Chapter 7

Critical Dynamics of Neural Networks with Spatially Localised Connections

Robert W. Kentridge

7.1 Psychology and time-dependent information processing

People and animals live in a changing world, yet often time and change are eliminated from psychological experiment and from the study of neural networks for reasons of simplicity. The stimulus in experimental psychology is typically unrealistically stationary (Gibson, 1966), while analytically tractable neural networks mainly deal with mapping static input patterns onto static output patterns. The psychological processes through which organisms deal with the changing world nevertheless often require some form of time-dependent dynamic information processing in which the output of a process is dependent not only on its current input, but also on its previous inputs and outputs (see e.g., Neisser, 1976). There is a fundamental equivalence between treating the inputs to a system in the context of its current output and processing sequences of inputs as time-dependent stimuli. In the former case the output at any time is dependent on the current input and the immediately preceding state of the system (and hence its output), but this state is in turn dependent on the input to and state of the system an instant previously and so on ad infinitum. The output of the context-dependent system can in fact be viewed as being dependent on the entire sequence of all its prior inputs. A few examples starting with "low-level" processes and proceeding "upwards" serve to underline the importance of time- and context-dependent processes in psychology. Sensory-motor co-ordination, at least in pursuit as opposed to
ballistic movements, requires sensory inputs to be treated in the context of the current output of the system as can be seen from the classic work of von Holst and Mittlestadt (1950, see Gallistel, 1975) on efference copy or Robinson’s (1965; Robinson, Gordon & Gordon, 1985) studies of feedback in smooth pursuit eye movements. Stimuli which are ambiguous when stationary can often be identified easily when they move — examples range from the separation of objects from their backgrounds through motion parallax, through the kinetic depth effect (Wallach & O’Connell, 1953) and the more general use of optic flow generated both by observer and object movement (see e.g., Gibson, 1966), extending to the use of gait cues to distinguish characteristics such as people’s gender (Cutting, Profit & Kozlowski, 1978) or their effort in lifting weights (Runeson & Frykholm, 1981). Navigation and map building are more explicitly driven by the preceding context — animals navigate mazes not by learning a fixed sequence of turns but by learning each turn in its individual context (O’Keefe & Nadel, 1978). These capabilities can be demonstrated in isolation in various context-driven rule-learning tasks (see e.g., Hirsh, 1980; and Rawlins, 1985). Our ability to deal with the recursive structure of language is clear evidence of very sophisticated information processing which must of necessity be driven both by inputs and outputs of a system (Chomsky, 1963). Finally, there is considerable evidence (e.g., Mullholland, Pellegrino & Glaser, 1980) that certain types of thought (e.g., goal-directed problem-solving, Newell & Simon, 1972) are explicitly serial processes in which each step in solving the problem is undertaken in the context of the current partial solution. In summary, if we want neural network theories to address psychological issues in general, and in particular “higher” faculties such as language, thought and reasoning, then an attempt to understand neurodynamics is surely called for.

7.2 Why study neural networks?

Having concluded that a number of psychological phenomena are best explained in terms of dynamic information-processing models we should now consider why we might want to use neural networks as the basis for those models. There are three basic reasons for using neural networks. First, we may aim to discover something about methods of information processing used in the brain by studying analogous processes in artificial neural network models. The construction of these models will be driven more by biology than by computational theory. Second, we may attempt to gain insights into the natural decomposition of a problem by comparing the details of an optimal or near-optimal machine solution to the problem (perhaps given certain constraints and
styles of input representation) with human or animal performance. This approach may offer insights into the constraints faced by natural systems and the forms of representation they use. For many complex non-linear problems optimal solutions cannot be identified analytically but artificial neural networks can provide near-optimal solutions numerically. The use of more traditional symbolic artificial intelligence methods in cognitive science can also be seen in this light. In this case optimality is in the hands of the system designer rather then being implicit in the numeric properties of the system. Finally, neural networks can be used to determine statistical properties of the input data to a problem which we may expect to be psychologically significant regardless of the methods or even goal of the solution (Baddeley & Hancock, 1991; Valentine, 1991).

