponding topographic map.

- *Connectedness* and *common region*: Connected elements tend to group. Similarly, elements that lie inside the same connected region tend to group.
- *Familiarity*: A set of elements that belong to the same familiar pattern tend to group.

Many of the principles are directly related to the computational requirements discussed in Section III. For real images, to apply grouping principles generally requires a separate process that extracts local features, which may be simply pixel values for intensity images or statistical features that characterize a textural pattern. In this section, we show that, in conjunction with feature extraction, oscillator networks can effectively perform challenging scene analysis tasks, including visual and auditory segmentation and selection, contour integration, clustering, and hardware implementation.

Before describing individual tasks, let us summarize the basic approach to scene segmentation. After a scene is presented, feature extraction first takes place, and extracted features form the basis for determining connection weights between oscillators, external inputs, or initial oscillator phases. The oscillator network then evolves autonomously. After a number of oscillation cycles required for the synchronization and desynchronization process, oscillator assemblies are formed and they represent resulting segments. Different segments emerge from the network at different times, and it is *segmentation in time* that distinguishes this approach from others.

We note that various aspects of scene analysis have been extensively studied outside of neural networks (see, e.g., [36]). The following survey is obviously limited to studies that explore the time dimension in neural networks.

A. Image Segmentation

The first successful study on segmenting real images was reported by Wang and Terman [148] on the basis of a LEGION network. Two issues immediately arise when handling real images: noise on an image and computing time required for integrating a large oscillator network. Image noise can lead to many fragments which, in the presence of a limited segmentation capacity, can deteriorate the result of LEGION segmentation. To address this problem of fragmentation, Wang and Terman [148] introduced a lateral potential for each oscillator, which allows the network to distinguish between major blocks and noisy fragments. The basic idea is that a major block must contain at least

one oscillator, denoted as a leader, which lies at the center of a large homogeneous region. Such an oscillator receives large lateral excitation from its neighborhood, and thus its lateral potential is high. A fragment does not contain a leader. The collection of all fragments is called the background. To alleviate the computational burden of integrating a large oscillator network, they abstracted an algorithm that follows main steps in the numerical simulation of LEGION dynamics, such as two time scales, jumping, and spread of activation. The resulting algorithm also removes the segmentation capacity. In their system, each oscillator is connected to its eight nearest neighbors, and the connection weight between two neighboring oscillators i and j is set proportional to $1/(1 + |I_i - I_j|)$, where I_i and I_j indicate the corresponding pixel intensity values. The key parameter for segmentation is the level of global inhibition: W_z [see (4)]. Larger values of W_z produce more and smaller segments. To further reduce sensitivity to image noise while preserving important features, Chen *et al.* [23] proposed the idea of adapting dynamic weights to perform feature-preserving smoothing before LEGION-based segmentation. Their weight adaptation method is insensitive to termination times—a common problem in image smoothing techniques. Moreover, they suggested to employ a logarithmic coupling term, which can be written as [see (4)]

$$S_i = \frac{\sum_{k \in N(i)} W_{ik} H(x_k - \theta_x)}{\log \left(\sum_{k \in N(i)} H(x_k - \theta_x) + 1 \right)} - W_z H(z - \theta_z). \quad (9)$$

The resulting algorithm performs better than other segmentation algorithms in image processing, such as nonlinear smoothing and multiscale segmentation [23], [74]. Fig. 12 gives one example of extracting hydrographic objects from satellite images from the United States Geological Survey (USGS). The original image containing water bodies is shown in Fig. 12(a). Fig. 12(b) shows the corresponding extraction results, where the water bodies are marked as white and superimposed on the original images. Fig. 12(c) provides the corresponding 1:24 000 map from USGS. A careful comparison between the extracted waterbodies and the map indicates that the former portray the images better, because the map does not accurately reflect geographical changes.

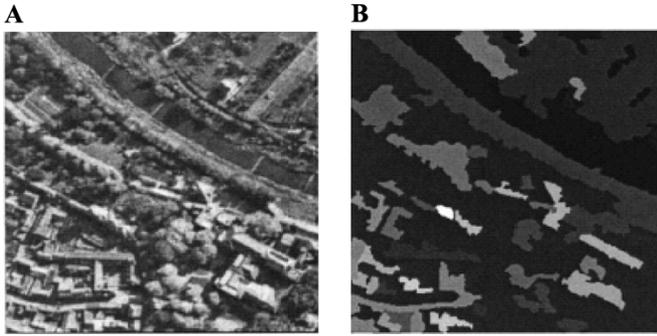


Fig. 13. Aerial image segmentation (from [122]). (a) Input image. (b) Result of segmentation. Here, each segment corresponds to a distinct gray level.

Using a spike oscillator model modified from that proposed by Eckhorn *et al.* [37], Kuntimad and Ranganath studied a laterally coupled network for image segmentation [69]. Their network uses an input image to set the external inputs of the corresponding oscillators. This encoding method entails that different pixel intensities convert to different spiking frequencies. Each oscillator is connected to neighboring ones, and a connection weight is set to the reciprocal of the square distance between two oscillators. From the computational perspective, the coupling here plays the role of image smoothing. When the network successfully separates an image, different regions correspond to different oscillator assemblies that are synchronized within assemblies but exhibit different frequencies for different region intensities. Kuntimad and Ranganath gave a successful example in which a synthetic image containing regions with overlapping ranges of intensities is correctly segmented. The mapping of intensity to spiking frequency creates some difficulty when dealing with real images where multiple objects have similar intensities. In a recent study, Stewart *et al.* [122] introduced a form of inhibition and coupling variations to avoid this difficulty and applied the resulting network to segmenting real images. Fig. 13 shows an example of their segmentation performance. Fig. 13(a) shows an aerial image, and the segmentation result is given in Fig. 13(b), where each segmented region corresponds to a different gray level. How to decode a segmentation result from an oscillator network is the issue of readout, to be discussed in Section VII-B. Eckhorn *et al.* [39] recently discussed related coding and coupling schemes in the visual cortex and suggested a number of biologically plausible mechanisms for scene segmentation.

Using Wilson–Cowan oscillators, Ursino *et al.* [130] recently proposed the idea of using detected contours to assist image segmentation. More specifically, they proposed a two-layer network: the first layer performs contour detection and the second layer segmentation. Oscillators activated by object contours in the first layer inhibit the corresponding oscillators in the second layer, leading to separation of different objects in the second layer. However, tests on real images are not conducted.

Other successful attempts include segmentation of motion, texture, and range images. Common motion, or common fate, is a major grouping cue, as we discussed earlier. Cesmeli and Wang [20] applied LEGION to motion-based segmentation, and their model considers both motion and intensity for analyzing image sequences. Recent studies by Zhang and Minai [158] and

Yu and Zhang [157] make use of spike oscillator networks. Texture is elusive to characterize, and texture-based segmentation has proven to be very difficult in computer vision. Baldi and Meir [6] gave an early demonstration of oscillation-based texture discrimination using phase oscillators. A recent model by Cesmeli and Wang [21] first extracts a set of texture features, which are then encoded as coupling strengths in a LEGION network. Liu and Wang [75] studied range image segmentation using relaxation networks. A range image is an image of depth values directly obtained from a range sensor. The basic feature in this study is a five-dimensional vector including depth, surface normal, and curvature, respectively.

B. Contour Integration

While the systems described above are primarily concerned with separating images into segments that are image regions, a closely related task is extract contours of different regions.⁸ Horn and Opher [56] proposed a network of spike oscillators for detecting borders between regions. Their network uses difference-of-Gaussian coupling, which corresponds to local excitation and lateral inhibition, the latter covering a broader area than excitation but not as extended as global inhibition. Their simulations suggest that at local minima of total network activity, firing oscillators tend to correspond to edges.

Yen and Finkel [156] used a network of phase oscillators to extract salient contours. Excitatory and inhibitory connections in their network encode orientation and distance relations between edge filters that correspond to oscillators. Specifically, the coupling term S_i to oscillator i on a 2-D network takes the following form:

$$S_i(\mu) = \sum_{k \in N(i)} \sum_{\nu} W_{ik}(\nu - \mu) x_k(\nu) - z \quad (10)$$

where μ indicates the dominant orientation corresponding to oscillator i , ν the orientation of oscillator k in the neighborhood of i , indicated by $N(i)$, and x_k the activity of oscillator k . The connection weight from k to i , W_{ik} , depends on two main factors. First, the orientation μ at location i specifies i 's preferred orientation in $N(i)$, which satisfies the collinearity rule as well as other linking conditions suggested from neurobiological and psychophysical data. As the orientation ν deviates from the preferred orientation, W_{ik} decreases as a Gaussian function. Second, the connection strength decreases as the distance between i and k increases, again following a Gaussian function. The term z indicates global inhibition, which plays the role of suppressing background activity that arises in response to scattered stimulus items. The salient contours in an image correspond to oscillator assemblies that can overcome global inhibition and emerge from the network. The saliency of a contour is indicated by the total activity of the corresponding assembly.

Fig. 14 shows an example of contour integration. Fig. 14(b) shows the output from the network on the input image shown in Fig. 14(a). The saliency of a contour is indicated by the brightness in the figure. In this case, the network successfully detects a

⁸Given segmented regions, contours are readily obtained. On the other hand, obtaining regions from contours is more complicated. It involves linking pieces of contours and forming closed contours that bound their regions, which can be quite challenging computationally.

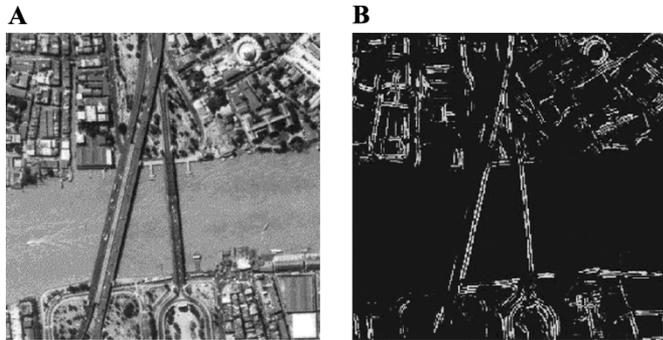


Fig. 14. Contour integration. (a) An intensity image with 385×385 pixels. (b) Result of contour integration, where extracted contours correspond to connected ones in the figure. The saliency is indicated by brightness. (Reproduced with permission from [155].)

number of salient contours. Note that the question of desynchronizing different contours is not addressed in this paper. They suggested that an attentional mechanism could be subsequently applied to select different contours, a topic to be discussed in Section V-D.

