THE DYNAMIC LINK ARCHITECTURE

Christoph von der Malsburg
Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum, Germany
Tel +49 (234) 32-27997, Fax 32-14210
E-mail: malsburg@ruhr-uni-bochum.de

and

Depts. for Computer Science, Biology and Physics
University of Southern California
University Park, Los Angeles, CA 90089-2520
Tel +1 (213) 740-9224, Fax -1317
E-mail: malsburg@usc.edu
INTRODUCTION: THE ARCHITECTURE

The dynamic laws governing the brain’s physical elements and their interaction enable it to fall into functionally useful states. The term “neural architecture” is taken here as referring to the shape of these dynamic laws. This article presents and discusses the “dynamic link architecture” (von der Malsburg, 1981, 1985, 1986). There are various ways in which the DLA has been couched in terms of equations (von der Malsburg, 1985; von der Malsburg and Schneider, 1986; von der Malsburg and Bienenstock, 1987; Bienenstock and von der Malsburg, 1987; Wiskott and von der Malsburg, 1996; Zhu and von der Malsburg, 2001). As the DLA has not received a canonical mathematical description yet it is described here in abstract verbal terms. The DLA is a construction site, and this article is an invitation to work at it.

According to the DLA, the brain’s data structure has the form of graphs, composed of nodes (called “units”) connected by links. The graphs of the DLA are dynamic: both units and links bear activity variables changing on the rapid functional time scale of fractions of a second. Graphs form a very versatile data format which probably is able to render the structure of any mental object. A particularly important feature is the ability of graphs to compose more complex data structures from simpler ones, an important requirement for the expression of cognitive structures (COMPOSITIONALITY IN NEURAL SYSTEMS).

The units of the DLA play the role of symbolic elements. This follows the tradition of associating neurons with elementary meaning (the identification of units with neurons is, however, not taken for granted here, see below). Units are endowed with structured signals changing in time. These signals can be evaluated under two aspects, intensity and correlation. Intensity measures the degree to which a unit is active in a given time interval, signifying the degree to which the meaning of the unit is alive in the mind of the animal. Correlations, on the other hand, quantify the degree to which the signal of one unit is related to that of others. The general idea is that identical signal patterns are strongly correlated, whereas statistically independent signal patterns have zero correlation. A correlation can be a binary relation, characterizing two units, or an n-ary relation, to be evaluated for n units at a time.

The strength of links can change on two time scales, represented by two variables called temporary weight and permanent weight. The permanent weight corresponds to the usual synaptic weight, can change on the slow time scale of learning, and represents permanent memory. The temporary weight is constrained to the interval between zero and the permanent weight and can change on the same time scale as the unit activity (hence the name dynamic links).

Dynamic links constitute the glue by which higher data structures are built up from more elementary ones. Conversely, the absence of links (temporary or permanent) keeps mental objects separate from each other and prevents their direct interaction. In the simplest case, a link binds a descriptor to an object. For example, a link may bind a unit representing a color to another unit that stands for a specific object. More generally, mental objects are formed by binding together units representing constituent parts. The infinite richness and
flexibility of the mind is thus made possible as a combinatorial game. The mental activity of familiar objects (like my grandmother, or a yellow volkswagen) may be reliably correlated with the activity of specialized units, but these objects still acquire their substance — their imagined visual appearance and so on — by the dynamical binding of appropriately structured arrays of other units. Units can be part of different functional contexts. They are integrated into a specific one by the activation of appropriate links. Dynamic links are the means by which the brain specializes its circuit diagram to the needs of the particular situation at hand.

Graph Dynamics

Under the influence of signal exchange, graphs and their units and links are subject to dynamic change, constituting a game of network self-organization (SELF-ORGANIZATION IN THE BRAIN). The dynamic links have a resting strength near the value of the permanent weight. When the units connected by a permanent link become active, there is rapid feed-back between the units’ signal correlations and the link’s strength, a strong link tending to increase signal correlation, a strong correlation controlling the link to grow in strength toward the maximum set by the permanent weight. This feed-back can also lead to a downward spiral, weak correlation reducing a link’s strength, a weak link losing its grip on signals which, under the influence of other links, drift apart toward lower correlation. Thus, links between active units tend to be driven toward one of their extreme values, zero or the maximum set by the permanent weight.