In this chapter I will concentrate on the first of these uses of neural networks: gaining insights into the mechanisms of natural neural computation. I will initially review work in artificial neural networks to find possible bases for dynamic information processing in the brain. This search needs to be guided by anatomical and physiological constraints. This adherence to biological constraints need not be slavish — aiming for models which are biologically realistic down to the last sodium channel is not necessarily the best course for brain theory; simpler models are more open to analytical understanding and also more practical to simulate numerically on a large scale, which may be very important in the study of collective phenomena. On the other hand, models which include features that are biologically impossible must be discarded.

7.3 A brief history of time in neural computation

I will now survey various approaches that have been taken to producing time-dependent computation in artificial neural networks, giving consideration to the applicability of these approaches as biologically plausible models of the mechanism of computation.

7.3.1 Time-independent feedforward systems

The most popular artificial neural network system of recent years, the multi-layer perceptron trained using the back-propagation learning algorithm (Rumelhart, Hinton & Williams, 1986a,b) has no dynamics in its response to stimulation — it is a purely feedforward system in which the instantaneous mapping of inputs onto outputs is independent of the prior state of the network. This system has also been justly criticised as being extremely biologically unrealistic (see e.g., Crick & Asanuma, 1986). However, it is only
the back-propagation training algorithm which imposes biological impossibilities such as bidirectional symmetric synapses on the system. Identical training can be achieved by more realistic reinforcement training, albeit much more slowly (Barto & Jordan, 1987, see also McClaren, 1989, for a biologically plausible method of implementing back-propagation). Multilayered perceptrons have been used successfully in time-dependent tasks such as production of speech from text (Sejnowski & Rosenberg, 1987), this time-dependence being achieved by segmenting the time-varying input to the system into a series of static spatial patterns. This approach can only succeed if some external executive segments and buffers real-time input in order to allow learning to occur and then performs similar segmentation during recall. This segmentation of the world into a series of snapshots seems psychologically unrealistic and requires a complex supervisor which requires time-dependent processing itself.

A variety of other systems also produce input–output mappings which are independent of the prior state of the network but, rather than being trained to make a specific set of desired mappings as the multilayer perceptron typically is, they are constructed so as to extract statistical regularities from their input data (e.g., principal components extracted by systems such as those of Linsker, 1989 or Sanger, 1989 or allied competitive learning systems such as that of Rumelhart & Zipser, 1985 or the orthogonalisation of input data in the work of Kohonen, 1988). These systems vary in their degree of biological realism. Although none can model the temporally extended tasks discussed earlier in the chapter, one or more of these systems may help describe certain types of early perception which appear, on the basis of the timing of associated single-cell evoked responses, to be dominated by feedforward processes which take place too quickly for feedback to have a significant effect (Oram & Perrett, 1992).

7.3.2 Back-propagation in recurrent networks

Although multilayered perceptrons using back-propagation cannot deal with temporally extended tasks directly, a variety of time-dependent derivatives of the system have been developed. The simplest of these allows networks which contain feedback connections, and hence are likely to produce a time-varying response to an initial time-invariant stimulus, to be trained using the back-propagation learning procedure by unfolding in time (Rumelhart et al., 1986 a,b). In this scheme the network is “unfolded in time” to produce a purely feedforward network in which each layer is a copy of the neurons in the original network at each successive instant in time. Connections within the
original net are now all feedforward as they join neurons at one instant to those they connected to in the original net, but at the next instant (and hence in the next layer). A very minor variation of the standard back-propagation learning rule can be used to train this unfolded network with time-invariant or time-dependent stimuli. When the resulting connection strengths between all of the copies of pairs of neuron in the original net are averaged and entered between those pairs in the original net then it will produce the response trained in the unfolded net. This scheme is, of course, quite impossible to implement biologically as it stands — even though actual unfolding is unnecessary, each unit must nevertheless store and recall all of its activation values over each trial. As we shall see, unfolding also requires unrealistic prior knowledge of the task to be learned.

The general principle behind back-error propagation is gradient descent of a surface describing error (the difference between the network’s actual and desired outputs) in a multidimensional space defined by the connection strengths of the network (weight space) achieved by iteratively computing the partial derivatives of error along each axis of weight space. Unfolding computes partial derivatives for each weight averaged over the whole duration of each trial — it is therefore necessary to know exactly when each trial starts and how long it last for. Two related algorithms can be used to train networks containing feedback connections to perform time-dependent tasks without the prior knowledge of the duration and timing of the task required by the unfolding approach. Instead of computing average error gradient over whole trials, these algorithms can compute error gradient at each instant, and hence prior knowledge of the timing of the trial is not necessary. Both the algorithms of Pearlmutter (1988) and Williams and Zipser (1989) are unfortunately impossible to implement biologically. The Pearlmutter algorithm requires that each neuron knows all of its prior states at any instant (or that time can be run backwards!), while Williams and Zipser’s algorithm requires that each neuron knows the state of all other neurons in the network at each instant whether or not they are synaptically connected.