Li [72] proposed a similar model to perform contour integration, and her study is intended to model the primary visual cortex where such a segmentation process likely occurs. The model incorporates neurobiological data, but results on real images are not available. A more recent study by Choe [24] uses a network of spike oscillators to address the issue of contour segmentation via desynchronized activity, in addition to contour integration. Instead of using a predetermined connectivity pattern, his model uses self-organization from input scenes consisting of orientation bars in order to arrive at appropriate connections. The model then uses organized connections and synchronized firing to simulate psychophysical data on contour integration.

C. Clustering

Rhouma and Frigui [103] studied networks of spike oscillators to perform data clustering. There is an intrinsic connection between segmentation and clustering; a clustering algorithm organizes a collection of data points into a small number of clusters so that there is similarity within each cluster and dissimilarity between different ones. They extended the analysis of Mirolla and Strogatz [87] to cases with certain inhomogeneity, and specified conditions where a fully coupled network synchronizes. They also introduced a form of global inhibition in order to achieve phase desynchrony among different assemblies.

To apply to data clustering, an oscillator represents a data point, and coupling between oscillators encodes similarity between the two corresponding data points. In addition, a global threshold dictates whether coupling is excitatory or inhibitory and it plays a role similar to the global inhibition discussed earlier. Rhouma and Frigui proposed two algorithms for two distinct types of clustering: relational and prototype-based [103]. For relational clustering, explicit pairwise coupling between oscillators is used, whereas for prototype-based clustering, a prototype for each cluster is used to represent an average data point of the cluster and it is updated constantly during the clustering process. The coupling method explicitly ensures that once an

assembly of oscillators is synchronized, it will remain synchronized in the future evolution of the network. In comparison with conventional clustering techniques, this clustering method has the advantage that it can produce a nested sequence of partitions with different numbers of clusters as one varies the global threshold systematically.

Rhouma and Frigui also applied their prototype-based clustering to color image segmentation. Here a data point corresponds to a single pixel and is an eight-dimensional vector consisting of color, texture, and 2-D location. Fig. 15 shows the result of segmenting a color image, where Fig. 15(a) shows the original image and Fig. 15(b) the segmented result. Segmented regions correspond to connected ones with different gray levels and are bounded by closed contours in Fig. 15(b).

Bohte *et al.* [12] also investigated data clustering using a spike oscillator network. Instead of emphasizing synchronization, this paper explores spike timing to encode each data point. Specifically, a multidimensional data point is represented by a population of spike oscillators whose spiking times are uniquely determined by the data point. With such encoding, an output layer that receives input from the oscillator population can be trained using unsupervised Hebbian learning so that each output oscillator is activated by a cluster of data points. Training here strengthens the connections from the input layer to the output layer that have appropriate delays, and eventually the data points in the same training cluster trigger collections of input spikes that arrive at the corresponding output neuron simultaneously. This way similar data points are clustered together after training is complete. They have applied their method to classifying a set of remote sensing data. Fig. 16 shows an example of classifying three-band RGB image. The original image is given in Fig. 16(a), and the classification is performed on a smaller image cut out from the center shown by the inset. The result of classification is shown in Fig. 16(b), where the number of classes is set to 17 in advance.

D. Visual Selection

A classic topic in neural networks is neural competition. Winner-take-all (WTA) networks have been extensively studied [2], [34], [41], [51], [111]. WTA dynamics is based on global inhibition, either in the form of a global inhibitor or mutual inhibitory connections, and produces a winner that has the highest input. Competitive dynamics has played a major role in modeling selective visual attention. Typically, a saliency map is introduced that integrates local feature extraction at different spatial locations, and WTA then takes place to select a winning location to direct attention to [66].

Niebur *et al.* [95] modeled selective visual selection using spike oscillators. In this model, oscillators whose receptive fields lie in the focus of attention generate correlated spikes, whereas those oscillators outside of the focus generate randomly distributed spikes. As a result, they employ coincidence detection to suppress responses associated with unattended stimuli. Niebur and Koch [94] later extended the study to address the issue of attentional control by including a saliency map, which summarizes input from the feature maps for detecting intensity, color, orientation, and change. Selection on the saliency map is implemented by a network of spike oscillators

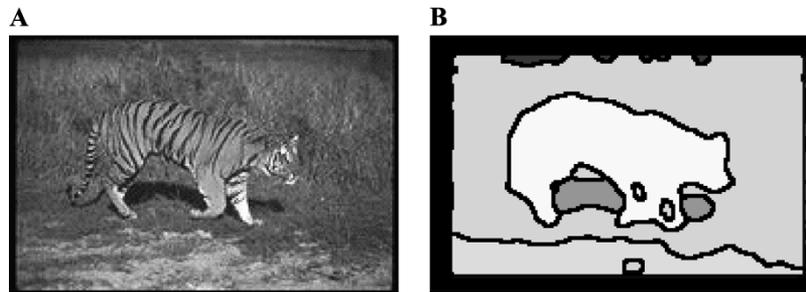


Fig. 15. Color image segmentation. (a) A color image with 128×192 pixels (shown in gray). (b) Segmentation result. Here distinct segments correspond to different gray levels, enclosed by closed contours. (Reproduced with permission from [103].)

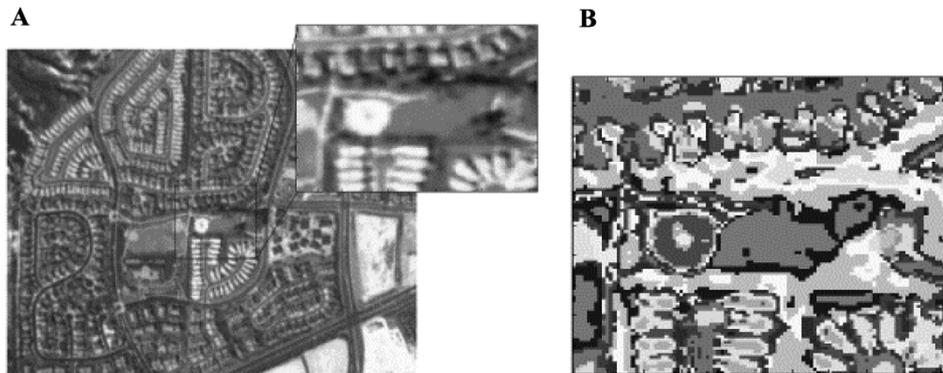


Fig. 16. Clustering of multidimensional data. (a) Full RGB image along with a cutout image shown in the set. The size of the cutout image is 99×103 , resulting in 10 197 data points. (b) Result of classification after unsupervised training using 70 000 randomly chosen data points from the cutout image. (Reproduced with permission from [12].)



Fig. 17. Modeling attentional shifts. The thick lines show the shifting trajectory of attentional focus, which jumps from the center of the image to three other locations of decreasing saliency. (Reproduced with permission from [94].)

with strong global inhibition. In addition, a winning location on the saliency map receives feedback inhibition that causes shifting of attention to other salient locations. An example of such attentional shifts is given in Fig. 17 on a real image. The figure shows three shifts starting near the center of the image that contains a banner on a wall, which the system considers to be the most salient. Location-based selection has also been modeled using phase oscillators [26], [154].

In WTA, individual neurons compete with each other, which corresponds to local representations. For perceptual processing, however, experimental data suggest that objects may act as wholes in competition [33], [35], [92]. But, in order to capture object-level competition, one must address the binding issue.

Exploiting the LEGION mechanism for binding and slow inhibition, Wang proposed an architecture for size-based object selection [141]. In other words, in an input scene with many objects (patterns), the network attempts to select the largest one. The basic idea is that after oscillator assemblies are formed, competition between assemblies takes place in time. When an assembly jumps to the active phase, it leaves an inhibitory trace via a slow inhibitor, which can be overcome only by larger assemblies. After a number of oscillation cycles, the largest assembly will be the only one that oscillates, while all the others are suppressed. A later study by Kazanovich and Borisyuk [62] addressed the same problem, size-based selection, using a different kind of oscillator network where oscillation frequency and amplitude play a major role in the selection process. The study shows that a number of simple objects can be selected consecutively using oscillatory correlation.

Note that size as an example of the saliency in the above selection models is an object-level property, whereas saliency encoded in a saliency map is a local, location-specific property [66], [94]. Object selection models are compatible with object-based theories of visual attention, while WTA models are compatible with location-based theories; see Pashler [100] and Parasuraman [99] for discussions on the two contrasting theories of visual attention.

E. Auditory Segregation and Selection

Similar to the visual domain, a listener in an auditory environment is exposed to acoustic wave from multiple sound sources. To understand the auditory environment, the listener must first

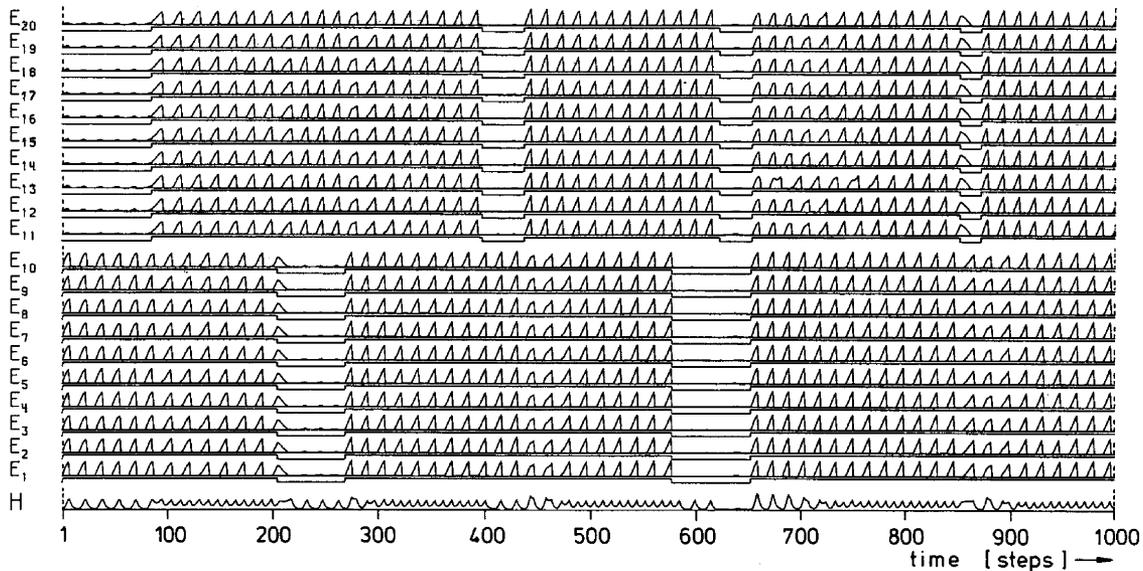


Fig. 18. Onset-based auditory segregation. Shown are the responses of 20 oscillators and the global inhibitor (indicated by H). The oscillators $E_1 - E_{10}$ are activated by one stimulus and $E_{11} - E_{20}$ by another one. As indicated by the traces underneath the two groups of oscillators, the stimuli switch on and off independently.