Links are subject to divergent and convergent competition: links converging on one unit compete with each other for strength, as do links diverging from one unit. This competition drives graphs to sparsity. Links are also subject to co-operation. Several links carrying correlated signal structure cooperate in imposing that signal structure on a common target unit, helping them all to grow. As the ultimate cause for all signal structure is random, correlations can only be generated on the basis of common origin of pathways. Thus, co-operation runs between pathways that start at one point and converge to another point. The common origin of converging pathways may, of course, be an event or a pattern in the environment.

Co-operation and competition conspire to favor certain graph structures. These are distinguished by being sparse (that is, activating relatively few of the permanent links in or out of units) and by having a large number of co-operative meshes — arrangements of alternate pathways from one source to one target unit. Beyond these statements, a general characterization of graph attractor states is an open issue. However, there are certain known graph structures that have been shown in simulations to be attractor states and that prove to be very useful, see the section on applications. All of these graph structures may be characterized as “topological graphs”: If their units are mapped appropriately into a low-dimensional “display space” (one- or two-dimensional in the known examples), the links of those graphs all run between units that are neighbors in the display space.

Slow Plasticity
In classical neural architectures, learning is modeled by synaptic plasticity: the change of permanent synaptic weights under the control of neural signals. This general idea is also part of the dynamic link architecture. However, the DLA imposes a further refinement in that a permanent weight grows only when the corresponding dynamic link has converged to its maximum strength, which happens only in the context of an organized graph structure. For a permanent link to grow it is thus not sufficient for the two connected units to have high intensity in the same brain state, but their signals must be correlated and their link must be active. This puts the extra condition on the growth of permanent connection weights that they be validated by indirect evidence, in the form of active indirect pathways between the units connected, and in the form of relative freedom from competition, the two conditions characterizing a well-structured dynamic graph. Thus, only the very few connections that are significant in this sense can grow.

**NEURAL IMPLEMENTATION OF DYNAMIC LINKS**

How can the units, links and dynamical rules of the DLA be identified with known neural structures? This is possible in several ways. It will turn out that to some extent the DLA can be seen as a fair interpretation of known structures, whereas there flow also some experimental predictions from it.

**Units are Individual Neurons**

On the most fundamental level, units are to be identified with neurons, links with axons and synapses, signals with neural spike trains, and permanent weights with conventional synaptic strengths. Signal intensity is evaluated as firing rate, averaged over intervals of length \( \Delta \), whereas the stochastic signal fine structure within that interval is evaluated in terms of correlations with a resolution time \( \tau \), two spikes arriving within \( \tau \) of each other being counted as simultaneous. The smallest reasonable choice for \( \Delta \) is probably 100 msec or a little less, the smallest choice for \( \tau \) may be 3 msec, as proposed in (von der Malsburg, 1981). Neural signals in the cerebral cortex have a very rich stochastic structure on all time scales, much of which is not correlated strongly with external stimuli in neurophysiological experiments (and is usually suppressed by averaging in a post-stimulus time histogram).

A point of contention at the present time is the precision with which nervous tissue can process temporal signal structure. Some authors (Shadlen and Movshon, 1999) believe that fine temporal structure cannot be transmitted by neurons and that meaningful signal correlations cannot be extracted. The proposed argument is, however, circular, as it was assumed that neural input signals are random and independent. If this assumption is violated in the brain, the argument falls flat. Indeed it has been shown (Mainen and Sejnowski, 1995) that spike timing of cortical neurons can be reliable with 1 msec precision if neural input is sufficiently structured, and similarly precise spike timing was found in response to temporally structured visual input seven synaptic generations behind the retina (Bair and Koch, 1995). The latter and other studies would encourage the assumption of a \( \tau \) of 1 msec. A heated discussion has also sprung up around the status of the interpretation of signal correlations in terms of dynamic binding (SYNCHRONIZATION, BINDING AND EXPECTANCY) as proposed in (von der Malsburg, 1981), see (Shadlen and Movshon,
1999; Gray, 1999; Singer, 1999; von der Malsburg, 1999) and other articles in the same issue of “Neuron.” At the present time, the issue is the subject of intensive experimental study in many laboratories.