7.3.3 Feedforward systems with context units

The most popular recent variants of multilayer perceptron back-propagation involve adding a set of connections to a standard three-layer architecture so that the state of either the hidden layer (Elman, 1988) or the output layer (Jordan, 1986a; but see also below) at one instant becomes part of the input to the network at the next instant. The network is then trained to produce desired responses to time-varying inputs using the standard back-propagation learning
rule. These techniques have been quite successful with small networks. However, it must be stressed that, unlike the back-propagation variants discussed previously, there is no analytical justification for the success of these algorithms. It is reasonable to suspect that as networks and the number of feedback connections get larger these systems will become more and more likely to become unstable. These potential scaling problems apply not only to the training phase (in common with all back-propagation procedures which scale rather badly) but also to simple responses to stimuli after training. Even if this is not the case, these systems share all the biological impossibilities of normal back-propagation. It is not clear whether they could be trained by more realistic reinforcement algorithms which rely on time-averaged statistical search for descending gradients rather than the guaranteed step-by-step descent of back-propagation — there would clearly be a requirement that the time scale of stimulus change was vastly slower than that of synaptic modification. All in all, these systems do not hold much promise as models of the dynamic information processing in the brain.

7.3.4 Back-propagation in time-delay networks
A second approach to modifying back-error propagation in multilayered perceptrons for temporal tasks is to make the behaviour of units or weights in the network time-dependent themselves rather than introducing feedback connections. The simplest type of time-dependent unit is one which is connected to itself; i.e., its activation at any time is a function of its current inputs and its prior output. Jordan (1986a,b) used fixed self-connections of this type in the additional input-layer units which received their activation from the previous output of his network. In this case their main function was to smooth variation in the activity of output feedback units in order to conform with speech phenomena being modelled. Stornetta, Hogg and Huberman (1988) and Mozer (1989) have self-connected units explicitly as the sole means of introducing time-dependence into feedforward networks. The drawback of this approach is that time-dependence is necessarily dominated by a recency effect — an unnatural performance limitation. A more sophisticated approach involves each unit being a function of its prior inputs over all time. In models of this type the activation of each unit is a time-integrated function of all its prior inputs multiplied by their respective connection weights, most importantly, however, those weights themselves are functions of time. This last characteristic eliminates the domination of time-dependence by recency. Networks of this type can be trained using back-propagation procedures and can be very successful in performing time-dependent tasks such as phoneme recognition.
Critical dynamics of neural networks 187

(Lang, Waibel & Hinton, 1990; Waibel, Sawai & Shikano, 1988). Time-delay methods can be computationally expensive for large problems (see Waibel et al., 1988). However, de Vries and Principe (1992) have recently proposed an approximation to generalised time-delay models with much better scaling characteristics. Again, it is doubtful whether any of these systems could form the basis for a model of the mechanism of time-dependent neural computation. However, the practical success and relative analytic tractability of these models indicates that they may be very useful in understanding time-dependent tasks.

7.3.5 Attractor networks
The final group of systems I will consider, attractor networks, can exhibit time-dependent behaviour, but in normal use the aim of adaptation rules is to avoid time-varying behaviours. These systems include Boltzmann machines (Hinton & Sejnowski, 1986), Hopfield machines (Hopfield, 1982, 1984) and more general attractor networks in which some of the strict constraints of the Hopfield machines are relaxed (see, e.g., Amit, 1989). The networks in all of these systems comprise both feedforward and feedback connections and hence they exhibit time-dependent behaviour after stimulation rather than the effectively instantaneous input–output mapping of purely feedforward systems. Both the Hopfield and Boltzmann machines have severely constrained architectures in which all connections between units are bidirectional and symmetrical. This constraint guarantees that although a network’s response to a stimulus will be time-varying it will eventually reach a stationary stable state (or point attractor). The learning algorithms used with these systems are designed so that useful information is stored in these stable states. This information will therefore crystallise out of these networks given a suitable initial stimulus.