disentangle the acoustic wave and organize it into streams corresponding to different sources. This process is referred to as auditory scene analysis [13]. According to Bregman [13], auditory scene analysis takes place in two stages. In the first stage, the acoustic mixture reaching the ears is decomposed into a collection of sensory elements. In the second stage, elements that are likely to have arisen from the same source are grouped to form a stream that is a perceptual representation of an auditory event. Important cues for auditory grouping include proximity in frequency and time, smooth temporal transition, onset and offset coincidence, and common location.

Von der Malsburg and Schneider [137] conducted the first study on auditory segregation using oscillatory correlation. As discussed in Section II, the grouping cue in this study is common onset. The model is based on a globally connected oscillator network plus global inhibition. Mutual excitation among the oscillators serves to synchronize a stream with a common onset, and global inhibition plays the role of desynchronizing streams. In order to form appropriate oscillator assemblies, they employ dynamic links that rapidly establish strong excitatory connections among the oscillators triggered by a common onset. Fig. 18 shows two oscillator assemblies thus formed by their network. The first assembly consists of the oscillators labeled as $E_1 - E_{10}$ and the second one those labeled as $E_{11} - E_{20}$. The onsets and offsets of a stimulus are indicated by a line underneath each assembly. The bottom trace of the figure shows the activity of the global inhibitor. Fig. 18 clearly indicates that oscillators responding to the same stimulus are synchronized and the two oscillator assemblies are desynchronized (compare oscillatory traces when both stimuli are on).

By extending LEGION to the auditory domain, Wang [140] proposed an oscillator network to address stream segregation. The basic architecture is a 2-D LEGION network: one dimension represents time and another one represents frequency. The network demonstrates a set of psychophysical phenomena,

including dependency of stream segregation on spectral and temporal proximity, sequential capturing, and competition among different perceptual organizations [13]. Wang and Brown [144] studied the speech segregation problem: separating target speech from its acoustic interference. Echoing Bregman's two-stage notion, their model consists of a stage that computes sensory elements or segments from an auditory scene and a subsequent stage that groups segments into a target speech stream and a background. Segment formation is based on temporal continuity and cross-channel correlation between filter responses of adjacent auditory channels. Grouping is based on the dominant pitch computed in consecutive 20-ms time frames with 10-ms frame shift. Segment formation is performed by a LEGION network, and grouping is carried out by a laterally connected network of relaxation oscillators. An example is given in Fig. 19, where the input scene is a mixture of male utterance and telephone ringing [see Fig. 19(a)]. The segregated target is shown in Fig. 19(b) and the background in Fig. 19(c).

Wrigley and Brown [153] recently proposed an oscillatory correlation model of auditory attention. After auditory peripheral processing, their model first extracts auditory segments using a LEGION array. The main part of the model is an interactive loop between an oscillator network that performs stream segregation and an attentional leaky integrator that selects one dominant stream from the output of stream segregation. The connections between the oscillator network and the leaky integrator are modulated by the attentional interest of the model. Their model simulates several perceptual phenomena, including two-tone streaming and sequential capturing. At the conceptual level, a main difference between this model and Wang's model [140] is that the former allows attention to be directed to only one stream at a time whereas the latter claims that attention may be shared by more than one stream. The issue of capacity will be further discussed in Section VII-C.

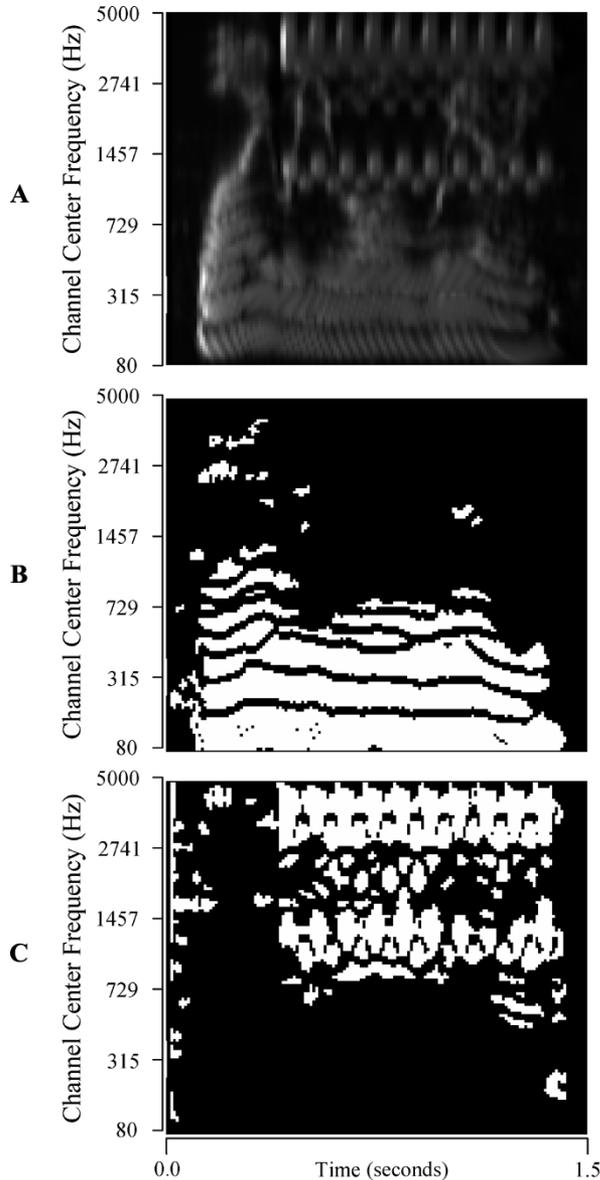


Fig. 19. Speech segregation (from [144]). (a) Filter responses to a mixture of male utterance and telephone ringing. The response pattern is generated by 128 filter channels, whose center frequencies range from 80 Hz to 5 kHz, over 150 time frames. (b) Segregated target speech, indicated by white pixels representing active oscillators at one time. (c) Segregated background, indicated by white pixels representing active oscillators at a different time.

F. Hardware Implementation

One of the advantages of computing with neural networks is that it would facilitate direct hardware implementation. Circuit implementation compared to computer simulation has attractive features such as small area, high speed, and low power consumption, which in turn would facilitate real-time autonomous processing of complex tasks [25], [82].

Numerous studies have been conducted to implement single oscillators. The fact that the van der Pol oscillator was originally proposed to model a triode circuit indicates a close relationship between mathematical oscillations and electronic circuits. Nagumo *et al.* [91], when studying the FitzHugh–Nagumo equations, also gave a circuit that exhibits relaxation oscillations described by the equations. Later, Keener [63] described an inte-

grated circuit that implements the FitzHugh–Nagumo oscillator. Mahowald and Douglas [79] designed a circuit that implements a conductance-based neuron model. Their silicon neuron includes dendritic and somatic compartments and emulates ionic conductance. As a result, the circuit generates a spike train in response to injected currents.

Kinser and Lindblad [64] gave a digital implementation of spike oscillator networks studied by Eckhorn *et al.* [37]. Their implementation uses a connected network of adapted processors, which is a parallel architecture optimized for pattern recognition and signal processing. The resulting system is a single-instruction multiple-data parallel computer, where each node processes its own data while all active nodes execute the same instruction at a time. They implemented a 128-node array processor and tested it on target detection tasks.

Ando *et al.* [3] described circuit implementation of a relaxation network for intensity image segmentation. They modified the original (3)–(5) to facilitate very large-scale integration implementation; in particular, the coupling term S_i is made bipolar before adding to the right-hand side of (3a). Their implementation employs pulse-width modulation (PWM) and pulse-phase modulation (PPM) techniques, which can realize arbitrary nonlinear dynamics. PWM and PPM circuits seem well suited for implementing relaxation oscillator networks, where an oscillation has a noninstantaneous active phase. Cosp *et al.* [28] also studied neuromorphic implementation of relaxation oscillator networks. The analog circuit design employs a hysteresis current comparator and a damped integrator for a single oscillator. Cosp and Madrenas [27] later fabricated a chip implementing a 16×16 LEGION network using CMOS technology, and evaluated the chip on a number of segmentation tasks. The evaluation results demonstrate that the chip performs segmentation with much higher speed as well as area and power consumption advantages in comparison to simulation on a general-purpose computer.

G. Further Remarks

To perform scene segmentation effectively, both feature extraction and grouping are important. Feature extraction is modality- and cue-specific, whereas a binding mechanism may be generally applicable. Taken together, the body of work reviewed in this section gives a strong indication that oscillator networks provide a general and effective mechanism for scene segmentation. We believe that the grouping mechanism has, for the main part, answered Rosenblatt's challenge regarding figure-ground separation. His challenge regarding spatial relations will be discussed in Section VII.

