Dimension Change, Coarse Grained Coding and Pattern Recognition in Spatio-Temporal Nonlinear Systems

David DeMaris
demaris@us.ibm.com
IBM Corporation
11400 Burnet Rd
Austin TX 78758 USA

Abstract

Several research programs employing spatio-temporal recurrent dynamics and changes in dimensionality have extended the dialog on neural computation and coding beyond classical frameworks such as feed forward and attractor neural networks and feature detectors. Some have emphasized spiking networks, while others emphasize oscillations and synchronization as the locus of computation and coding. In this paper, the formalism of locally connected homogeneous coupled map lattices is described. Its deployment in an extended version of the dynamical recognizer framework is described, and is compared with density coding, computational mechanics, and liquid state machine frameworks for neural computation. A population coding strategy based on coarse graining the continuous valued distribution of all sites in the lattice is developed and examined as a form of dimension reduction. Results on recognition of 3-D objects are reported. In order to better understand the dynamics supporting recognition, measures suggested by these other research programs and computational frameworks were examined. Dynamics trajectories from object recognition trials were examined for correlation with recognition rates and measures of the distance of the representation space statistics between the target objects and noise initial conditions, and the intrinsic separation between different objects in the set to be classified were performed. These results raise questions about the efficacy of density coding as an explanation for the results, and on the validity of recent criticisms that chaotic systems cannot satisfy separation requirements required for real time computation.

Keywords: population code, coupled map lattice, object recognition, population coding, liquid state machine, computational mechanics, chaotic neural network, synchronization, density coding, symbolic dynamics
Introduction

The combination of a coding scheme, network topology, and computational operators (typically multiplicative scaling on weights and gating by inhibition), can be considered as a computational framework. Such a framework will guide the design of experiments, data analysis, and simulation work. The vast majority of work in neuroscience and neural computation relies on the concept of one or more varieties of neural coding. Diverse perspectives can be found in (5; 21; 17; 44; 1). Localized rate codes were commonly assumed based on the discovery of feature detectors, and underlies the coding of computational units in connectionist models. Along with prevailing assumptions about the locus of coding, there were assumptions about the localization of microcircuits. Specific computational roles were assigned to microcircuits, which performed a task by virtue of specific structure – often in feed forward models. Even though it was known that recurrent (feedback) connections are pervasive at various scales in the nervous system, the difficulty of analysis of such systems has discouraged the use of recurrent networks until recently. Ideas imported from symbolic artificial intelligence and signal processing were more easily mapped onto microcircuits where local state was identified with a rate code.

Microcircuit computational frameworks rely on rate codes and task specific topologies to accomplish biologically relevant computations. In contrast, nonlinear dynamics frameworks have exploited the rich structure of state flows in dynamics of recurrent networks (38). While much of the best known work in neural applications of nonlinear dynamics focuses on motor coordination (51; 25) and olfaction (47), there have been applications in pattern recognition (57) and gestalt grouping processes (54). Other applications of dynamics are more explicitly computational, adapting recurrent networks to function in symbolic domains such as language recognition. Work to apply recurrent networks to language problems has been based on low dimensional systems, in which single state variables are used in the decision process (18; 43; 22).

This paper will be concerned with computational frameworks based on cooperative processes in spatially extended systems. These systems add interactions between spatially extended recurrent units to the possibilities inherent in complex state flows of recurrent systems. Cooperative processes (42) are those in which computation and coding phenomena arises from the interaction of a large number of similar components, with homogeneous dynamics at each component, and local laws of interaction. The symmetries of large scale patterns are not inherent in the local dynamics and are not easily predicted.

One issue with cooperative computations is how to utilize or readout the distributed state from a large regular array or lattice of unit. Some form of dimension reduction is required for to turn this high dimensional state vector into a perceptual discrimination, or into a motor control signal. Cooperative systems exhibit a variety of mechanisms for dimension change beyond the explicit connections to a reduced number of output units typical of connectionist and microcircuit based strategies. These include:
Synchronization processes between local oscillating pools, which reduce the effective dimensionality of oscillating subsystems through coupling and mutual entrainment.

Reduction of dimension via readout pools driven by complex spiking dynamics.

Cooperative interactions reducing the complexity of spatio-temporal patterns, which have been characterized as particle interactions.

Reduction of dimension through coarse graining, in which continuous state variables in many units are clustered into a smaller number of bins as in an inter-spike interval histogram.

These mechanisms will be reviewed in more depth below; one key theme of this paper is to highlight the common ground in all of these approaches, which is a change in effective dimension. Apart from some intrinsic mathematical and physical interest, and as an experimental measurement technique associated with nonlinear dynamics, some investigators have explored functional roles for dimension change

- The selection of a subspace correlated with a percept and its conditioned meaning.
- A sequence of emergent intermediate states in a decision process.
- Projection of inputs into a higher dimensional space, similar to support vector methods.

The topic of dimensionality and synchronization in nonlinear systems is a very rich one, and only a brief treatment is given here to set the stage for discussion of coupled map lattice models (CML.) and their use in pattern recognition computations. A more extensive description is given in (10). In the following discussion, dimension as usual simply means the number of state variables in a dynamical system. Readout dimension refers to some reduced number of variables obtained in a measurement of the full dimension (i.e. all sites or state variables). Effective dimension will refer to conditions where the full set of state variables are correlated due to coupling and partial synchronization; the distribution of states diverges from an expected probability distribution if the variables were uncoupled. (A reduction in effective dimension will have an impact on other formal measures of fractal dimension of the state space, such as correlation dimension, but no use of formal dimension measures is made in this discussion).

The methods section describes a system designated Synchronization Opponent Cooperative Activity (Soca) network, for computing metric similarity of spatial forms and recognition of 3-D objects. The system explicitly incorporated effective dimension changes by a sequence of different bifurcation and coupling parameter epochs. The original motivation for this sequence was the hypothesis that the effective dimension change correlated with respiration investigated by Freeman and colleagues (47) might be replicated in higher levels of the visual system. It was hypothesized that slower (delta-beta frequency) excitatory impulses from sub-cortical or other cortical regions serve to modulate the fast (gamma) waves which carried the state (9). These impulses would temporarily change some bifurcation parameter or the local inter-column coupling dynamics from a task-specific baseline value. It was expected that evolution would select networks with initial epochs resulting in expanding the effective dimension, by spreading out from tightly concentrated intervals of phase space to cover the whole state space,
through spatially distributed interactions. A second set of parameters would concentrate
the allowable states into a pattern-determined subspace via increased synchronization.
Some adaptive learning algorithm would be required to find the exact coupling and
nonlinearity parameters to map a particular set of object worlds to a representation space,
but the general trend in the epochs would correspond to this “synchronization opponent”
dynamics embodied in the name.

However, data reported below indicate that these expected dimension changes are not critical to the computation; the system performs nearly as well with only a single parameter epoch. How, then, do they function? The function appears to be a
combination of recurrent state flows similar to the dynamical recognizer framework, and
dimension expansion and reduction effected through coarse grained coding. Only in
reviewing the range of other spatio-temporal frameworks reviewed below has the
importance of population coding with partition cell statistics as a dimension reduction
method been clarified. Computation takes place in a dimensionality defined by the
continuous state variables of the coupled map formalism and the network, but readout
occurs in a lower dimension controlled by partial synchronization processes, wave
interaction in the spatial network derived from the sampled input, and the number and
width of bins (state space intervals or partition cells) whose population density
constitutes the code. The biological interpretation of this binning is that readout
microcircuits or assemblies are sensitive to inter-spike interval statistics, but with limited
precision.