The requirement for symmetric connections in these systems both limits their biological plausibility as brain models and eliminates any time-dependent behaviour beyond a simple descent to a stationary equilibrium state. Amit (1989) reviews the effects of removing this requirement. The effects of diluting the proportion of symmetric synapses in a network can be studied from two points of view; first, by studying the effect of dilution on the original stationary storage behaviour of the network, and second, by studying new time-dependent behaviours which dilution allows. Although very interesting, the first of these points of view need not concern us here. The general approach taken in the studying “useful” time-dependent behaviour in asymmetric networks is to consider connections as having separate symmetric and asymmetric components. The symmetric components support point attractors in networks just as they did in the symmetric case while the asymmetric
connections induce transitions between attracting states. Early attempts (Hopfield, 1982, p. 558) at encoding sequential behaviour using this approach had limited success because the transition components were constructed to give a single time-step delay between transitions to the next attractor in the sequence. This did not allow the attracting states (strictly quasi-attracting states) to stabilise and hence was very susceptible to noise. This approach failed when interaction between neurons in the network was asynchronous because of this susceptibility to noise. The situation is markedly improved by introducing delays into the actions of the transition synapses thereby allowing each quasi-attractor in the sequence to stabilise somewhat before the transition term begins to dominate. Of course, we now have two different types of synapse, fast and slow, and have reinstated a symmetry requirement for the fast synapses. This approach can be extended to include synapses whose behaviour is time-dependent in a similar manner to those used in more sophisticated time-delay back-propagation systems. In general these networks can be viewed as sequences generators rather than discriminators. Amit (1989) demonstrates, however, that a third term can modulate the transition between quasi-attractors in a recall sequence. Although one can imagine a number of such terms differentially modulating such sequences and terms within sequences being susceptible to different modulators producing a network capable of differential sequence recognition, this seems unwieldy and very likely to succumb to noise and capacity limitations of this type of network.

A number of alternative approaches to producing time-dependent behaviour in attractor networks without fast synapses based on noise-induced transitions between quasi-attractors have also been investigated. Buhmann and Schulten (1987) describe a noise-driven system in which the symmetric portion of weights is constructed from both supportive excitatory-based connections based on covariance of units within patterns to be recalled and suppressive inhibitory connections determined by covariance between those patterns. Controllable sequential behaviour can be produced by making two modifications to these connections. First, the substitution of excitatory connections for the suppressive component of connections between units in successive patterns produces associations between patterns. Second, this association is made unidirectional by enhancing the suppression of successor patterns by the predecessors. This system suffers the severe limitation, however, that it is only capable of recalling sequences of patterns which overlap with each other. Buhmann (1989) and Buhmann and Schulten (as described in Amit, 1989, pp. 256–262; see also Buhmann & Schulten, 1987) have also investigated variants of attractor networks whose time-dependence is limited to
oscillation either around or between stored patterns. In the Buhmann and Schulten model two separate attractor networks are coupled by connections between these nets which are inhibitory in one direction and excitatory in the other direction. An oscillation whose period is dependent on the level of noise can be set-up between two quasi-attractors in this system. Buhmann (1989) describes a system in which a symmetrically connected network of excitatory cells is modulated by inhibitory connections with spatially localised receptive fields (in contrast to the unconstrained connectivity of the excitatory connections). The effect of these inhibitory neurons reduces the overall level of activity in the network and causes stored states to be recalled as oscillatory firing of the pattern’s constituent units. In itself this is not a very interesting type of time-dependent behaviour. However, the motivation for investigating oscillation in artificial neural networks is its presence in brains and the possibility that it is stimulus-dependent and hence has functional significance in information processing (Eckhorn et al. 1988; Gray et al., 1989; but see also Young, Tanaka & Yamane, 1992).

At this point it seems wise to summarise this survey of time in artificial neural networks and to then move onto more biologically inspired systems. Although there is a wide range of networks capable of producing time-dependent behaviour, none of them are particularly satisfying bases for models of biological information-processing mechanisms. Recurrent variants of back-propagation either require that neurons have access to non-local information or that they have infinite memories. Feedforward networks with context units have unnatural architectures, an unnatural learning rule and may well scale badly as they rely on being approximations to stationary systems rather than having a truly dynamic basis. Feedforward systems which use time-delay units are probably the most successful time-dependent artificial neural networks in practical terms. However, they too have biologically unrealistic architecture and learning rules and synapses for which it is difficult to envisage a biological implementation. Attractor networks with asymmetric connection can have architectures which are not entirely at odds with neuroanatomy, but unfortunately they would fail in a noisy biological implementation. Eliminating this sensitivity using slow synapses reinstates the biologically unrealistic requirement for symmetric synapses.
7.4 Cortical neurodynamics

In this section I describe a series of models I have studied in order to evaluate the plausibility of particular classes of dynamic behaviour as bases for cortical information processing.