Given a diverse range of oscillator models that have been investigated for oscillatory correlation and applied to scene analysis, an important question is to what extent do these models differ from the computational perspective? In other words, are there substantial differences when used as a binding mechanism? In order to answer this question, one needs to examine an oscillator network at multiple levels. At the level of a basic oscillator, as discussed in Section IV, a number of distinct oscillator models, properly coupled, can all meet the computational requirements specified in Section III. Hence, we consider that the details of an individual oscillator do not matter

as long as the model meets the computational requirements. A clear example is offered by Campbell *et al.* [18], where a segmentation task achievable using a network of relaxation oscillators can be similarly achieved using that of spike oscillators. That said, there can be differences in terms of computational efficiency and characteristics such as the segmentation capacity [16], [18]. At the level of network architecture, probably dictated by the computational requirements, most of the network models studied so far take some form of local connection to achieve grouping within the same object and long range connections to achieve segmentation between different objects. At the level of scene representation, we see distinct schemes that could lead to large differences at the computational level. Regarding how to represent different objects of a scene, two different schemes are phase (see, e.g., [148]) and frequency (see, e.g., [69]). Regarding how to represent different feature values, say, pixels, three methods have been studied: connection strength (e.g., [148]), initial phase (e.g., [55]), and frequency (e.g., [69]). Although amplitude has not been systematically studied, one can imagine that different oscillation amplitudes could also be used to encode different objects or pixel values (see [62]).

As discussed at the beginning of this section, the familiarity principle of grouping requires a memory recall and may be viewed as a top-down process. Memory-based segmentation has been previously studied [57], [65], [77], [120], [145]. But, without a primitive segmentation stage, the computational performance of these models is very limited. A general solution to the scene analysis problem seems to require an approach that integrates bottom-up segmentation and top-down analysis. Stoecker *et al.* [123] proposed a multilayer model that combines an invariant pattern representation, top-down connections, and primitive segmentation. However, the performance of this complex model is uncertain. Recently, Wang and Liu investigated the integration of primitive segmentation and associative memory [146]. Their model consists of initial primitive segmentation, multimodule associative memory, and a short-term memory (STM) layer. Primitive segmentation separates an input scene into multiple segments, each of which then activates the memory layer, and potentially multiple recalls interact in the STM layer, resulting in a common part. The top-down input from the STM layer performs memory-based grouping and segmentation. The systematic evaluation shows that memory-based organization is responsible for a significant improvement in scene analysis performance.

Nonoscillatory methods have also been studied for the scene analysis problem in neural networks. Besides those discussed in Section I, Grossberg and Wyse [52] proposed a model that iterates between a feature contour system and a boundary contour system. The feature contour system detects local image features and performs diffusion within a region, while the boundary contour system detects local edges and performs contour completion. For grouping, a filling-in process is introduced to spread a region label until reaching a boundary. Opara and Wörgötter [97] proposed an energy-based cluster update algorithm for image segmentation. In this method, each image feature (e.g., a pixel) is represented by a spin variable that has a fixed set of discrete spin states. Each spin state corresponds

to a region label. The labeling process then corresponds to updating spin variables, and the update procedure is guided by an energy function that encodes intensity similarities between nearby pixels. Neumann and Sepp [93] addressed boundary processing with a recurrent network that models the V1–V2 interaction in the visual system. Their model successfully simulates a number of psychophysical results concerning visual boundary perception. A recent model by Wersing *et al.* [150] segments multiple salient contours in a competitive layered network, where each contour corresponds to a separate layer. To perform contour segmentation, the model uses local cooperation within each layer, competition among different layers at the same location, and global inhibition to ensure that remote contours are segmented. In comparison to oscillatory correlation methods, nonoscillatory methods tend to be less flexible since the number of labels needs to be predetermined. On the other hand, oscillator networks tend to be difficult to control without substantial knowledge in dynamical systems.

VI. BIOLOGICAL RELEVANCE

There is an extensive and growing body of physiological evidence that supports the existence of coherent oscillations in various cortical regions as well as their potential role in feature binding. Experimental results come from different modalities and different animal species and humans, in both anesthetized and awake conditions. Early evidence of neural oscillations was obtained from sensory evoked potentials in the olfactory system [44] and the auditory system [46]. The accumulation of evidence has accelerated following the discovery of synchronous oscillations from cell recordings in the cat visual cortex [38], [50]. Detailed reviews are given by Singer and Gray [117], Usrey and Reid [131], and Varela *et al.* [134] and are not considered here. Two points are worth noting. First, theoretical investigation [84], [135] on the binding problem predates and directly influences empirical work that uncovered visual coherent oscillations. Second, neural oscillations have been a widely debated topic in neuroscience; see, for example, the *Neuron* issue mentioned in Section II. It is fair to state that neural oscillations and synchrony are clearly present [108], and that the debate has largely been shifted from whether coherent oscillations exist to whether they play a major role in binding.

The following summarizes several important aspects of the experimental data on cortical oscillations and neural synchronization. Frequencies of coherent oscillations from different modalities and animal species generally range from 30 to 70 Hz, often referred to as 40 Hz oscillations. This frequency range is compatible with that of electroencephalograph gamma rhythms, and hence such oscillations are also called gamma oscillations. Cortical oscillations depend on external stimulation, but not on oscillating input. They are thus referred to as stimulus-dependent, not stimulus-driven. Synchrony in neural oscillations occurs across a considerable extent of the cortex, beyond the distance of direct connections between cortical cells. The presence or absence of coherent oscillations correlates with perceptual organization for a broad range of perceptual stimuli (e.g., [45]), including the influence of top-down processing [40]. Finally, recent evidence from awake animals

(e.g., monkeys) suggests that neural synchrony may occur rapidly without oscillations. Note that fast synchronization and desynchronization are emphasized in Section III, where we have pointed out that model oscillators may not produce periodic or rhythmic activity, particularly with changing input.

Much of the physiological evidence on coherent oscillations is uncovered in the visual system. In the auditory system, 40-Hz oscillations in localized brain regions have also been recorded both at the cortical level and at the thalamic level, and these oscillations are synchronized over considerable cortical areas [76], [104]. Joliot *et al.* [59] found evidence that directly ties coherent 40 Hz oscillations to perceptual grouping of clicks. Cell recordings in the auditory cortex show that neurons exhibit synchronous firing activity [31], [32], [80]. The study by Barth and MacDonald [9] suggests that oscillations in the auditory cortex originate within the cortex and synchrony is produced by intracortical interactions. The suggested anatomical substrate for coherent oscillations well agrees with that from the visual domain [117]. Recently, Brosch *et al.* [14] found stimulus-dependent gamma oscillations from the monkey primary auditory cortex using multielectrode recording. The occurrence of the gamma oscillations correlates with the match between stimulus frequency and a unit's preferred frequency. In addition, gamma oscillations at different recording sites can synchronize depending on spatial distance and the match between preferred frequencies.

VII. DISCUSSION

In this section, I discuss a number of issues pertinent to temporal coding and oscillatory correlation. I also explore the potential of the time dimension for computing spatial relations and resolving the problem of compositionality.

A. Internal Time versus External Time

In the oscillatory correlation theory, time plays the role of binding: different segments unfold in time. Let us refer to this putative role of time as *internal time*. Time is also a dimension in the physical world, and indeed a defining dimension for auditory and other temporal patterns. To complicate the matter further, time, in the form of common onsets and offsets, is also a grouping principle [13], [70], [71]. Let us refer to this use of time as *external time*, which is external to the organism. A potential difficulty for the double use of time has been raised [15] and used as an argument against the temporal correlation theory [114].

Onset/offset detectors have been identified in the auditory system [102] and have been used in auditory models. With such detectors, grouping based on common onsets/offsets is not unlike that based on other cues. A similar idea can be extended to the visual domain. How to distinguish internal time from external time as used in temporal patterns depends on how time is represented in temporal patterns. In neural network modeling of temporal patterns, time is usually coded by delay lines, decay traces, or exponential kernels [8], [143]. Delay lines convert time to space and hold the most recent patterns for a certain period of time. This way of representing time has also been used

in the context of auditory segregation (see Section V-E), and as demonstrated in these studies, the potential conflict of the double use is not present. Decay traces encode time implicitly and compactly but have limited discriminative power; for instance, it is unclear how they could underlie a variety of auditory functions, such as pitch and rhythm perception. The use of exponential kernels is a compromise between these two cases, and it converts time into a logarithmic axis of space so that more recent traces are represented with higher temporal resolutions. The distinction between internal time and external time can be made similarly as in the case of time delays. Both delay lines and exponential kernels form a shifting representation. Given the high resolution of temporal processing in the auditory system [88], internal time would need to be preserved during the shifting process. This raises the interesting issue of how internal time can be maintained during neural transmission. One such neural mechanism is the synfire chain proposed by Abeles [1], which refers to the notion of a chain of diverging and converging links that maintain synchronous activity during transmission.

B. Readout

An issue often cited as a problem for the oscillatory correlation representation concerns how a synchronized code is decoded in a later processing stage (see [48]). The readout problem is not really unique to temporal coding and, as pointed out by Roskies [108], it is "one of the most puzzling and fundamental problems for systems neuroscience in general." If later processing, say, recognition, requires information across a large part of the visual field, whether that information is encoded via temporal correlation or any other means, it must somehow be decoded in a corresponding way.

While the readout issue in temporal coding is an open question in neuroscience, some computational considerations may be helpful. As illustrated in Fig. 9, the basic claim of oscillatory correlation is that each segment pops out at a distinct time from the network and different segments alternate in time. In the case of relaxation oscillator networks, a segment is in the active phase when it pops out. As a result, all of the features of the segment, but none of the features from competing segments, are simultaneously available for later processing tasks such as selective attention and recognition. This way of encoding a pattern, i.e., activating all of its features, is very commonly used in neural models for pattern recognition, e.g., perceptrons.

When different segments are encoded as different oscillation frequencies (see, e.g., [69]), the readout becomes more complicated since a frequency analysis would need to be performed as part of decoding. Also, different segments may exhibit comparable frequencies and this needs to be dealt with as well. However, no fundamental difficulty seems to exist even for frequency encoding.

C. Capacity

A basic implication of the oscillatory correlation theory is a capacity limitation on segmentation and binding (see Section IV-C). When studying memory-based segmentation in an oscillator network, Wang *et al.* [145] made the first link between the model capacity of memory segmentation and the so-called

magic number, 7 ± 2 , of human STM capacity [83]. Subsequently, Lisman and Idiart [73] developed a more detailed, oscillation-based STM model, where the 7 ± 2 capacity results from the interaction between the gamma oscillation and a slower rhythm in the theta-alpha range (5 to 12 Hz). The magic number seven symbolizes the existence of a limited capacity and should not be taken literally. A recent, comprehensive examination concludes that the capacity is actually about four [30].