Dynamic links are realized on the single neuron level as rapid reversible synaptic plasticity (RRP). Starting from a resting value, the temporary weight of a synapse is increased by correlations between the pre- and postsynaptic signals, and is decreased if both signals are active in a given time period but are not correlated. The resting weight of a synapse is probably not too far from the maximum set by the permanent weight (so that RRP will manifest itself mainly in the form of rapid weight reduction). The interactions between temporary synaptic strength and signals is such as to constitute a positive feed-back loop. Changes in temporary synaptic weights must take place on a fast time scale to be of functional significance, possibly as quickly as within 10 msec. In the prolonged absence of presynaptic or postsynaptic activity, the temporary weight rises or falls back toward its resting value, with a time scale that corresponds to short-term memory (perhaps a few dozen seconds), or it is reset by an active mechanism (for example, in the visual cortex during saccades). Convergent synaptic competition (competition between synapses at the same postsynaptic neuron) could be implemented by the signals arriving on one synapse or one set of synapses spoiling the postsynaptic activity for others. Divergent competition could be implemented with the help of inhibition between the target cells, making it difficult or impossible to synchronize them all with the same presynaptic signal.

The existence of rapid reversible changes in synaptic strength in cortex is a broadly documented experimental fact (see Hempel et al., 2000 for examples and a review of the experimental literature and TEMPORAL DYNAMICS OF BIOLOGICAL SYNAPSES). What has not been investigated experimentally in any detail is the dependence of rapid synaptic change on postsynaptic signals, and without such study the type of control postulated in RRP cannot be ascertained. There are many open details which must be determined experimentally. Among them is the identity of the relevant postsynaptic signal (membrane potential, some second messenger, e.g., Ca++, action potential or other) and the precise definition of the dynamics of synaptic strength (which could require a delay between the presynaptic and postsynaptic signals, as described for long-term potentiation in Senn et al., 2001, and experimental work reviewed therein).

As was pointed out, implementation of the DLA on the single-neuron level can be realized on a hierarchy of time scales $\Delta$ and the concomitant resolution time $\tau$. So far I have discussed the faster end of the hierarchy. If $\Delta$ is taken to be a large fraction of a second or longer, we are in the domain of overt attention and the well-studied phenomena of the mind shifting context sequentially on smaller and larger time scales. There is no doubt that a very important function of attention is keeping topics separate if their simultaneous activation would lead to confusion, and thus to provide temporal binding. A proper understanding of the mechanisms of attention will have to provide an answer to the question how the focus of attention is formed. Part of the answer will be, of course, that it must unite elements that have something to do with each other (as recorded by the links between them), and not to activate simultaneously what would lead to confusion. The conceptual
framework of the DLA and its network self-organization is therefore appropriate for the description of attention dynamic.

**Multi-Cellular Units**

Just as the DLA interpretation of neural dynamics can be applied at different temporal scales, it can also be applied at different spatial scales, by either identifying units with single neurons, as above, or by identifying them with groups of neurons. In this perspective, all individual neurons in a group, called multi-cellular unit (MCU), are interpreted to have the same meaning. They differ, however, in the synaptic connections they have to neurons in other MCUs. Signal intensity of a MCU is the combined neural activity of all of its neurons. Signal correlations, however, are calculated by paying attention to the distribution of activity in MCUs and determining the combined synaptic weights of all connections between currently active cells in a pair of MCUs. By changing the distribution of activity over its neurons, a MCU can control the connectivity pattern it has to other MCUs. An important example of MCUs are hypercolumns in visual cortex. All neurons in a hypercolumn have the same theme, sub-serving, by definition, one point in visual space. The neurons differ, however, in how they are connected with afferent neurons (which gives them different meaning on a more fine-grained level), and how they are connected to neurons in other hypercolumns.