Methods

This section will first treat in slightly more detail several examples of spatially
extended nonlinear systems and computational techniques which have preceded or
evolved concurrently with the present author’s work. After a comparison of the present
strategy with other investigations, details of the model, learning, and recognition are
presented along with some discussion of dimensionality issues.

Background and Review of Related Systems

One of the pioneering efforts emphasizing oscillations, dimension change, and
neural function beyond the scale of rate coded microcircuits is the experimental and
modeling work of Freeman and colleagues(19). In this work, known as the KIII model,
rate coding is acknowledged at the peripheral sensory and motor subsystems, but
perception and memory are embodied in large scale oscillatory activity. Spike level
coding is important only for input, but dendritic field current patterns - as measured by
the local field potential (LFP) electroencephalogram - is the “output” of the perception
process. The entire history and retained memory of sensory input and conditioning is
retained in global oscillatory patterns. During perception, sensory input controls gain in
local microcircuits to perturb the system and reduce the dimensionality of the global
chaotic attractor, which is referred to as forcing the system into a “wing of the attractor.
The neural correlate of the percept is asserted to be a sequence of spatial amplitude
pattern in the gamma frequency range during the burst.

In a recent summary (20) of this family of models and future directions, the term
biocomplexity models is introduced, and several characteristic features defined:
1. They are high dimensional but not fixed – they are dynamically changing in (effective) dimension.
2. The models are not autonomous, but continuously interact with “environments” selected from infinitely complex world
3. They are adaptive – a network adapts and modifies its internal organization, and changes the external environment to meet the organism’s own goals.
4. Such models are a distributed object evolving in both space and time to meet goals shaped by cumulative experience stored in memory.
5. They are driven and stabilized by noise of internal origin.

This KIII olfactory research program has emphasized experimental work on localized regions of paleocortex and limbic systems, and has been accompanied by complex hierarchical computational models. The recent and expanding body of imaging data suggests that most activity, and certainly higher level cognition and behavior, are performed by large scale interactive networks. The scale of such networks exceeds the scope of computational models thus far attempted in microcircuit frameworks, particularly spiking networks. The interactions and control linking memory, intentional states, and perception involved in behavior to oscillatory phenomena across classical Brodman areas is thus even less well understood and modeled, but recent experimental data has begun to shed light on so-called coordination dynamics (4). For example, analysis of multi-channel local field potentials in an awake behaving monkey found that coherent signals in lower frequency (β, approximately 15-17 Hz) at several prefrontal sites indicate that the sites participated in a synchronized network, and power and coherence in this network were correlated with response time in a discrimination task (26). This synchronization strength also predicts strength and timing of event related components correlated with task performance. Methods have also been developed to assess the causal influence and direction as a function of time on the same base data set (27). These indicate that processing involving striate (V1) and extrastriate (V2-V4) areas may consist of a mixture of feed-forward dynamics from V1 followed by recurrent or feedback influences.

Little investigation and modeling has occurred to date on the functional or computational role of such coordination activity in mixed feed forward and oscillatory recurrent systems; this can only take place in the context of some common understanding on coding and operators. In hopes of developing such an understanding, I will turn now to models which are similar in the use of spatially extended recurrent networks, but which are more explicitly computational in their orientation than the KIII model. While the KIII experimental data and models were seminal in expanding the neuroscience dialog to include oscillations and dimensionality change, their origins in olfaction meant that key issues of interest to psychology and artificial intelligence communities were not addressed, including similarity measures (46) and the problem of invariance in recognition of spatial or temporal patterns.

Another framework for real time computation known as liquid state machines (LSM) recently emerged which also challenges the prevailing view of structural microcircuit functionality and localized rate coding. LSM retains a commitment to
spikes and spike timing as the carriers of relevant state in real time computations (29), in contrast to the local fields (as in the KIII model) and large scale oscillation patterns of coordination dynamics. A key point in the LSM framework is that *enduring states* are dispensed with in the liquid circuit. The liquid circuit is a generic, unstructured microcircuit which is nevertheless crucial to computation. This liquid circuit is perturbed by spiking inputs, possibly with regular spatial structure, and serves to increase the dimensionality of the perturbing signal through recurrent spiking interactions. The LSM dimensionality is equal to the number of neurons in the liquid pools, which are considered as analog state variables in their instantaneous spiking frequencies.

To accomplish a computational task, the transient interactions initiated by a perturbation in the liquid pool are read by readout spiking neuron pools, whose activation (the fraction of spiking neurons in the pool in a 20 ms window) constitutes the analog state variable for some decision or control task. The readout pools undergo supervised learning with the p-delta rule in order to extract information present in the high dimensional state space of all the liquid pool neurons.

The LSM group notes that previous computational frameworks store information about past states. In contrast, within the LSM, no persistent state is maintained – all information about past inputs is decaying. The system can ignore decay of information as long as inputs produce states that are *separable* by the readout pools. Further, there is nothing task specific in the liquid state, in contrast to finite state machines where the persistent states and transitions are designed for a task. Given this flexibility and lack of task commitment, multiple computations can proceed in parallel in the same loosely structured liquid network of generic microcircuits.

The final type of dynamical network to be introduced involves recurrent map networks which use the *coarse graining* principle underlying the field of *symbolic dynamics* (28). Partition cells (intervals in the state space) of the dynamical system are identified with symbols in a formal language. The residence of the system in a particular interval or subspace at some readout time may be used to implement a decision process. Alternatively, some time interval of the dynamical history may be coarse grained to produce a sequence of symbols indicating the intervals are visited. When the number of intervals is smaller than the network state variable dimension, coarse graining of the continuous state variables implies a dimension reduction. To my knowledge, the first advocate for the use of this coding and computational style with chaotic networks was Nicolis (41; 40). However, from a biological standpoint, this is essentially the use of firing rate densities (inter-spike histogram intervals) as a code, which has been suggested as early as 1963 (55).

The *dynamical recognizer* of Pollack was an early demonstration of coarse graining in low dimensional networks. The state a discrete recurrent network is perturbed by an input symbol sequence mapped to intervals in the dynamic state. A genetic algorithm is used to find parameters (weights) such that the resulting dynamics will, with high probability, accept only symbol sequences belonging to a regular language. Acceptance is characterized by the system’s output unit state residing in a particular interval after the last input symbol is applied.

More recently, the present author has used coarse grained interval statistics in spatially extended networks of coupled maps to perform similarity computations, associating the state space intervals with axes of a representation space. Coupled maps
are a discrete space and time recurrent network introduced by Kaneko (24). One topology which has been investigated consists of a regular rectangular array or lattice with nearest neighbor sites coupled; this network is known as a coupled map lattice (CML). Some of the work described here has been briefly reported in conference proceedings. (8) (12) (11).

The occupancy of these intervals at a readout time locates a point in the space. This can be considered as a statistical extension of the dynamical recognizer principle supporting more general categorization. It is also similar to the LSM system in several senses:

1. Dimensionality is expanded (through projection of constrained input state intervals into the full continuous range of allowable network state values) and contracted, (through the coarse graining in the readout process).
2. The computation does not rely on specific task related structure or state variables – the connections and network parameters are regular and homogeneous, except that there is some orientation specificity.
3. Readout of transient states is required – there are no task specific output units or assemblies.
4. Computation is rapid, exploring transient activity rather than waiting for stable attractors to emerge.