The notion of a limited capacity arises naturally from relaxation oscillations, which have a noninstantaneous active phase. This characteristic of relaxation oscillation is not shared by spike oscillators [18] or chaotic maps [159], which in theory are not subject to capacity limits. Though the existence of such a capacity is sometimes viewed as a computational limitation [150] [159], capacity limitation is a fundamental property of cognitive processing. Capacity limits arise from a variety of information-processing tasks, including memory retrieval, attention, mental operations (e.g., addition), enumeration (subitizing), multiobject tracking, etc. (for reviews, see [30] and [100]). Arguments have also been made that limited capacity is a strength rather than weakness for information processing (e.g., [61], [78]).

It is well documented that both STM and attention exhibit a limited capacity. What is the relation between them? Though often discussed in the literature as related, few studies directly address the question. The clearest answer is attempted by Cowan [29], [30], who provides a theoretical framework that ties a large number of studies from a variety of empirical paradigms. According to him, the focus of attention has a capacity about four, and this is the only source of capacity limitations in cognitive processing; in other words, capacity limits exhibited from STM and other tasks result from capacity-limited attention. Attention provides a “global workplace” for mental operations [5], [30].

D. Spatial Relations

In terms of addressing Rosenblatt’s challenge, our discussion so far is exclusively on the figure-ground separation problem. The other problem is how to compute topological or spatial relations among objects. I asserted earlier that one needs to first solve the separation problem in order to compute geometrical relations. Given the challenge of solving the figure-ground separation problem, very little research has been conducted to compute spatial relations.

Building on the LEGION ability to perform figure-ground separation, Chen and Wang [22] addressed a related question: how to tell whether a dot belongs to area *A* or area *B*, as illustrated in Fig. 20. This has been formulated as how to compute the inside/outside relation [128]. The solution by Chen and Wang is to first separate the two areas apart, and then decide whether the oscillator corresponding to the dot is in the assembly representing *A* or that representing *B*. It is interesting to contrast this solution with that offered by Ullman [128], [129] in his framework of visual routines. His suggested solution to the inside/outside problem is a visual routine called coloring, which spreads activity from the dot location until the boundary of an area is reached. Although LEGION dynamics for synchronization has some resemblance to a coloring process, there are

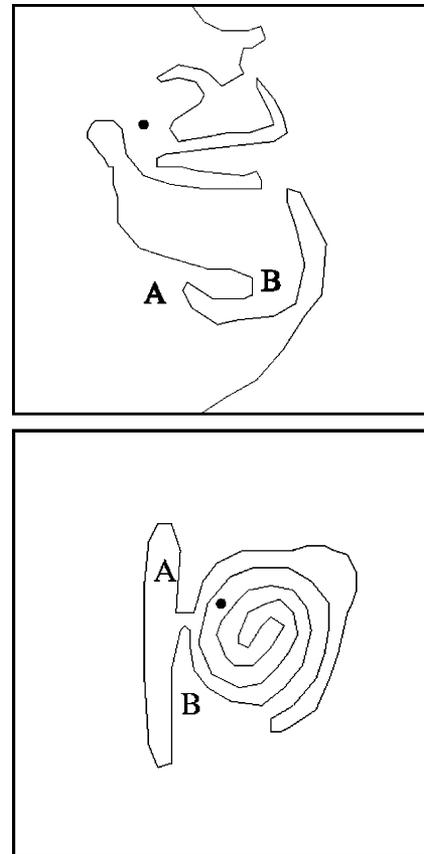


Fig. 20. Inside/outside relation. The top panel shows an example where the boundary between area *A* and area *B* is not very convoluted, whereas the bottom panel shows another example with a highly convoluted boundary.

several differences between the two solutions. First, the coloring routine is described as a serial algorithm, while for LEGION synchronization emerges from a network of interacting oscillators. Second, the time course of synchronization enables Chen and Wang to explicitly distinguish between effortless perception with simple boundaries (see the upper panel of Fig. 20) and effortful perception with convoluted boundaries (see the lower panel of Fig. 20). It is not clear how a qualitative distinction can be made in the coloring process.

I believe, for the following reasons, that the oscillatory correlation theory may lay a general foundation to compute spatial relations between multiple objects. Arbitrary objects can be segmented in time, and relevant ones for computing a specific relation can be further selected (see Section V-D). As a result, the relevant objects are all activated and yet separated in phase; this provides a workspace [5], [30] to calculate attributes from each object and compare them across different ones. Representing each object as an oscillator assembly facilitates the derivation of object-level properties, such as its center, size, spatial extent, etc., which are important for computing geometrical relations (e.g., *left of*). In essence, this way of dealing with spatial relations endows different objects with a time axis, or converting space to time, which, interestingly, runs counter to standard representations of temporal patterns in neural computation where time is converted to space. How spatial relations can be systematically computed, however, is an important topic for future research.

E. Compositionality

The issue of compositionality has received much attention in the debate between connectionism and symbolism (e.g., [43] and [118]). The question is whether neural networks possess the representational power to deal with combinatorial structure that is manifested in our ability to process relational and syntactical information. A trivial “yes” answer can be derived from the fact that neural networks are general-purpose computing devices that can simulate universal Turing machines, but such a recourse does not address the critical assessment that neural networks are merely an implementation theory and cannot account for cognitive functions. The key issue is whether a fixed network can encode and process hierarchical relations in a flexible way.

The introduction of the time dimension may open an entirely new avenue to address the issue of compositionality. With network architecture fixed, the time dimension provides the critical flexibility: hierarchy could be encoded in time. In the context of range image segmentation, Liu and Wang [75] showed how a range-defined object can be hierarchically decomposed into its parts (or surfaces) by gradually decreasing the level of global inhibition in a fixed LEGION network; related parts may synchronize at one level of global inhibition but become desynchronized at an increased level. Similarly, Rhouma and Frigui showed that a nested structure of clustering can be formed in their network of spike oscillators [103]. With the ability to select a segment for further analysis as explained in Section V-D, arbitrary hierarchies could be embodied. For example, an embedded tree structure $((AB)(CD))$ could be coded by first forming two assemblies corresponding to (AB) and (CD) , respectively, and then each assembly is selected and further decomposed into two assemblies corresponding to two terminal symbols. This way of representing syntactical structure converts structural complexity into temporal complexity, and time being an infinitely extensible dimension allows the system to have *in principle* unbounded capacity to deal with the compositionality of data structures. This latter property has been argued to be a defining characteristic of symbolic architecture, not shared by connectionist architecture [43]. Furthermore, embedding combinatorial structure in time makes *processing time* a relevant quantity—problem-solving is viewed as a temporal process, and a processing device naturally takes more or less time to solve a particular problem, depending on the difficulty of the problem.

To describe spatial relations between objects in a scene, discussed in Section VII-D, would require that the system have the capability to deal with syntactical structure. So how to compute spatial relations should have significant bearing on the compositionality problem. In a related study on natural language representation, Shastri and Ajjanagadde [115] described the use of oscillatory correlation to dynamically bind arguments and constants in order to perform reasoning with predicates and rules. This is a form of instantiation that binds an abstract slot (say “recipient”) and a specific filler (say “John”); see von der Malsburg [136] for a general discussion on instantiation as an application that can benefit from a solution to the binding problem.

VIII. CONCLUSION: VERSATILE COMPUTING REQUIRES THE TIME DIMENSION

The substrate for diverse mental functions in perception, reasoning, and action is a gigantic network of neurons whose common language is a neuronal signal. The fundamental claim of the temporal (and oscillatory) correlation theory is that binding is manifested in the time structure of such a signal.

If there is one difference that stands out between natural intelligence and artificial intelligence, it is the versatility of the former. Furthermore, natural intelligence emerges from a concrete neural network—an individual brain—whose architecture is more or less fixed after development. As pointed out by von der Malsburg [136], typical practice in neural computation, and artificial intelligence in general, is: “Give me a concrete problem and I will devise a network that solves it.” This is the principle of universality, in the sense of universal Turing machines or multilayer perceptrons. The problem that faces the brain is a rather different one: “Given the concrete network, learn to cope with situations and problems as they arise.” Let me call this the principle of versatility (see also [116]). In other words, the difference between universality and versatility comes down to “first the problem then the network” versus “first the network then the problem.”

How can such a network give rise to versatility as wide-ranging as from sensory response, scene analysis, to language processing and long-term planning? I believe that time provides a necessary dimension for the network to fulfill its various functional requirements. The time dimension is flexible and infinitely extensible—a characteristic not shared by spatial organization of the network, no matter how complex it is.

APPENDIX

ON THE NUMBER OF CONNECTED PATTERNS

For a 1-D image R with n binary pixels, it is easy to see that the number of connected patterns is $n(n+1)/2$. For a 2-D R , the number of connected patterns has been stated to be an exponential relation with respect to $|R|$ ([53, p. xiv], [142, p. 132]). However, I cannot find a proof to this conclusion in the literature, and it is not straightforward to give the precise number of connected patterns on an image of the size $m \times n$, where $m > 1$. Thus, I furnish in this Appendix a proof on a $2 \times n$ figure, and from it to an $n \times n$ image.

Theorem 1: The number of connected patterns on a $2 \times n$ binary image R increases exponentially with $|R|$.

Proof: Let us arrange R in a two-row n -column layout, as shown in Fig. 21(a). Denote the number of connected patterns on such R as $N(n)$. Thus, we have $N(1) = 3$, $N(2) = 13$. For $n > 2$, consider R as formed by appending a $2 \times (n-1)$ image with an additional column to the right of the image [see Fig. 21(a)]. One can divide all connected figures into two sets: those containing no black pixel in column $n-1$ and the remainder. The size of the first set is simply $N(n-2) + N(1)$, counting those to the left of column $n-1$ and those to the right. The second set must have at least one black pixel in column $n-1$ and is determined by possible ways of appending column n .