7.4.1 Time-dependent behaviour in biologically inspired systems

One class of behaviour in the cortex currently attracting a great deal of interest is oscillatory neural activity. Although this is a form of time-dependent behaviour it is rather too impoverished to meet the types of problems discussed earlier. Oscillation is, in fact, most often posited as the solution to a problem in static representation — the binding of disparate feature representations into unified object representations (Wang, Buhmann & von der Malsberg, 1991). The second type of dynamic behaviour often studied in biologically inspired models is central pattern generation (e.g., Selverston, Miller & Wadepuhl, 1983). One problem with this type of work is that it tends to deal with rather small and specialised non-vertebrate systems while the psychological problems that motivate this chapter are more likely to be tackled by large-scale, relatively general purpose hardware.

The topic of pattern generators raises a more general issue: What is the relationship between systems that can autonomously generate time-dependent sequences of states and those that can discriminate between time-dependent sequences of stimuli? In the psychological issues discussed earlier the latter ability was far more important. In order to address this question it is easiest to consider the behaviour of networks as rather general dynamical systems. In networks like Hopfield (1982) machines the whole network state is determined by a stimulus. The eventual state of the network depends on which of a number of basins of attraction the stimulus lies in. Each basin of attraction is the set of states from which the network will eventually converge to one particular attracting state — in this case a stationary equilibrium state. In networks where the state of only a subset of units are determined by a stimulus, the basin of attraction into which the network falls is determined both by its prior state and by the stimulus (which can be viewed as a perturbation of the network’s state). If the attractors of the network are not equilibrium points but rather paths in phase space (limit cycles or strange-attractors) then the effect of a particular perturbation may change as the state of the network evolves around (or toward) this path. Particular sequences of perturbations may be consistent with the paths of particular attractors (or particular segments of those paths) while other sequences of perturbations may move the state of the network into
Critical dynamics of neural networks

different basins of attraction. In some regions the system will be much more sensitive to perturbations in switching between attractors or between the lobes (near looping paths) of individual structurally complex attractors. Networks with complex state dynamics can therefore act as discriminators between stimulus sequences.

Complex dynamical systems can exhibit attractor structures which are rich enough potentially to support the types of task discussed earlier. These tasks also require a process which is not just time-dependent but also predictable and capable of handling generalisations. In essence this means that the attractors governing a network's behaviour should occupy relatively small and compact regions of the available phase space (so that behaviour on the attractor is relatively predictable) and have reasonably wide basins of attraction (so that reasonably disparate sets of stimuli can produce similar behaviours). In systems with very large numbers of neurons, and hence very high dimensional phase spaces, these criteria are only likely to be met by attractors with relatively low dimensionality — such attractors may be expressed in the network as cooperative behaviour between large number of neurons forming coherent spatio-temporal patterns of activity. The problem of reducing the very large number of degrees of freedom in a neural network into a predictable low dimensional behaviour will be addressed later in this chapter.

Before investigating the types of dynamics that a physiologically inspired neural network model can support, it is worth pointing out that types of complex dynamics similar to those discussed above have already been inferred in animal and human brain activity. Freeman has studied attractor dynamics in models of rabbit olfaction for some time (see e.g., Freeman, 1991, for a recent review) and has found considerable evidence that electrical activity in the brain is best characterised by complex, probably chaotic, attractors. Babloyantz and Destexhe (1987, 1988) reach similar conclusions from analysis of human electroencephalographic (EEG) data. Freeman has based his simulation work around equations describing the behaviour of populations of neurons rather than network simulations in which individual neurons are represented (see e.g., Yao & Freeman, 1990). In the following section of this chapter I will consider the generation of time-dependent behaviour in network models.