There are several points of contrast:

1. Input is applied only as an initial condition, suggesting a sample and iterate paradigm in which recurrent function is combined with feed forward processing from more conventional feature and edge extraction modules. This is not a requirement of CML modeling; other investigators have used input to modulate coupling from baseline values, for example (54).
2. Rather than training readout units, the bifurcation and coupling parameters of the network are adjusted during learning.
3. Readout is an intrinsic response to a histogram of states in the entire lattice, corresponding to the inter-spike interval statistics of lattice sites sensed by a tuned microcircuit with high fan-in. Readout has been assumed to occur at a specific time, rather than continuously (while recognizing the possibility of the latter).
4. The units are recurrent chaotic oscillators, intended to model localized pools of excitatory and inhibitory neurons.
5. Spatial patterns formation and interactions can be observed in the computation, similar to the results in the computational mechanics group. Due to the rapid nature of the computations (constrained to at most 14 iterates) and statistical readout, the role of particle interactions is less prominent in the present work.

Work by Milton and Mackie (34) and independently by this author (10) recently called attention to the rapid (a few iterations) convergence of densities in high
dimensional chaotic systems, such as CMLs. This convergence is similar to that developed analytically for large ensembles of random initial conditions applied to a single low dimensional chaotic map (13); this statistical response of an uncoupled system is the strict definition of a density as understood by dynamicists. The term collapsed density was used to describe the distribution in CML systems (34). Initialized with a uniform random state, the many units of a large CML with homogeneous bifurcation and coupling parameters also show some of the characteristics of the infinite ensemble density, though convergence may be slowed by arbitrary initial conditions. CML evolution was shown experimentally (i.e. numerical simulation) to converge rapidly (5-10) iterations to state distributions either time invariant or of low periodicity. This convergence is visible only in the statistical or coarse grained space, not in time series of individual units. The resulting distributions also shows some sensitivity to spatial patterns embedded in the otherwise uniformly distributed noise initial condition.

The rapid convergence property may play a role in the rapid recognition of biological systems. Since they allow for the possibility of coding and selective influence within Brodman area style modules, coarse grained statistical coding and variations on a “sample and iterate” computational paradigm may preserve a role for function specialization in modular regions of column like computing units. This is in contrast to the classical dynamics approach as seen in the KIII network, which apparently conceives of neural dynamics as more broadly integrated, with sensory systems perturbing a unified high dimensional attractor across broad regions of cortex subsuming multiple modules.

**Coupled Map Lattice Formalism**

This report presents new results from the ongoing study of oscillations in coupled map lattices (CML) used for the recognition of three dimensional objects. CML neural networks are a discrete space, discrete time model in the tradition of oscillatory neuronal ensemble models (56), where each unit (lattice site) stands for the collective behavior of a mixed pool of excitatory and inhibitory neurons. The entire network or lattice is a homogeneous spatial ensemble of locally connected column-like units. By homogeneous, I mean that the parameters are the same at each site.

The discrete time update nature is not intended to imply a global clock for the biological network, but simply indicates that a mixed pool can be modeled as series of temporal windows imposed on the continuous rate; a transfer function mapping the rate at time t to the rate at time t+1 can then be defined. The time variable plays a role like the temporal windows used to derive ensemble rates from repeated single unit measurements, not like the time step in an ordinary differential equation simulation.

Early studies of randomly connected threshold firing models resulted a single humped curve for input and output ensemble rate transfer functions of such mixed pools, with thresholds and the ratio of excitatory and inhibitory elements serving as a bifurcation parameter, raising and lowering the curve (2). The logistic map described below approximates this behavior. Figure (1) below indicates the recurrent nature of each site, the mixed pool underlying each site, and the regular lattice arrangement of the model.
Figure 1.  a) Each lattice site is a recurrent network mapping the ensemble average firing rate at \( t \) to the rate at \( t+1 \), according to equations 1 and 2 below.  b) A mixed excitatory inhibitory network underlying the dynamics in each node.  c) The lattice consists of such maps connected to nearest neighbors by a coupling function; a fraction of neighboring states is added to each site; high coupling values will eventually lead to totally synchronized behavior in small laps.  Some orientation sensitivity is implied by the lack of connections to diagonal neighbors. Information on initial conditions or perturbations of increasingly distant sites reaches each node over time, resulting in an interference pattern when regular spatial forms are iterated in the lattice.

Previously it was reported that such systems could be trained, via a genetic algorithm, to map various viewpoints of paperclip objects to approximately the same distribution at some readout time(12). Recognition rates of 85% were obtained with complete training on a set of 20 objects using a set of recognizer CML trained for each object, and choosing as the matching object the recognizer producing the nearest matching distribution to the characteristic distribution for the object. Important features of the network and population coding method include rapid processing (under 14 iterations), ease of learning, translation invariance, and generalization between outlines and filled objects.

The short range diffusive or local mean field coupled logistic map was used. The term diffusive is historical (reflecting the use of CML for fluid dynamics), but should be interpreted as synchronization between local pools for neural systems.

In the locally coupled system, each unit is connected to nearest neighbors. The computation is divided into two steps; first a diffusive coupling step is applied:

\[
S_d(x,y) = (1 - c)S_t(x,y) + \frac{c}{4}[S_t(x,y + 1) + S_t(x,y - 1) + S_t(x + 1,y) + S_t(x - 1,y)]
\]

\{1\}

where \( S_d \) is a temporary array holding the state of the diffusion computation, \( t \) is the current time step, \( x \) and \( y \) are the spatial indices of the lattice site \( S \) at the center of the connection neighborhood, \( S \) is the continuous state variable at each pixel of the array restricted to the range \([-1.0 \text{ to } 1.0]\), and \( c \) is the coupling constant restricted to the range
The factor $1-c$ is a squashing function, preventing the sum of the surround from exceeding the stable range of the map.

The second computational unit applied in each time step is the logistic map. This dynamical equation has the underlying single humped parabolic curve format, corresponding to the mixed pool firing rate transfer function described above:

$$S_{t+1}(x,y) = 1 - bS_d(x,y)^2$$

(2)

where $S$, $t$, $x$, and $y$ are as above and where $b$ is the bifurcation parameter, restricted to the interval $[0.0, 2.0]$. Initial and intermediate states $S$ are restricted to the interval $[-1.0 < 1.0]$.

Readers familiar with the physics CML literature will note that I have departed from the widely used notation of $\alpha$ for the nonlinearity or bifurcation parameter and $\varepsilon$ for coupling. This is motivated by two concerns. First, when discussing CML models in the context of brain oscillations and pattern recognition based on representation space concepts, these symbols have other common uses. Second, I want to encourage thinking of bifurcation and coupling as operators in computations, which network designers consciously exploit, and which experimentalist should look for in their data or in formulating experiments.

During learning and recognition, the logistic dynamics are applied as a parameter epoch at every lattice site defined by the triplet $\{b, c, t\}$. Two parameter epochs are applied sequentially during training and recognition, with the parameters $b$ and $c$ changing instantaneously at time $t_1$ and $t_2$. This is interpreted in biological terms as a periodic modulation of the local and short range synchronization dynamics in a cortical column and its neighborhood, induced by bursting behavior from long range excitatory connections from memory areas which control the computation. Compared with constant parameters, these modulations have been demonstrated to provide small gains in both recognition rate and average recognition time (See table 1 below).

During evolutionary search for the parameter sets, each epoch is constrained to 7 iterations. This upper bound was motivated by biological recognition time constraints of around 150 ms. One can interpret each iteration as corresponding to a half cycle at 40Hz, with the state values interpreted spike density within a particular cycle, and with the parameter changes associated with slower (5-20 Hz) modulations of the gamma dynamics. The total iterations required to create the representation space across all classifiers ranges from 6 to 14 iterations, with a mean of 12.7 across 50 learning trials. For the network learning strategy described next, a learning trial in this network refers to learning parameters and a coarse grained signature for all objects via separate genetic search which maps all provided views to the same point in the representation space.