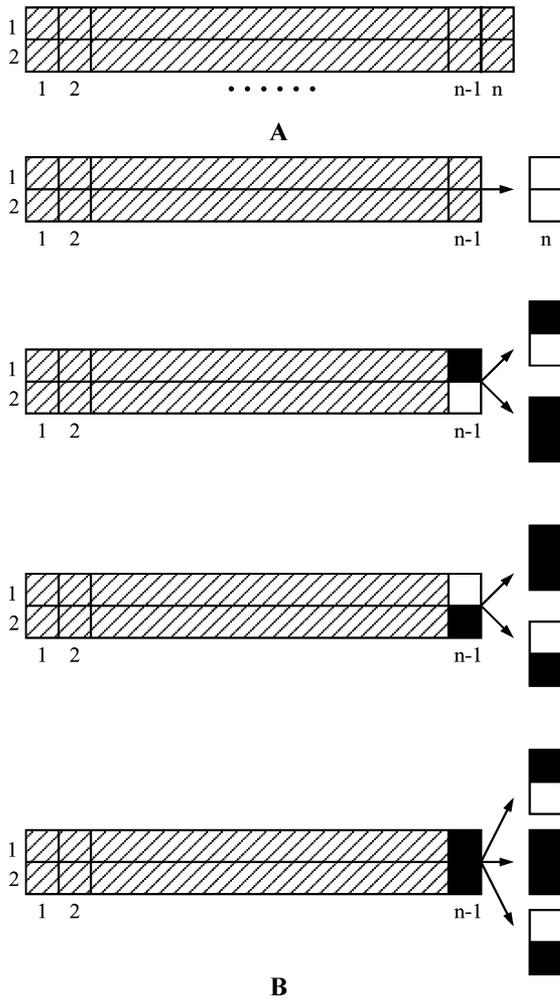


Fig. 21. Counting the number of connected patterns on a $2 \times n$ image. (a) A $2 \times n$ grid. (b) Four different ways of expanding from a $2 \times (n - 1)$ image to include column n .

There are four different ways of expanding to column n , as illustrated in Fig. 21(b). First, it involves no black pixel in column n and the number of such patterns is just $N(n - 1) - N(n - 2)$. Second, it involves expanding when the upper pixel in column $n - 1$ is black and the lower one is white, and this leads to two distinct subsets of connected patterns depending on whether column n has one or two black pixels; see Fig. 21(b). The third way is a symmetrical case when the upper pixel in column $n - 1$ is white but the lower one is black. The fourth way involves expanding when both pixels in column $n - 1$ are black, and this leads to three distinct sets of connected patterns, as shown in Fig. 21(b). Even counting only two such subsets, the total number of connected patterns that involve black pixels in both column $n - 1$ and column n is $2[N(n - 1) - N(n - 2)]$. Thus, we have the inequality

$$\begin{aligned}
 N(n) &> [N(n - 2) + N(1)] + [N(n - 1) - N(n - 2)] \\
 &\quad + 2[N(n - 1) - N(n - 2)] \\
 &= 2N(n - 1) + [N(n - 1) - 2N(n - 2)] + 3 \\
 &> 2N(n - 1) + [N(n - 1) - 2N(n - 2)].
 \end{aligned}$$

Because $N(2) - 2N(1) > 0$, we have the following recurrence inequality:

$$N(n) > 2N(n - 1) \quad (A1)$$

Thus, we have $N(n) > 2^n$. This completes the proof.

Given that all connected patterns on a $2 \times n$ image are also connected patterns on an $n \times n$ image for $n \geq 2$, we have the following corollary.

Corollary 1: The number of connected patterns on an $n \times n$ image R increases exponentially with $|R|$.

ACKNOWLEDGMENT

The author thanks N. Robertson for very helpful discussions concerning the Appendix, G. J. Brown and C. von der Malsburg for insightful comments on an earlier draft, and G. Hu and Y. Shao for assistance in manuscript preparation.

REFERENCES

- [1] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*. New York: Cambridge Univ. Press, 1991.
- [2] S. Amari and M. A. Arbib, "Competition and cooperation in neural nets," in *Systems Neuroscience*, J. Metzler, Ed. New York: Academic, 1977, pp. 119–165.
- [3] H. Ando, T. Morie, M. Nagata, and A. Iwata, "A nonlinear oscillator network for gray-level image segmentation and PWM/PPM circuits for its VLSI implementation," *IEICE Trans. Fund. Electron. Comm. Comp. Sci.*, vol. E83A, pp. 329–336, 2000.
- [4] M. A. Arbib, Ed., *Handbook of Brain Theory and Neural Networks*, 2nd ed. Cambridge, MA: MIT Press, 2003.
- [5] B. J. Baars, *A Cognitive Theory of Consciousness*. London: Cambridge Univ. Press, 1988.
- [6] P. Baldi and R. Meir, "Computing with arrays of coupled oscillators: an application to preattentive texture discrimination," *Neural Comp.*, vol. 2, pp. 458–471, 1990.
- [7] H. B. Barlow, "Single units and cognition: a neurone doctrine for perceptual psychology," *Perception*, vol. 1, pp. 371–394, 1972.
- [8] G. D. A. Barreto, A. F. R. Araujo, and S. C. Kremer, "A taxonomy for spatiotemporal connectionist networks revisited: The unsupervised case," *Neural Comp.*, vol. 15, pp. 1255–1320, 2003.
- [9] D. S. Barth and K. D. MacDonald, "Thalamic modulation of high-frequency oscillating potentials in auditory cortex," *Nature*, vol. 383, pp. 78–81, 1996.
- [10] I. Biederman, "Recognition-by-component: A theory of human image understanding," *Psychol. Rev.*, vol. 94, pp. 115–147, 1987.
- [11] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
- [12] S. M. Bohte, H. L. Poutré, and J. N. Kok, "Unsupervised clustering with spiking neurons by sparse temporal coding and multilayer RBF networks," *IEEE Trans. Neural Netw.*, vol. 13, pp. 426–435, 2002.
- [13] A. S. Bregman, *Auditory Scene Analysis*. Cambridge, MA: MIT Press, 1990.
- [14] M. Brosch, E. Budinger, and H. Scheich, "Stimulus-related gamma oscillations in primate auditory cortex," *J. Neurophysiol.*, vol. 87, pp. 2715–2725, 2002.
- [15] G. J. Brown, "Auditory scene analysis," in *Handbook of Brain Theory and Neural Networks*, 2nd ed, M. A. Arbib, Ed. Cambridge, MA: MIT Press, 2003, pp. 132–135.
- [16] S. Campbell, "Synchrony and desynchrony in neural oscillators," Ph.D. dissertation, Dept. Physics, Ohio State Univ., Columbus, 1997.
- [17] S. Campbell and D. L. Wang, "Synchronization and desynchronization in a network of Wilson-Cowan oscillators," *IEEE Trans. Neural Netw.*, vol. 7, pp. 541–554, 1996.
- [18] S. R. Campbell, D. L. Wang, and C. Jayaprakash, "Synchrony and desynchrony in integrate-and-fire oscillators," *Neural Comp.*, vol. 11, pp. 1595–1619, 1999.
- [19] —, "Synchronization rates in classes of relaxation oscillators," *IEEE Trans. Neural Netw.*, vol. 15, pp. 1027–1038, 2004.