7.4.2 A minimally plausible model for cortical dynamics

I have studied a series of numerical models in order to investigate the types of dynamic behaviours which might support information processing in the brain. In contrast to most artificial neural networks the construction of these models was not guided by an attempt to produce one particular type of behaviour.
Rather, features of real physiology and anatomy were adhered to if their exclusion could be expected to reduce the range of behaviours open to the model. On the other hand, as complex collective behaviour often only emerges in large systems every effort was made to keep the models simple in order to make large-scale simulations practical.

A few of the classes of dynamics we may expect to find in cortical networks were mentioned in the discussion of attractor networks above. There are, however, a number of important differences between the architecture of the cortex and that of typical artificial attractor network models. For the present I will ignore the layered third dimension of cortical structure. In the other two dimensions, in addition to a variety of functional architectural features characterised by the organisation of receptive field characteristic of neurons such as microcolumns (Mountcastle, 1978) or blobs (Hubel & Livingstone, 1987), there is a spatially dependent organisation of connectivity between neurons (Braitenberg & Schütz, 1990; Lowel & Singer, 1991). I will not attempt to model functional architectural features since these are the types of characteristic that we might hope would emerge from a model. The space-dependent connectivity of neurons, however, has important implications for network dynamics. In randomly connected systems such as attractor networks the relative spatial position of neurons had no functional significance. The characterisation of these systems' behaviour is entirely temporal. For example attractor networks might settle to stationary states, cycles of states, chaotic sequences of states, or quasi-stationary states followed by rapid transitions (see e.g., Kürten & Clark, 1988). In networks in which space has significance we can classify behaviour both in terms of spatial and temporal characteristics. There is an obvious sense in which stationarity, cyclicity and chaos have spatial as well as temporal meaning (formally we might define them in terms of the spatial covariance function of a system). Uniform, cyclic or chaotic stationary patterns can therefore exhibit simultaneous fixed, oscillatory or chaotic temporal behaviour (although we should not expect to see all of these behaviours in one particular system). In addition to these behaviours, more complex interactions between spatial and temporal instabilities can occur involving coupling between diffusive spatial processes and more localised processes in which particular states become dominant (see e.g., Nicolis & Prigogine, 1989, pp. 147-192). There are clear analogues of the physical examples discussed by Nicolis and Prigogine (1989) in neural networks — first, activity in one group of neurons will tend to be transmitted to neighbouring cells; and second, competition between excitatory and inhibitory processes within a localised group of cells (including perhaps interneural inhibition and intraneural effects
such as membrane decay and refractoriness) can produce localised selection of sustained activity, sustained quiescence or even more complex cyclic or chaotic behaviours. In physical and chemical systems the result of these coupled processes is the appearance of large-scale spatio-temporally organised dissipative structures (e.g., non-linear wave-fronts) produced through the amplification of small perturbations. The occurrence of this coupling depends on a number of characteristics of a system. First, the system must be far from equilibrium (thus allowing attractors other than a single universal equilibrium point attractor to govern the system's evolution). Second, it must be in a regime near to a bifurcation between single and multiple attractors. Large-scale structure will appear through a process of self-organisation if the selection of one of these attracting states in a small element of the system causes similar selections of that particular attracting start to be made by distant elements of the system. The required long-distance interactions become possible as the characteristic correlation length between states of different parts of the system diverges to infinity at the phase transition corresponding to the emergence of the new attractors. Finally, the time-scales over which the local selective processes and the diffusive transport processes operate must be similar. These requirements sound quite onerous. However, apart from the last, all are met by spatially distributed systems maintained near phase transitions.

In all the systems described in this section, the architecture of connections was determined by spatially gaussian probability distributions. That is, the probability that any neuron was connected to any another neuron was proportional to the gaussian of the distance between those neurons. This isotropic connectivity is more similar to that found by Braitenberg and Schütz (1990) in the basal dendritic fields of cortical pyramidal cells than to the anisotropic connectivity found by Lowel and Singer (1991) in long-range tangential fibres in area 17 which were probably related to ocular dominance columns. The evolutionary trend of increasing physical isolation between the basal and apical fields of cortical pyramidal cells (see e.g., Goffinet, 1990) suggests a functional separation of apical and basal systems. Figure 1 shows the input connections to a small percentage of neurons in a typical model network.