### Dimension and Synchronization

Synchronization between a high dimensional time series, such as the state sequence of sites in a coupled network, admits many definitions. A network is defined as a system of $N$ variables $S_i$ with some coupling between state evolution. Total synchronization is defined as the case in which $|S_i - S_j| \rightarrow 0, t \rightarrow \infty$. The system of $N$ variables is effectively one dimensional when totally synchronized. This is the only rigorously defined case, with all other situations defined as some variety of partial synchronization. One definition of partial synchronization addresses a situation where
some units obey the synchronization condition above while others do not. The network may be characterized by clusters $C_1, C_2, \ldots C_n < N$, where $N$ is the number of units. These may be spatially segregated into contiguous oscillating domains.

The dynamics of the CML system are, by this definition, operating in a regime of partial synchronization, but for large networks full synchronization or clustering occurs after a lengthy transient regime; computational studies commonly begin clustering measures after 10000 iterates.

**Recognition**

The goal of the 3-D recognition task is to recognize one of a family of similar paperclip objects, described in more detail below. Seven images of each object are provided, seen from different viewpoints. The objects are generated in a 3-D modeling system and are rotated in depth, i.e. they are rotated about the z axis.

After generating classifiers and corresponding signatures by the learning process described below, recognition of the objects is implemented by simply providing each target view of an object as an initial condition to a CML with the parameter set \{b1, c1, t1, b1, c2, t2\} defining each classifier. To map an image into the continuous space, a binary threshold is applied to the original source gray scaled image, and the lattice sites for black pixels are assigned initial state 0.0, white pixels are assigned 0.9999. Since the range of each lattice site is [-1, 1], the value zero can be considered the background firing rate of the ensemble, and 1 as the maximum firing rate. Applying the dynamics results in an evolving spatial wave front orthogonal to the object boundary.

The configuration of the lattice is read out after $t_1 + t_2$ steps and a histogram is generated; 64 bins of equal width were used in this study. This is the coarse graining step. The bin with the largest occupancy is zeroed (corresponding to the region outside the object boundary and the dynamics evolution wave front).

All classifiers are applied in parallel, then the Euclidean distance from the signature for each classifier is computed. The object is assigned to the classifier which resulted in the minimum distance to the original signature.

This signature, obtained during the learning process described below, is the mean occupancy for each partition cell across all training views of the corresponding object corresponding used for training. The number of partition cells (64 in the current implementation) is an implicit parameter, and implies a dimension reduction from the continuous states of each lattice site. For example, the 6400 dimensions of an 80x80 lattice are reduced to 64. This is the maximum possible dynamical dimensionality, which may be further reduced by the synchronization process.

**Learning**

To produce such a recognizer which can capture the differences in orientation statistics in the family of images, we employ an evolutionary computing strategy. Since this learning strategy is not claimed as biologically relevant, the details are omitted in this paper. Briefly, the training views of each object are presented to a classifier 6-tuplet parameterized CML, and the resulting coarse grained signature is evaluated according to the objective function defined below. This process is repeated for each of 100 population members in each generation, with crossover and mutation operators applied to the surviving parameter sets. In contrast to typical practice, the generated network is not
tested on the recognition task, but is evaluated on several abstract measures on the signature (distribution over partition cells) in an objective function. This makes the learning more like biological learning conditions, in which objects are encountered sequentially and not necessarily in the context of a specific task. In other recent 3-D object recognition work achieving similar or higher performance levels, choices of learning parameters (Gaussian centers in radial basis function (RBF) networks ([14], weighting coefficients on statistics [33]) are determined with a global knowledge of object statistics.

In the present work, a cross entropy metric described below is used to ensure even coverage of the representation space. If one were to interpret this or construct it with dynamical neural activity, the correct strategy would seem to be to employ anti-Hebbian learning, with a coarse grained population code of objects in long-term memory steering the emerging invariant population code for newly learned objects to sparsely occupied subspaces.

The objective function used in learning must balance two goals: 1) tight clustering of different views of an object and 2) separation of classes corresponding to various views of other objects. A view normalization strategy (adapted from the radial basis function approach to view based recognition [15]) serves the tight clustering goal, with a cross-entropy term comparing current output with previously stored representations enforcing separation of objects in the representation space.

The objective function returning a fitness \( f \), with low values indicating higher fitness, takes the form

\[
 f = W_d D - W_e (H_c + H_s) + P_{synch} \tag{3}
\]

where \( W_d \) is the inter-view distance or normalization weight and \( W_e \) is the entropy weight; \( D \), the inter-view distance sum for \( j \) views is

\[
 D = \sum_j D_j = \sqrt{\sum_j \left( \sum_{p=1}^k (v_{i,p} - v_{j,p}) \right)^2} \tag{4}
\]

where \( v_p \) are the occupancies of \( k \) partition cells for each of the \( j \) views.

\( H_c \) is the cross entropy or Kullback-Lieber information measure between the current reference view distribution \( C \) and the database of \( N \) object distributions (signatures) \( S \) with \( k \) bins:

\[
 H_c = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k C_j \log_2 \frac{C_j}{S_{i,j}} \tag{5}
\]

\( H_s \), the Shannon entropy of the current signature with \( k \) bins is

\[
 H_s = \sum_{i=1}^k S_i \log_2 S_i \tag{6}
\]

and

\[
 P_{synch} = \begin{cases} 1000 & \text{if } \max(S_i) < \text{synchPenThresh} \\ 0 & \text{otherwise} \end{cases} \tag{7} \]
The parameter synchPenThresh was empirically determined as .15 after a series of experiments reported in (10).

**Recognition and Readout: Theory**

While a brief algorithmic description of recognition process was given above, I wish to give a more theoretical picture and historical background, and to situate the methods in context. The problem of learning to accept positive exemplars of a language while rejecting negative exemplars is known as language induction. Classical machine learning approaches to this problem construct a finite state automaton to achieve recognition. Formally, a finite state recognizer is a quadruple \( \{Q, \Sigma, \delta, F\} \), where \( Q \) is a set of states (with \( q_0 \) denoting the initial state), \( \Sigma \) is some finite alphabet, \( \delta \) is a transition function mapping \( Q \times \Sigma \to Q \), and \( F \subset Q \) is a set of final or accepting states. A string of tokens from alphabet \( \Sigma \) is accepted by the recognizers if, starting from initial state \( q_0 \), the sequence of state transitions indicated by the tokens in the string ends up in one of the final states in subset \( F \).

Pollack made an early attempt to formulate the language induction problem as a dynamical system in the form of a recurrent neural network (43). The dynamical recognizer is a quadruple \( \{Z, \Sigma, \Omega, G\} \), where \( Z \subset \mathbb{R}^k \) is a state space; \( z_k(0) \) is the initial condition. \( \Sigma \) is the input “alphabet”, where a particular closed interval in \( Z \) corresponds to each element in this alphabet. \( \Omega \) is the dynamic, a sequence of transformations \( \omega_i: Z \to Z \) (one for each token) with an associated set of dynamical parameters; these parameters are fixed for a particular recognizer during the induction (training) process. \( G(Z) \to \{0,1\} \) is the decision function which maps one or more states in the sequence produced by the dynamic to an accept / reject decision. In Pollack’s work, only the final state and token are used in the decision function. Within this general framework, the dynamics and decision function are normally much weaker in computational power than a Turning machine.

Pollack notes that \( G \) may be generalized to a graded function indicating “fuzzy” acceptance, or could return a more complex categorization or representation. The Soca network and recognition method is quite consistent with this extended dynamical recognizer framework. A key difference is that the Soca net operates on an entire spatial configuration in parallel (thus the state space has higher dimensionality \( \mathbb{R}^N \), where \( N \) is the number of pixels or sampled image elements), and the tokens are used only once as the initial state. Such a parallel recognizer framework for *picture languages* with constrained states and transformations was introduced by Rosenfeld (45). In contrast to the string based dynamical recognizer, the transition function at each site is now a function of surrounding sites in some spatial neighborhood.