MCU implementation of the DLA differs in important points from the single neuron implementation. The variables that constitute dynamic links are not temporary synaptic weights but neural spike activity, and correlations are not computed by time-consuming temporal integration over pairs of neural signals but by the instantaneous and parallel evaluation of the signals and connection weights of all active neurons in the MCUs involved. In consequence, with MCUs there is a much greater capacity to express highly structured graphs than in the single-cell implementation. The price for this greater power is much reduced flexibility, because appropriately specialized connectivity patterns within and between MCUs must first be installed.

Implementations of the DLA on different temporal and spatial scales are not mutually exclusive and are probably realized concurrently in our brain. The single cell version is indispensable due to its great flexibility and the absence of any need for specialized pre-existing connectivity structures, but is limited in its capacity to distinguish detailed link structures in limited time. The MCU version is very powerful, may be seen as just an unconventional view of networks in classic architecture, but requires highly specialized connectivity structures and appropriately tuned activity dynamics.

**APPLICATIONS**

Aim and purpose of the DLA is to serve as framework for understanding brain functions. The conventional neural network architecture, lacking the equivalent of dynamic binding, may be a universal medium for realizing individual functions when they are defined ahead of time (such that appropriate combination-coding neurons and connectivity patterns can be defined and binding ambiguities avoided), but in decades of modeling attempts this
architecture has shown itself as too narrow to go beyond elementary functions. The DLA has the full functional repertoire of the conventional neural network architecture, but goes beyond it in being able to build up structured objects, have them interact in a structured way or keep them from interfering. The full potential of the DLA is far from realized, but a certain collection of applications have already been modeled, as briefly reviewed in (von der Malsburg, 1999).

Figure ground segmentation in visual scenes or other modalities is most naturally modeled in the DLA (VISUAL SCENE SEGMENTATION). So far, most concrete models have employed temporal signal correlations to bind all elements of a figure together and to keep them separate from elements belonging to the ground. For this type of models there is experimental evidence, as reviewed in (Gray, 1999; Singer, 1999). Also MCU implementations have been realized, in which each unit is subdivided into sub-populations, one for figure, one for ground, and in a final state all units belonging to the figure restrict their activity to the “figure” neurons, all units in the ground just activate their “ground” neurons.

Many mental objects are met first as sensory arrays of local features. They are most naturally handled, stored and recognized if the neighborhood relations between features are expressed as bindings and stored and retrieved as dynamic links. This has been realized as “Dynamic Link Matching” for the purpose of invariant visual object recognition (reviewed in von der Malsburg, 1999 and FACE RECOGNITION, NEUROPHYSIOLOGY AND NEURAL TECHNOLOGY). Dynamic Link Matching, implemented in terms of temporal binding, has rightly been criticized as too slow to account for object recognition in adults. However, a recent implementation employing direct interaction between links (to be implemented with the help of multi-cellular units, for instance) was shown to be very fast, requiring only one or a few iterations (Zhu and von der Malsburg, 2001).

As to the potential of the DLA for modeling brain function and cognitive processes the cited applications are but the tip of the iceberg. Processing and learning the syntactical structure of natural language has proved very difficult to impossible on the basis of the conventional neural architecture. The reason is that the flexibility to analyze or to form novel sentences requires dynamic binding. It is particularly important here to realize the general process of instantiation, in which an abstract syntactical structure is applied to a concrete set of elements. Instantiation requires the manipulation of dynamic links between abstract roles and concrete role fillers, and requires the recognition of structural relations between abstract structures and concrete instances. Both of these functions are not part of the repertoire of conventional neural networks.

REFERENCES