- [20] E. Csemeli and D. L. Wang, "Motion segmentation based on motion/brightness integration and oscillatory correlation," *IEEE Trans. Neural Netw.*, vol. 11, pp. 935–947, 2000.
- [21] —, "Texture segmentation using Gaussian-Markov random fields and neural networks," *IEEE Trans. Neural Netw.*, vol. 12, pp. 394–404, 2001.
- [22] K. Chen and D. L. Wang, "Perceiving geometric patterns: from spirals to inside-outside relations," *IEEE Trans. Neural Netw.*, vol. 12, pp. 1084–1102, 2001.
- [23] K. Chen, D. L. Wang, and X. Liu, "Weight adaptation and oscillatory correlation for image segmentation," *IEEE Trans. Neural Netw.*, vol. 11, pp. 1106–1123, 2000.
- [24] Y. Choe, "Perceptual grouping in a self-organizing map of spiking neurons," Ph.D. Dissertation, Dept. of Computer Sciences, Univ. of Texas at Austin, 2001.
- [25] L. Chua, *CNN: A Paradigm for Complexity*, Singapore: World Scientific, 1998.
- [26] S. Corchs and G. Deco, "A neurodynamical model for selective visual attention using oscillators," *Neural Netw.*, vol. 14, pp. 981–990, 2001.
- [27] J. Cosp and J. Madrenas, "Scene segmentation using neuromorphic oscillatory networks," *IEEE Trans. Neural Netw.*, pp. 1278–1296, 2003.
- [28] J. Cosp, J. Madrenas, and J. Cabestany, "A VLSI implementation of a neuromorphic network for scene segmentation," in *Proc. MicroNeuro*, 1999, pp. 403–408.
- [29] N. Cowan, *Attention and Memory: An Integrated Framework*. New York: Oxford Univ. Press, 1995.
- [30] —, "The magic number 4 in short-term memory: a reconsideration of mental storage capacity," *Behav. Brain Sci.*, vol. 24, pp. 87–185, 2001.
- [31] R. C. deCharms, "Information coding in the cortex by independent or coordinated populations," *Proc. Nat. Acad. Sci. USA*, vol. 95, pp. 15 166–15 168, 1998.
- [32] R. C. deCharms and M. M. Merzenich, "Primary cortical representation of sounds by the coordination of action-potential timing," *Nature*, vol. 381, pp. 610–613, 1996.
- [33] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Ann. Rev. Neurosci.*, vol. 18, pp. 193–222, 1995.
- [34] R. L. Didday, "The simulation and modeling of distributed information processing in the frog visual system," Ph.D. dissertation, Stanford Univ., 1970.
- [35] J. Driver and G. C. Baylis, "Attention and visual object recognition," in *The Attentive Brain*, R. Parasuraman, Ed. Cambridge, MA: MIT Press, 1998, pp. 299–326.
- [36] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [37] E. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex," *Neural Comp.*, vol. 2, pp. 293–307, 1990.
- [38] R. Eckhorn *et al.*, "Coherent oscillations: a mechanism of feature linking in the visual cortex," *Biol. Cybern.*, vol. 60, pp. 121–130, 1988.
- [39] —, "Different types of signal coupling in the visual cortex related to neural mechanisms of associative processing and perception," *IEEE Trans. Neural Netw.*, vol. 15, pp. 1039–1052, 2004.
- [40] A. K. Engel, P. Fries, and W. Singer, "Dynamic predictions: Oscillations and synchrony in top-down processing," *Nature Rev. Neurosci.*, vol. 2, pp. 704–716, 2001.
- [41] G. B. Ermentrout, "Complex dynamics in winner-take-all neural nets with slow inhibition," *Neural Netw.*, vol. 5, pp. 415–431, 1992.
- [42] R. FitzHugh, "Impulses and physiological states in models of nerve membrane," *Biophys. J.*, vol. 1, pp. 445–466, 1961.
- [43] J. A. Fodor and Z. W. Pylyshyn, "Connectionism and cognitive architecture: A critical analysis," *Cognition*, vol. 28, pp. 3–71, 1988.
- [44] W. J. Freeman, "Spatial properties of an EEG event in the olfactory bulb and cortex," *Electroenceph. Clin. Neurophysiol.*, vol. 44, pp. 586–605, 1978.
- [45] A. Gail, H. J. Brinksmeyer, and R. Eckhorn, "Contour decouples gamma activity across texture representation in monkey striate cortex," *Cereb. Cortex*, vol. 10, pp. 840–850, 2000.
- [46] R. Galambos, S. Makeig, and P. J. Talmachoff, "A 40-Hz auditory potential recorded from the human scalp," *Proc. Nat. Acad. Sci. USA*, vol. 78, pp. 2643–2647, 1981.
- [47] W. Gerstner, "Spiking neurons," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. Cambridge, MA: MIT Press, 1999, pp. 1–53.
- [48] G. M. Ghose and J. Maunsell, "Specialized representations in visual cortex: a role for binding?," *Neuron*, vol. 24, pp. 79–85, 1999.
- [49] C. M. Gray, "The temporal correlation hypothesis of visual feature integration: still alive and well," *Neuron*, vol. 24, pp. 31–47, 1999.
- [50] C. M. Gray, P. König, A. K. Engel, and W. Singer, "Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties," *Nature*, vol. 338, pp. 334–337, 1989.
- [51] S. Grossberg, "Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors," *Biol. Cybern.*, vol. 23, pp. 121–134, 1976.
- [52] S. Grossberg and L. Wyse, "A neural network architecture for figure-ground separation of connected scenic figures," *Neural Netw.*, vol. 4, pp. 723–742, 1991.
- [53] G. Hinton and T. J. Sejnowski, Eds., *Unsupervised Learning: Foundations of Neural Computation*. Cambridge, MA: MIT Press, 1999.
- [54] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol. (Lond.)*, vol. 117, pp. 500–544, 1952.
- [55] J. J. Hopfield and A. V. M. Herz, "Rapid local synchronization of action potentials: toward computation with coupled integrate-and-fire oscillator neurons," *Proc. Nat. Acad. Sci. USA*, vol. 92, pp. 6655–6662, 1995.
- [56] D. Horn and I. Opher, "Collective excitation phenomena and their applications," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. Cambridge, MA: MIT Press, 1999, pp. 297–320.
- [57] D. Horn and M. Usher, "Parallel activation of memories in an oscillatory neural network," *Neural Comp.*, vol. 3, pp. 31–44, 1991.
- [58] J. L. Johnson and M. L. Padgett, "PCNN models and applications," *IEEE Trans. Neural Netw.*, vol. 10, pp. 480–498, 1999.
- [59] M. Joliot, U. Ribary, and R. Llinas, "Human oscillatory brain activity near to 40 Hz coexists with cognitive temporal binding," *Proc. Nat. Acad. Sci. USA*, vol. 91, pp. 11 748–11 751, 1994.
- [60] D. Kahneman, A. Treisman, and B. Gibbs, "The reviewing of object files: Object-specific integration of information," *Cognit. Psychol.*, vol. 24, pp. 175–219, 1992.
- [61] Y. Kareev, "Through a narrow window: working memory capacity and the detection of covariation," *Cognition*, vol. 56, pp. 263–269, 1995.
- [62] Y. Kazanovich and R. Borisyuk, "Object selection by an oscillatory neural network," *BioSyst.*, vol. 67, pp. 103–111, 2002.
- [63] J. P. Keener, "Analog circuitry for the van der Pol and FitzHugh-Nagumo equations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 1010–1014, 1983.
- [64] J. M. Kinser and T. Lindblad, "Implementation of pulse-coupled neural networks in a CNAPS environment," *IEEE Trans. Neural Netw.*, vol. 10, pp. 584–590, 1999.
- [65] A. Knoblauch and G. Palm, "Pattern separation and synchronization in spiking associative memories and visual areas," *Neural Netw.*, vol. 14, pp. 763–780, 2001.
- [66] C. Koch and S. Ullman, "Shifts in selective visual attention: Toward the underlying neural circuitry," *Human Neurobiol.*, vol. 4, pp. 219–227, 1985.
- [67] K. Koffka, *Principles of Gestalt Psychology*. New York: Harcourt, 1935.
- [68] P. König and T. B. Schillen, "Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization," *Neural Comp.*, vol. 3, pp. 155–166, 1991.
- [69] G. Kuntimad and H. S. Ranganath, "Perfect image segmentation using pulse coupled neural networks," *IEEE Trans. Neural Netw.*, vol. 10, pp. 591–598, 1999.
- [70] S. H. Lee and R. Blake, "Visual form created solely from temporal structure," *Science*, vol. 284, pp. 1165–1168, 1999.
- [71] U. Leonards, W. Singer, and M. Fahle, "The influence of temporal phase differences on texture segmentation," *Vis. Res.*, vol. 36, pp. 2689–2697, 1996.
- [72] Z. Li, "A neural model of contour integration in the primary visual cortex," *Neural Comp.*, vol. 10, pp. 903–940, 1998.
- [73] J. E. Lisman and M. A. P. Idiart, "Storage of 7 ± 2 short-term memories in oscillatory subcycles," *Science*, vol. 267, pp. 1512–1515, 1995.
- [74] X. Liu, K. Chen, and D. L. Wang, "Extraction of hydrographic regions from remote sensing images using an oscillator network with weight adaptation," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 207–211, 2001.
- [75] X. Liu and D. L. Wang, "Range image segmentation using a relaxation oscillator network," *IEEE Trans. Neural Netw.*, vol. 10, pp. 564–573, 1999.
- [76] R. Llinás and U. Ribary, "Coherent 40-Hz oscillation characterizes dream state in humans," *Proc. Nat. Acad. Sci. USA*, vol. 90, pp. 2078–2082, 1993.
- [77] C. Lourenco, A. Babloyantz, and M. Hougardy, "Pattern segmentation in a binary/analog world: Unsupervised learning versus memory storing," *Neural Netw.*, vol. 13, pp. 71–89, 2000.