The decision function is necessarily modified by this larger state space. Rosenfeld proposed several possibilities:

- every spatial element reaches an accepting state
- any element reaches an accepting state
- one particular spatial element reaches an accepting state.

In the Soca network and recognition strategy, my approach was not to simply reach an accepting state but to form a metric representation space. The dimensions of the
space are defined by the coarse grained partitions, and the location of a particular representation is derived from ensemble statistics (i.e. the histogram of the entire CML) measured instantaneously during a high dimensional, parallel dynamics conditioned by the spatial arrangement of input feature responses. These statistics naturally support an acceptance function; simply define some threshold distance for each classifier, and accept an object as an instance of language L if it satisfies this distance test. The distances might vary by class, depending on the cluster density of that class in the representation space. Another contribution of the Soca network, then, is adding another type of decision function to the repertoire defined by Rosenfeld. While such a distance threshold decision function is common in statistical pattern recognition, the use of coarse grained statistics considered as a representing space is novel for dynamics based approaches.

Other researchers have recently been concerned with decision functions over spatial patterns processed by cellular automata (CA), a form of spatially-extended dynamical systems closely related to those used in the present work ((37),(36),(23)). CA typically use boolean states and transition functions, or integer valued states and simple numerical transition rules. Genetic algorithms were used to generate and test particular one dimensional cellular automata (CA) which decide, for example, whether a random initial condition has majority ones or zeros. The investigators then examined space-time plots (i.e. plots of successive iterations of a 1-D spatial array) of the resulting successful computations and developed an explanatory framework based on physical metaphors; this computational framework is designated as Computational Mechanics.

Computational mechanics seeks to describe the computations embedded in high dimensional space-time behavior in terms of regular domains, particles, and particle interactions. Regular domains are regions visible in space-time plots consisting of words (spatial configurations) in the same regular language, i.e. regions that are computationally homogeneous. Particles are localized boundaries between such domains; they serve as information carriers. Collisions between particles are the loci of information processing. This processing can be conceived in terms of operators such as decay of one particle to many, reactions (state transitions between language domains at collision sites), and annihilations (the disappearance of an interface as one language domain dominates future spatial evolution at a collision site). The computational strategy can then be expressed in the more concise language of particles and their interactions, substituting for a more verbose description in the language of CA rule lookup tables and raw spatial configurations.

While the computational mechanics group does not explicitly state this, the decision function in the majority task can be considered a type of synchronization. Until all cells reach the same language domain (which, in this simple case, is all 0 or 1) the system is undecided. Consider a k-block, defined as a set of k adjacent cells. The entire CA state consists of overlapping sets of such k-blocks. Each k-block of cells in this automaton is defined by the state transition graph of a finite state automaton (FSA) with k+2 states. As the automata evolves, at each time step we can count for each k-word the fraction of k-blocks which currently hold that value (we say they occupy the state). Synchronization in this context implies that over time, the occupancy statistics in this k-word space converge to sharp peaks, or an unchanging sequence of sharp peaks. We could also consider a graph of k-words, with transition probabilities between them. Over time, particular sub-graphs of this state-transition graph are active, while others become
blocked, as their predecessor states become unreachable within increasing spatial “territories”.

Note that particles have a characteristic velocity, and for certain kinds of terminating conditions (such as a particular site or region reaching a value in a set of accepting states \( F \)) one possibility for variance in the temporal processing is dependence on the emergent particle velocities on initial configurations in a family of inputs, when a “synchronization” decision function is reached.

In summary, the Soca system extends the formalism of dynamical language recognizers to spatial configurations, and attempts to unify this approach with traditional metric representation space of statistical pattern recognition. The decision functions used here also involve *subspace synchronization* in the sense defined above, but it is not limited to contiguous regions as in the regular domains. Another key distinction of the Soca work from the research in the computational mechanics group is that the search strategy is constrained to solve the decision problem in a bounded number of iterations, rather than an open ended synchronization process.

**Results**

**Recognition Rates**

Much recent experimental visual psychology and neuroscience research on the on 3-D object recognition has worked with a family of objects commonly referred to as paper clips. For continuity with such research, I performed network simulations on a set of paperclip objects previously studied in psychophysics. The image set was designed in the visual psychology lab of M. Tarr, and consists of 39 paper clip objects, with seven views provided for each object rotated in depth. Each object is a chain of 5 cylinders, with a variable joint angle connecting each pair. The views are separated by 30°, ranging from –90° to 90°. The objects were originally developed to answer questions regarding the recognition by components or *geon* theory of Biederman (3). The set consists of four complexity groups with 0, 1, 3 or 5 unique geons substituted at some position in the chain.

Similar paperclip objects (corresponding to the low complexity set) were used in human psychophysics experiments (6). In these experiments, subjects were trained with motion sequences of 2-D views, giving an impression of a 3-D object through the kinetic depth effect. In a two-alternative forced choice task on their object set, with single static views of a target or distractor, the miss rate (failure to indicate a match when the target was shown) averaged 30%, indicating that the task is rather difficult. The Soca network exceeded this rate when emulating a two alternative choice paradigm, but most of the set consists of higher complexity objects which were found to result in higher recognition rates.

---

1 One object in the last group was duplicated in the original set, hence 39 rather than an even 40.
Figure 2. Two silhouette views of the same paperclip object, illustrating the extreme nature of the viewpoint variation due to rotation in depth. The left view is the 0°, the right view is +90°. The raw paperclip image set was provided courtesy of Michael Tarr, Dept. of Psychology, Brown University.

The specific object set used here was also used in a study by Tarr and colleagues attempting to discriminate between view-based and structural theories (50). In that study no training period was provided; subjects simply had to judge whether two views shown briefly (200 and 100 ms, separated by a mask stimulus) were the same or different. Under these conditions, the baseline set of shapes (all tubes with no geons inserted) were essentially not recognizable by subjects...
psychophysical findings. Peak recognition rates for training on all seven views ranged from 60% for 39 objects to 88% for 20 objects. During recognition trials, all views of every object were presented to compute error rates; the following table summarizes the results in the 20 object case with all seven training views provided. The recognition process in this case does not measure generalization of each classifier, but rather the tightness of tuning of the classifiers; i.e. whether similar views of other objects will mismap into the representation space.
Table I. Comparison of CML Recognition with 20 objects
For Stationary and Two Parameter Epochs

<table>
<thead>
<tr>
<th></th>
<th>Average Rec. rate</th>
<th>Maximum Rec. rate</th>
<th>Average Time (iterations)</th>
<th>Time variance</th>
<th>Learning Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary</td>
<td>75.2</td>
<td>82.9</td>
<td>13.4</td>
<td>1.39</td>
<td>50</td>
</tr>
<tr>
<td>Two epochs</td>
<td>78.8</td>
<td>87.9</td>
<td>12.7</td>
<td>2.7</td>
<td>51</td>
</tr>
</tbody>
</table>

Distance from Noise Response

A further series of computational studies was undertaken based on the resulting large set of network parameter 6-tuples generated by these learning trials, with a goal of improved understanding of underlying network operating principles. While the network is understood from a computer science standpoint as a generalization of the dynamical recognizer framework, a better physical understanding might give insight into the variations in recognition performance, and could lead to performance improvements within the genetic algorithm learning framework, or might suggest more direct adaptive learning methods.

During recognition, a particular network characterized by a parameter 6-tuple \( \{b_1, c_1, s_1, b_2, c_2, s_2\} \) must achieve two goals – to map diverse views of an object to a point in the space whose axes are defined by state space intervals, and to map other objects to different points in space. Differential rates in the approach to statistical equilibrium of the trajectories derived from the target object and from other stimulus classes might serve the latter. It is known the state occupancy statistics of a large set of uncoupled maps at fully developed chaos may reach a stationary distribution in only a few iterations (13); with coupling or other nonlinearity conditions, the systems may not reach a stationary distribution so rapidly, but rather reach a state of statistical periodicity, in which a cycle of low period is observed in the population statistics. The ensemble evolution will depend on the stimulus, which might be applied as an initial condition or coupled over time.

The first set of measurements reported was formulated to test this hypothesis that the network’s ability to reach a unique subspace for a particular object is predicated on an anomalous time course in an approach to statistical equilibrium.

I measured the Euclidean distance between the characteristic distribution produced by each object’s recognizer CML applied to all target views and the distribution created by the same recognizer’s parameters applied to a “noise” initial configuration: a 300x300 matrix of states uniformly distributed over the interval \([-1,1]\). This distance was computed and averaged separately for correct and error object presentations in a given recognition trial, and the ratio between distances of matching and non-matching trials was evaluated against the error rate. For 50 trials, no significant correlations were found \( r = .13 \).

This negative result suggests that the density coding hypothesis of by Milton and Mackey (34) may not be applicable to highly structured initial conditions, where an
upstream line extraction subsystem is presumed to present initial conditions along object boundaries constrained to small regions of state space (i.e. highly active or at background). However, their calling attention to the density evolution phenomena and the rapid approach to equilibrium is welcome and closely related to the present work. Instead, the deeper underlying mechanism for separation of classes may be similar to the Markov relaxation and information bottleneck framework recently described (52). In Markov relaxation, clustering of stimulus configurations into classes is supported by the differential rates of mutual information loss, with stimuli in the same class reaching temporary plateaus in loss at a characteristic time. The present process proceeds orders of magnitude more rapidly, presumably due to the rapid state mixing of chaotic systems.

**Sum of Configuration Entropies**

A well known measure of dynamical systems is the information dimension, closely related to Shannon entropy and the capacity dimension of a dynamical system (39). Normally, this measure is defined for infinite times, and practically measured for a long time series for a low dimensional system. Given the transient nature of the processing in the present system, a different but related measure is developed.

The measure employed here is the sum of the Shannon information for each configuration (lattice state at iteration i, computed over partition cells in the usual manner for information dimension: \( H_i \), the Shannon entropy of the current signature (occupancy distribution at the readout time of a recognizer) with \( k \) bins is

\[
H_s = \sum_{i=1}^{k} S_i \log_2 S_i
\]

and the measure I designate as sum of configuration entropies is

\[
H_c = \sum_{i=1}^{t} H_s = \sum_{i=1}^{t} H_s
\]

where \( t \) is the total number of iterates, in this case over two parameter epochs \( t_1 + t_2 \).

It is conjectured that the reason for the slightly higher recognition rate of the two epoch dynamics is attributable to a higher entropy obtainable by increasing the reachable states, increasing the chance of finding dynamical trajectories leading to the common subspace from different views, and of diverting undesired configurations and partial trajectories (of distractor memories) away from the subspace corresponding to the best match.

In Fig. (3) below, a least squares fit of this measure is plotted against recognition rate for a set of 16 two epoch trials, exhibiting a low-moderate positive correlation (\( r=0.28 \)). In Fig. (4) the ratio of this measure for correct recognition trials / error trials within each learning trial, in order to assess whether this entropy measure contributes to class discrimination. The correlation is negative, again with a low-moderate correlation (\( r=0.33 \)). This inverted relationship - with matching views having slightly lower entropy than error views - suggests that one error mechanism occurs when misidentified views spread out in state space during recognition dynamics. Successful recognition is
enhanced by limited configuration entropy for distractor objects, indicating that local chaos must be balanced by the synchronizing tendencies of coupling.

Figure 3. Total sum of configuration entropy over all iterations vs. recognition rate for 16 two parameter epoch trials of the CML recognition system applied to 20 paperclip objects.
Figure 4. Ratio of sum of configuration entropy measure for Match / Non-matching recognition trials is plotted against recognition rate for 16 two epoch learning trials of the CML recognition system applied to 20 paperclip objects.

**Separation of Input Conditions in CML Classifiers**

In this section, I present some results on separation of states for different views presented to the recognizer ensembles, inspired by state vector separation studies performed by Maas and colleagues for the liquid state machine model (29). They applied input spike trains $u$ and $v$, with known L2 norm distance between them, to the liquid ensemble with unstructured but diverse connectivity and weights. For each such pair, they then measured L2 norm distances between vectors of integrated firing rates of all neurons in the pool. These stimulus-reponse trials were compared to a baseline separation derived from applying the same spike train with a different random initial state. Since readout in the LSM system may occur continually from the time of stimulus application, it is appropriate to measure separation over the entire time course of the simulation.

The LSM systems differs from the CML system in coupling of inputs, in the timing of readout, and in the use of a parallel ensemble of networks rather than a single liquid pool. In the CML system, a sample and iterate computational style is used; input state configurations are applied to the lattice as an initial condition and iterated for up to 14 steps prior to readout. In addition, readout in LSM samples from a single network, while the readout in the current Soca system is from an ensemble of individual CML classifiers. Within the limitations for comparison due to system differences, some preliminary findings on state vector separation in the CML model are presented, with
distances computed at the readout time between configurations produced by all the classifiers in the ensemble.

Recognition is achieved in the Soca CML system by applying a target object as an initial condition to an ensemble of classifiers, each of which learns to map different training views of an object near the same point in a space formed by partition cells sampled at readout time after a dynamical evolution. Each such application is termed a recognition trial. A learning trial consists of a set of target views applied to the classifier ensembles generated by a particular run of the genetic learning, where the history and population encoding of each classifier influences subsequent encodings.

Some idea of the separation of inputs can be obtained by examining the difference between states generated by all classifiers during each recognition trial. The minimum, median, and maximum distances between the output of each classifier in each recognition trial were measured and summed across all presented object views in each such learning trial; the correlations between these measures and recognition rate were then computed. These distances and measures were performed for two different spaces – the coarse grained occupancies actually used for recognition, and the “raw” state configurations with spatial units corresponding across classifiers. The computations were performed on a set of 50 learning trials, representing 7000 recognition trials. Only single parameter epoch learning trials were computed.
Table II. Correlations between distance measures and recognition rate
(50 learning trials)

<table>
<thead>
<tr>
<th></th>
<th>Next Min</th>
<th>Max</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarse grained (64 bins)</td>
<td>0.2509</td>
<td>0.3271</td>
<td>0.1025</td>
</tr>
<tr>
<td>Continuous</td>
<td>0.2200</td>
<td>0.1265</td>
<td>0.1759</td>
</tr>
</tbody>
</table>

The first observation is that no strong correlation (i.e. r > .5) of recognition rate to any of these measures is found. Given that learning explicitly attempts to cover the coarse grained space at readout time and that recognition rates varies within a relatively small range (66-84%), this may not be too surprising.