- [78] J. N. MacGregor, "Short-term memory capacity: Limitation or optimization?," *Psychol. Rev.*, vol. 94, pp. 107–108, 1987.
- [79] M. Mahowald and R. Douglas, "A silicon neuron," *Nature*, vol. 354, pp. 515–518, 1991.
- [80] P. E. Maldonado and G. L. Gerstein, "Neuronal assembly dynamics in the rat auditory cortex during reorganization induced by intracortical microstimulation," *Exp. Brain Res.*, vol. 112, pp. 431–441, 1996.
- [81] J. L. McClelland and D. E. Rumelhart, Eds., *Parallel Distributed Processing, Vol. 2: Psychological and Biological Models*. Cambridge, MA: MIT Press, 1986, vol. 2.
- [82] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [83] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, pp. 81–97, 1956.
- [84] P. M. Milner, "A model for visual shape recognition," *Psychol. Rev.*, vol. 81, no. 6, pp. 521–535, 1974.
- [85] M. L. Minsky and S. A. Papert, *Perceptrons*. Cambridge, MA: MIT Press, 1969.
- [86] —, *Perceptrons (Expanded Edition)*. Cambridge, MA: MIT Press, 1988.
- [87] R. E. Mirollo and S. H. Strogatz, "Synchronization of pulse coupled biological oscillators," *SIAM J. Appl. Math.*, vol. 50, pp. 1645–1662, 1990.
- [88] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 4th ed. San Diego, CA: Academic, 1997.
- [89] C. Morris and H. Lecar, "Voltage oscillations in the barnacle giant muscle fiber," *Biophys. J.*, vol. 35, pp. 193–213, 1981.
- [90] M. C. Mozer, R. S. Zemel, M. Behrmann, and C. K. I. Williams, "Learning to segment images using dynamic feature binding," *Neural Comp.*, vol. 4, pp. 650–665, 1992.
- [91] J. Nagumo, S. Arimoto, and S. Yoshizawa, "An active pulse transmission line simulating nerve axon," in *Proc. IRE*, vol. 50, 1962, pp. 2061–2070.
- [92] K. Nakayama, Z. J. He, and S. Shimojo, "Visual surface representation: a critical link between lower-level and higher-level vision," in *An Invitation to Cognitive Science*, S. M. Kosslyn and D. N. Osherson, Eds. Cambridge, MA: MIT Press, 1995, pp. 1–70.
- [93] H. Neumann and W. Sepp, "Recurrent V1-V2 interaction in early visual boundary processing," *Biol. Cybern.*, vol. 81, pp. 425–444, 1999.
- [94] E. Niebur and C. Koch, "Control of selective visual attention," in *Advances in Neural Information Processing Systems*, D. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, pp. 802–808.
- [95] E. Niebur, C. Koch, and C. Rosin, "An oscillation-based model for the neuronal basis of attention," *Vis. Res.*, vol. 33, pp. 2789–2802, 1993.
- [96] M. S. Olufsen, M. A. Whittington, M. Camperi, and N. Kopell, "New roles for the gamma rhythm: population tuning and preprocessing for the beta rhythm," *J. Comp. Neurosci.*, vol. 14, pp. 33–54, 2003.
- [97] R. Opara and F. Wörgötter, "A fast and robust cluster update algorithm for image segmentation in spin-lattice models without annealing—visual latencies revisited," *Neural Comp.*, vol. 10, pp. 1547–1566, 1998.
- [98] S. E. Palmer, *Vision Science*. Cambridge, MA: MIT Press, 1999.
- [99] R. Parasuraman, Ed., *The Attentive Brain*. Cambridge, MA: MIT Press, 1998.
- [100] H. E. Pashler, *The Psychology of Attention*. Cambridge, MA: MIT Press, 1998.
- [101] C. S. Peskin, *Mathematical Aspects of Heart Physiology*. New York: NYU Courant Inst. of Mathematical Sciences, 1975.
- [102] A. N. Popper and R. R. Fay, Eds., *The Mammalian Auditory Pathway: Neurophysiology*. New York: Springer-Verlag, 1992.
- [103] M. B. H. Rhouma and H. Frigui, "Self-organization of pulse-coupled oscillators with application to clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 180–195, 2001.
- [104] U. Ribary *et al.*, "Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans," *Proc. Nat. Acad. Sci. USA*, vol. 88, pp. 11 037–11 041, 1991.
- [105] L. C. Robertson, "Binding, spatial attention and perceptual awareness," *Nature Rev. Neurosci.*, vol. 4, pp. 93–102, 2003.
- [106] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, pp. 386–408, 1958.
- [107] —, *Principles of Neural Dynamics*. New York: Spartan, 1962.
- [108] A. L. Roskies, "The binding problem," *Neuron*, vol. 24, pp. 7–9, 1999.
- [109] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1986, pp. 318–362.
- [110] D. E. Rumelhart and J. L. McClelland, Eds., *Parallel Distributed Processing, Vol. 1: Foundations*. Cambridge, MA: MIT Press, 1986, vol. 1.
- [111] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1986, vol. 1, pp. 151–193.
- [112] T. B. Schillen and P. König, "Stimulus-dependent assembly formation of oscillatory responses: II. Desynchronization," *Neural Comp.*, vol. 3, pp. 155–166, 1991.
- [113] T. J. Sejnowski and G. E. Hinton, "Separating figure from ground with a Boltzmann machine," in *Vision, Brain, and Cooperative Computation*, M. A. Arbib and A. R. Hanson, Eds. Cambridge, MA: MIT Press, 1987, pp. 703–724.
- [114] M. N. Shadlen and J. A. Movshon, "Synchrony unbound: a critical evaluation of the temporal binding hypothesis," *Neuron*, vol. 24, pp. 67–77, 1999.
- [115] L. Shastri and V. Ajjanagadde, "From simple associations to systematic reasoning: a connectionist representation of rules, variables and dynamic bindings using temporal synchrony," *Behav. Brain Sci.*, vol. 16, pp. 417–494, 1993.
- [116] W. Singer, "Neuronal synchrony: a versatile code for the definition of relations?," *Neuron*, vol. 24, pp. 49–65, 1999.
- [117] W. Singer and C. M. Gray, "Visual feature integration and the temporal correlation hypothesis," *Ann. Rev. Neurosci.*, vol. 18, pp. 555–586, 1995.
- [118] P. Smolensky, "On the proper treatment of connectionism," *Behav. Brain Sci.*, vol. 11, pp. 1–74, 1988.
- [119] D. Somers and N. Kopell, "Rapid synchrony through fast threshold modulation," *Biol. Cybern.*, vol. 68, pp. 393–407, 1993.
- [120] H. Sompolinsky and M. Tsodyks, "Segmentation by a network of oscillators with stored memories," *Neural Comp.*, vol. 6, no. 4, pp. 642–657, 1994.
- [121] O. Sporns, G. Tononi, and G. M. Edelman, "Modeling perceptual grouping and figure-ground segregation by means of active re-entrant connections," *Proc. Nat. Acad. Sci. USA*, vol. 88, pp. 129–133, 1991.
- [122] R. D. Stewart, I. Fermin, and M. Opper, "Region growing with pulse-coupled neural networks: An alternative to seeded region growing," *IEEE Trans. Neural Netw.*, vol. 13, pp. 1557–1562, 2002.
- [123] M. Stoecker, R. Eckhorn, and H. J. Reitboeck, "Size and position invariant visual representation supports retinotopic maps via selective backward paths: A dynamic second order neural network model for a possible functional role of recurrent connections in the visual cortex," *Neurocomputing*, vol. 17, pp. 111–132, 1997.
- [124] D. Terman and D. L. Wang, "Global competition and local cooperation in a network of neural oscillators," *Phys. D*, vol. 81, pp. 148–176, 1995.
- [125] A. Treisman, "Features and objects in visual processing," *Sci. Amer.*, Nov. 1986.
- [126] —, "Solutions to the binding problem: progress through controversy and convergence," *Neuron*, vol. 24, pp. 105–110, 1999.
- [127] A. Treisman and H. Schmidt, "Illusory conjunctions in the perceptron of objects," *Cognit. Psychol.*, vol. 14, pp. 107–141, 1982.
- [128] S. Ullman, "Visual routines," *Cognition*, vol. 18, pp. 97–159, 1984.
- [129] —, *High-Level Vision*. Cambridge, MA: MIT Press, 1996.
- [130] M. Ursino, G.-E. La Cara, and A. Sarti, "Binding and segmentation of multiple objects through neural oscillators inhibited by contour information," *Biol. Cybern.*, vol. 89, pp. 56–70, 2003.
- [131] W. M. Usrey and R. C. Reid, "Synchronous activity in the visual system," *Ann. Rev. Physiol.*, vol. 61, pp. 435–456, 1999.
- [132] W. R. Uttal, *An Autocorrelation Theory of Form Detection*. Hilldale, NJ: Lawrence Erlbaum, 1975.
- [133] B. van der Pol, "On 'relaxation oscillations'," *Philos. Mag.*, vol. 2, no. 11, pp. 978–992, 1926.
- [134] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: phase synchronization and large-scale integration," *Nature Rev. Neurosci.*, vol. 2, pp. 229–239, 2001.
- [135] C. von der Malsburg, "The correlation theory of brain function," Max-Planck-Institute for Biophysical Chemistry, Internal Rep.81-2, 1981.
- [136] —, "The what and why of binding: the modeler's perspective," *Neuron*, vol. 24, pp. 95–104, 1999.
- [137] C. von der Malsburg and W. Schneider, "A neural cocktail-party processor," *Biol. Cybern.*, vol. 54, pp. 29–40, 1986.
- [138] D. L. Wang, "Pattern recognition: neural networks in perspective," *IEEE Expert*, vol. 8, pp. 52–60, Aug. 1993.
- [139] —, "Emergent synchrony in locally coupled neural oscillators," *IEEE Trans. Neural Netw.*, vol. 6, no. 4, pp. 941–948, 1995.
- [140] —, "Primitive auditory segregation based on oscillatory correlation," *Cognit. Sci.*, vol. 20, pp. 409–456, 1996.

- [141] —, “Object selection based on oscillatory correlation,” *Neural Netw.*, vol. 12, pp. 579–592, 1999.
- [142] —, “On connectedness: a solution based on oscillatory correlation,” *Neural Comp.*, vol. 12, pp. 131–139, 2000.
- [143] —, “Temporal pattern processing,” in *Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA: MIT Press, 2003, pp. 1163–1167.
- [144] D. L. Wang and G. J. Brown, “Separation of speech from interfering sounds based on oscillatory correlation,” *IEEE Trans. Neural Netw.*, vol. 10, pp. 684–697, 1999.
- [145] D. L. Wang, J. Buhmann, and C. von der Malsburg, “Pattern segmentation in associative memory,” *Neural Comp.*, vol. 2, pp. 95–107, 1990.
- [146] D. L. Wang and X. Liu, “Scene analysis by integrating primitive segmentation and associative memory,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, pp. 254–268, 2002.
- [147] D. L. Wang and D. Terman, “Locally excitatory globally inhibitory oscillator networks,” *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 283–286, 1995.
- [148] —, “Image segmentation based on oscillatory correlation,” *Neural Comp.*, vol. 9, pp. 805–836, 1997.
- [149] P. J. Werbos, *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*. New York: Wiley-Interscience, 1994.
- [150] H. Wersing, J. J. Steil, and H. Ritter, “A competitive-layer model for feature binding and sensory segmentation,” *Neural Comp.*, vol. 13, pp. 357–388, 2001.
- [151] M. Wertheimer, “Untersuchungen zur Lehre von der Gestalt, II,” *Psychol. Forsch.*, vol. 4, pp. 301–350, 1923.
- [152] H. R. Wilson and J. D. Cowan, “Excitatory and inhibitory interactions in localized populations of model neurons,” *Biophys. J.*, vol. 12, pp. 1–24, 1972.
- [153] S. N. Wrigley and G. J. Brown, “A computational model of auditory selective attention,” *IEEE Trans. Neural Netw.*, vol. 15, pp. 1151–1163, 2004.
- [154] Z. Wu and A. Guo, “Selective visual attention in a neurocomputational model of phase oscillators,” *Biol. Cybern.*, vol. 80, pp. 205–214, 1999.
- [155] S.-C. Yen and L. H. Finkel, “Salient contour integration by temporal binding in a cortically-based network,” in *Proc. Neural Information Processing Systems (NIPS-96)*, 1997.
- [156] —, “Extraction of perceptually salient contours by striate cortical networks,” *Vis. Res.*, vol. 38, pp. 719–741, 1998.
- [157] B. Yu and L. Zhang, “Pulse-coupled neural networks for contour and motion matchings,” *IEEE Trans. Neural Netw.*, vol. 15, pp. 1186–1201, 2004.
- [158] X. Zhang and A. A. Minai, “Temporally sequenced intelligent block-matching and motion-segmentation using locally coupled networks,” *IEEE Trans. Neural Netw.*, vol. 15, pp. 1202–1214, 2004.
- [159] L. Zhao and E. E. N. Macau, “A network of dynamically coupled chaotic maps for scene segmentation,” *IEEE Trans. Neural Netw.*, vol. 12, pp. 1375–1385, 2001.



DeLiang Wang (M'90–SM'01–F'04) received the B.S. and M.S. degrees from Peking (Beijing) University, Beijing, China, in 1983 and 1986, respectively, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1991, all in computer science.

From July 1986 to December 1987, he was with the Institute of Computing Technology, Academia Sinica, Beijing. Since 1991, he has been with the Department of Computer Science and Engineering and the Center for Cognitive Science, The Ohio State University, Columbus, where he is currently a Professor. From October 1998 to September 1999, he was a Visiting Scholar in the Department of Psychology, Harvard University, Cambridge, MA. His research interests include machine perception and neurodynamics. He is President-elect of the International Neural Network Society (President in 2006).

Prof. Wang received the 1996 U.S. Office of Naval Research Young Investigator Award.