Figure 5. The distances to the next nearest neighbor in the coarse grained space for each recognition trial are summed and plotted against the recognition rates for an entire learning trial; a learning trial consists of a complete set of classifiers for each of 20 objects.
Perhaps the most interesting finding is that the sign of correlation of recognition rates with summed distances to the next nearest neighbor is opposite between the coarse grained space and the continuous CML state space (see Figs. 5 and 6). While correlation is weak to moderate, the trends for the dependence of recognition rate on the distance spread between classifier responses are in opposite directions for continuous state values (like those measured in LSM) and the coarse grained values used for classification. Specifically, the recognition rate improves with greater L2 norm distance spread in the coarse grained space, but decreases slightly with distance spread in the continuous state space. This suggests that chaotic processes degrade performance in the continuous realm but not necessarily when coarse graining is taken into account. For another measure, the sum maximum distance of any classifier to each target view, both continuous and coarse grained state are negatively correlated, but the continuous state more so.
Discussion

The method presented in the previous section was motivated by work in biology, computer vision, and the theory of computation in complex systems. I now consider the limitations of this study and opportunities for improvement from the perspective of biological systems, and some connections with specific anatomical regions, measurements, and analysis techniques in hopes of facilitating dialog with experimentalists.

Admittedly, the Soca framework is completely hybrid in character, combining oscillations with algorithmic computations on stored memory; it has a long way to go to be a complete biological theory of invariance in vision. Yet there are intriguing correspondences between this computational framework and aspects of biological systems absent in connectionist systems. Major distinguishing aspects of the Soca network approach from classical connectionist models are as follows:

1. A column like recurrent neuronal group of mixed excitatory and inhibitory neurons is taken as the basic functional unit, with macrostate variables representing the ensemble firing rate. Computations and representations are formed by collective measurements (distributions) over spatially uniform lattices of these units.

2. The typical connectionist unit transfer functions (sigmoidal activation and threshold output) is replaced with a non-monotonic, highly nonlinear (chaotic) function. There is no threshold, since the state variable represents some collective measure such as ensemble frequency. Connections between lattice sites are not excitatory or inhibitory, but serve to construct intrinsic state flows in the space of the recurrent network dynamics; these can be tailored to the statistics of input configurations. In Marr’s hierarchy of processing levels (32), the algorithmic level involves spatio-temporal cooperative processes ordering these state flows.

3. The code for a family of input patterns (i.e. outline shapes) is a coarse grained sample of the states in the entire network at a specific point in time, rather than activation of an optimal unit (place codes), a connectionist sparse or distributed output layer, or a recurrent network in a static (fixed point) attractor. The code is computed via orientation sensitive coupling processes operating on stimulus-linked aperiodic oscillations with local cooperative interactions.

4. This computation effectively performs a dimension reduction by coarse graining states, after first projecting from a constrained subspace (i.e. the sampled pattern with strong excitation or background values, arranged in spatial configurations) into higher dimensional spaces (the continuous range of each lattice site). This projection is hypothesized to occur by relaxing subspace constraints created by more peripheral areas of the visual system, after a sample time in an ongoing sample and iterate process. In this higher dimensional space, the orientation sensitive coupling process transforms spatial information into a population codes reflecting higher order co-occurrence statistics of the input stimulus. Note the implication is that a key role of early ventral system layers is to set up these subspace constraints, through edge extraction mechanisms.

5. The network is time varying. These changes in the dynamics are interpreted as corresponding with major signal bands and modulations as observed in
local field potential EEG and multi-unit neuron correlation studies. For the most part, connectionist accounts of cognitive processing make little contact with medium and large scale electrophysiology; rather, they focus on the level of neurons, while stressing that these may stand for groups. In contrast, the present work investigates computational processes involving temporal changes in coupling (to modulate synchronization) and in bifurcation parameters (to modulating synchronization rates and the underlying unstable periodic structure of chaotic oscillations, which governs the distribution). This emphasis is motivated by observed changes in coherence and spectral sharpening at various space and time scales of biological neural systems.

While these attributes are all unusual in neural modeling which attempts to address the perceptual or psychological level, there are many grounds for criticism in terms of biological realism. The arbitrary starting, stopping and injecting an image as an initial condition into the network is problematic. Without abandoning the CML formalism, realism would be improved by an input stage to induce a perturbation of an ongoing state vector, by increasing coupling parameters or changing the bifurcation state directly. Other assumptions, such as homogeneity of lattice sites and absence of longer range connections are easily relaxed.

In a biological network, the sampling of a transient process (as performed in readout and recognition here) can not be the end result, but must be a starting point for further encoding in long term memory, and for matching against some version in working memory during search tasks. If Soca-like computation and coding occurs in the IT complex (or related form-processing areas like V4) it is must work in conjunction with memory formation networks and comparison networks. Soca-like networks might play roles in view interpolation, invariant recognition, and similarity mapping. Memory formation might be handled by recurrent attractor networks, or by high dimensional complex networks which collapse to subspaces similar to the KIII model.

Since the population code of Soca does not lead to asymptotic spatially localized activity in certain units, some alternative explanation is required to explain the observations of Tanaka and coworkers on spatial localization in anterior IT, in which a small set of activated regions for specific feature combinations, with some small displacements of the spots by changing stimuli (48; 49). One possible explanation for those results would re-envision a role for locally coded, combination code column “units” as proposed by Tanaka. Rather than being selective for feature combinations, such units might instead be responding to specific ensemble frequency combinations (i.e. inter-spike interval distributions) in the transformed signal, which might be computed upstream in the TOE, pIT or V4 regions. This detection could be on the basis of recognizing specific temporal patterns, such as a preferred set of ensemble frequencies occurring in a short post-stimulus time window, or more distributed response where a local column responds to the ensemble frequency change in the last two fast cycles at the optimal readout time. The latter columns might then be linked by learning procedures. The role of such activated regions is confused by uncertainty over whether sub-regions of IT represent the locus of a feed-forward representation network, or an area of comparison of incoming dynamics with memory representation by unknown processes.

Is there a biological equivalent of coarse graining? This paper stresses the importance of coarse graining as a dimension reduction technique in high dimensional systems, in contrast to the initial identification of coarse graining as a coding element in
dynamical computing frameworks by Nicolis, and its previous deployment in low dimensional recurrent systems for formal language recognition. The successful use of coarse graining in invariant pattern recognition suggests that studies in how microcircuits might achieve coarse grained readout and resonant responses to specific densities should be explored. There are indications at the spike level that the range of fluctuations in inputs to model ensembles serves as a spike timing sensitivity parameter (53). This suggests another potential computational role for the slower frequency LFP coherence changes: they could function to controlling sensitivity of readout assemblies. This could occur in conjunction with controlling readout timing, as coherence increase might simultaneously push readout assemblies closer to threshold at the same time their sensitivity properties were changed. However, it is difficult to combine the evidence from LFP and spike levels, given the predominance of sub-threshold contributions to the LFP signals.

Does this work have anything to say about modulation and coordination dynamics? The chosen task of recognizing the equivalence of differing viewpoints is emblematic of task-specific, modular memory coupled computations, where the role of memory may consist of programming a general network topology with an implicit network of state flows between dynamical partition cells. Coordination dynamics between memory and recurrent modular computational regions may involve such “programming” of task specific, modular computations and possibly also the control of readout windows, sensitizing readout assemblies to perform their tasks at specific times in concert with the dynamics evolution to solve invariance problems.

A few remarks on limitations of the system and future work are in order. If viewed from the perspective of practical computer vision and visual psychology, the study has many limitations. I will simply list these; more detail and suggestions for improvements are contained in a dissertation (10).

• The system cannot distinguish widely separated discontiguous points, or line collections with common first order statistics.
• The system does not handle scale invariance well.
• Ecological realism is lacking; there is no behavior, or requirement to behave while learning.
• Testing has been performed on a limited range of objects.
• The system is limited to recognition of isolated objects, rather than conjunctions of objects or scenes.
• Recognition was degraded substantially for 40 objects.

The degradation of recognition with noise added to CML sites has not yet been investigated, but has been an issue for low dimensional dynamical recognizers (7; 30). It is possible that given the reliance on density-like evolution, noise will produce less degradation than in low dimensional systems.

It would strike most biologists and neural net investigators as more biologically plausible to achieve good recognition in a unified network, rather than a family of classifiers. The latter seems to resemble largely discredited template based pattern recognition; however the work of Edelman and colleagues has shown that template-like classifiers can build an effective representation space spanning the space between templates (16). Attempting to find a dynamics which would compute invariance in a
A unified network with the dynamical recognizer framework is an interesting problem, but the nature of learning and recognition dynamics is not obvious. Initial attempts to construct such a dynamics with locally connected homogeneous CML were not successful.

Regarding the current representation strategy, changing the partition size to finer granularity or using non-uniform intervals (bin widths) should be studied. Non-uniform intervals may better match the density of state occupancy in partitions shared by classifiers, increasing the recognition performance by separating those classes. The number, width and regular spacing of the partitions have been fixed in the Soca CML investigations to date. This could be a design parameter of importance, particularly as larger scale networks are built to explore coordination dynamics between CMLs.

With the exception of the KIII network, the spatio-temporal computing frameworks described here involve some form of supervised learning to solve invariance problems in spatial or temporal domains. One open question is how can such networks reach adaptive states for more generalized categorization problems in a more self-organized fashion. Can the desirable aspects of invariance and representation spaces be preserved?

**Conclusion**

From the perspective of pattern recognition, a method involving locally coupled, discrete time nonlinear oscillator arrays was demonstrated to form a partition cell representation space representation of objects, with partition cells in the dynamical phase space serving as the dimensions of the space. Clustering of objects with similar structure in a representation space emerges from learning to normalize views. The formalism of dynamical recognizers was extended from low dimensional to high dimensional dynamics.

This spatio-temporal style of computing a representation contributes to resolving the classic dilemmas for representing local features, feature conjunction, and binding, while affording rapid recognition. The method was demonstrated to reach recognition rates as high as 88% in a nearest neighbor match scenario with 20 objects. Cooperative interactions of transient trajectories in nonlinear oscillator support view based recognition with recognition rates comparable to a recent feed-forward model (14), and a statistical model with a rich input feature space (33), when the training conditions and limitations to spatial features only are taken into account (10).

*Non-stationary* dynamics outperform a single preferred dynamical regime by a small margin. At the outset of this work, it was hoped that a single dynamical stage might be found which solved the similarity problem for many objects, but with the present architecture and learning methods no such universal classifier was found. However, it is intriguing that an evolutionary scenario of improving performance with subtle improvements in non-stationary control is suggested. Measures of the sum of configuration entropy suggest that increasing the number of visited and interacting states is important. This seems to confirm the insight from the LSM group that having high dimension supported by random weights and structure gives better performance, but demonstrates that it can be effective even when non-uniformity in the basic network structure is minimal.
The lack of correlation between performance and the distance between states reached from noise initial configurations and the native objects for each classifier suggests that the recent proposal of density coding does not fully explain the underlying dynamical mechanism of the present system. That proposal is close in spirit to the present work due to the use of coarse graining and CML dynamics; the use of noisy backgrounds suggests that some pattern discrimination can still occur in highly noised systems. Further study in this area is anticipated.

The LSM work distinguishes itself from previous recurrent network paradigms such as dynamical recognizers, in part because these seem to require un-biological clocking mechanisms. Also it is claimed that, particularly when the dynamics are chaotic, the networks are not robust in the face of noise. However, the latter criticism appears to be based on the assumption that the system is low dimensional, or that state values at the micro-circuit level are used for readout as in LSM. While virtually no work has been done on noisy versions of such systems, it is possible that the rapid expansion and mixing of states that result in rapid convergence to period densities identified by Milton and Mackie (34) will reduce the effects of noise when coarse grained readout is performed on the continuous high dimensional state.

After examining the recognition dynamics via noise response, configuration entropy, and separation measures, the nature of the processing in the homogeneous CML seems best conceived of as a parallel, distributed stochastic switching network in the state space traversed by the recurrent spatio-temporal dynamics. It is tempting to describe it as a Markov chain, with the states corresponding to the partitions and the probabilities, but the conditions under which a chaotic flow satisfies the strict definition of a Markov chain are limited (40). I know of no similar analysis for maps but it is likely to be the case as well.

Progress on issues such as invariant perception is important, but ultimately neural computation frameworks should hope to scale up to more complex interactions between memory, attention and behavior characteristic of cognitive neuroscience. Many investigators in the nonlinear dynamics school have been explicitly and vigorously anti-computational, suggesting that the underlying concepts of state, coding, and representation have systematically mislead researchers. State transitions are seen as rapid and global switches to different large scale oscillatory regimes associated with perceptual or cognitive states, but otherwise localized concepts of state and coding are to be avoided.

Other researchers in psychology and computation have argued that both dynamics and computational or representational approaches can and must be taken into account. Mitchell (35) argues that adaptive behavior can only be supported in dynamical frameworks. when functional, information bearing components can be identified. However, she stresses that these need not be localized, structural or persistently coded; in the spatially extended cellular automata models studied by the computational mechanics group, representations are not static, passive, or symbolic – they are dynamic, active, and numerically describable, with descriptors like “particle velocities” correlated with classification performance and the mean time to reach a decision. In the Soca network, because of the bounded iteration times, high dimension, and loss of spatial information through coarse graining, it is more difficult to decompose the state flows into such discernible intermediate objects. Instead, I have sought to identify mechanisms rooted in statistical behavior - such as pattern dependent temporal approaches to invariants, or
trends in separation which correlate with recognition rate. Strong correlations with performance have been elusive.

Markman and Dietrich (31) argue that representation can and should be updated to account for the perspective of perceptual symbol systems, situated action, embodied cognition and dynamical systems. They note that all approaches agree on the need for mediating states, but classical assumptions such as enduring states, symbolic states, amodal states, and independence from sensory and motor systems of the agent are not characteristic of representation in general. Both Mitchell and Markman suggest that dynamics is chiefly oriented toward real time action and may be unable to scale up to cognitive tasks involving reasoning about the past. If issues of robustness in the face of noise are resolved, the principles of coarse grained population coding and representation spaces introduced in this work could offer one avenue of approach to such problems.

The research program described here bridges biologically inspired phenomena (spatio-temporal oscillations, population coding, synchronization, dimension changes mediated via interactions at multiple frequencies) with symbolic and representational constructs (symbolic dynamics, recognizers, and a novel form of representation spaces). Classical concepts in similarity studies, such as metric spaces, may be reconstructed on a dynamical foundation. This work provides one example of a possible reconciliation between the dynamics of biocomplexity models and classical symbolic representational computational strategies.

Like the LSM computational paradigm, little is needed in the way of microcircuit structure to produce results such as invariant pattern recognition; the control of generic, column like computational units by projected signals constructs “virtual” networks in which state flows perform computation and steer the global continuous state to coarse grained subspaces which support representations spaces, symbols, or motor behavioral sequences.

Since the large space of possible networks employing these phenomena is only now being investigated, there is reason to be hopeful for progress in scaling up to the problems addressed by integrative and cognitive neuroscience. The coupled map formalism allows such research to be conducted at low simulation cost relative to spiking networks, and many underlying architectural principles could appear at both spike level dynamics and interactions between modules. The basic issues of dimension change addressed by all the computational frameworks discussed are still relatively novel in neural computing; experimentalists might benefit from considering the issues raised and the way in which they may interact with cognitive level constructs and with task performance measures.

Acknowledgements: G. Werner, C. Peck and referees have provided useful comments on earlier drafts.

References