In addition to this architectural organisation I alluded to the role of time-dependence in the transmission and decay of neural activity in the potential formation of organised dissipative structures. Time-dependence in co-operation between neurons has been studied by Abeles (1982a, 1990). He demonstrates the increased efficacy of synchronous as opposed to asynchronous presynaptic action potentials in firing a neuron. If such effects are to occur it is necessary to model individual action potentials rather than the average firing rate of
neurons. This could be achieved using a multi-compartment neuronal model. However, a simple threshold firing leaky integrator also exhibits sensitivity to the relative synchrony of its inputs as described by Abeles (1982b, 1990) while requiring much less computational effort. It must be accepted that this approach leads to spatio-temporal processes within neurons being ignored. However, for the present purpose it is local time-dependence which is of particular importance. The model neuron used in all the systems described in this chapter was therefore a spiking leaky integrator which also exhibited absolute refractoriness after firing. It is described formally as follows:

\[ \text{if } v_j(t) > th_j \quad x_j(t) = ap_j \quad \text{and } v_j(t + 1 + r) = \sum w_{ij} x_i(t + r) \quad (1a) \]
if $v_j(t) \leq th_j$, $x_j(t) = 0$ and $v_j(t+1) = \sum w_{ij} x_i(t) + d_j v_j(t)$ (1b)

where $v_j(t)$ is the "potential" of neuron $j$ at time $t$, $th_j$ is a constant excitation threshold for neuron $j$, $x_j(t)$ is the activity of neuron $j$ at time $t$, $ap_j$ is the strength of an action potential (which may be negative for inhibitory units), $d_j$ is the time decay constant of the potential variable for neuron $j$, $w_{ij}$ is the strength of the synapse of neuron $i$ on neuron $j$ and $r$ is the refractory period (apart from the time-step during which the neuron fires, the output $x_j(t)$ is zero during the refractory period).

In preliminary experiments the network included both excitatory and inhibitory units whose proportions and connectivity reflected statistics reported in Braitenberg and Schütz (1990). Most of the later results described here, however, use purely excitatory networks.

7.4.3 Dissipative structure formation and time-dependent cortical information processing

The model I have described should support a number of different types of dynamic behaviours depending on the relative values of its controlling parameters. Only a few of these behaviours are potentially useful for information processing. One requirement is that the behaviour be time-dependent. Additionally, the behaviour must be reasonably predictable and produce activity which could form usable input to subsequent systems. The prime candidate behaviour is dissipative structure formation. These structures are clearly time-dependent yet relatively predictable in the way in which fronts of dissipation retain coherence as they travel through a system. As dissipative structures are formed from small perturbations they also provide a mechanism whereby the stimulation of relatively small numbers of sensory neurons can produce large-scale systematic responses. It is a plausible hypothesis that small systematic spatial variations in connection strengths between neurons could favour the production of structures with specific characteristics by particular sequences of perturbations. In addition, the way in which the consequences of stimulation are exhibited at points in the network which are distant in both space and time from the site of stimulation is biologically convenient. We do not have to conceive of processes in which the whole cortex adapts to one stimulus at a time. Rather, once a stimulus has been received in one location that location is free to receive new stimuli while the effects of previous stimuli still persist in the net. Finally, the coherent nature of dissipative structures makes output interpretation in subsequent systems relatively straightforward.
Freeman (1991) has described an alternative candidate behaviour based around low-dimensional almost limit cycle-like chaotic behaviour. Although proposed in the context of a time-independent task — odour identification — the dynamical system shares a number of features with one capable of generating dissipative structures. In this scheme the behaviour of the olfactory system is governed by a chaotic attractor with a relatively high-dimensional ground state and a number of lobes with highly constrained low-dimensional orbits. Small systematic perturbations (stimuli) move the system out of its chaotic ground state into one of these lobes. When the system is in one of these highly constrained orbits, the consequence is that activity in the olfactory system is organised into a systematic spatial pattern. The transition from the chaotic ground state to one of the more constrained orbits is described in terms of a phase transition, just as dissipative structure formation takes place at the bifurcation from single to multiple attractors which may be associated with a phase transition. Similarly, the selection of one particular attractor on the multi-stable side of the transition is determined by the amplification of a small perturbation into a coherent system-wide behaviour. Finally, the selection of one particular lobe in Freeman’s scheme is expressed as a systematic spatio-temporal pattern of activity in the neural network generating the attractor. These similarities indicate that the same type of underlying mechanism is being described in two alternative ways. Freeman’s work has been based on systems of equations which describe the behaviour of populations of neurons. The models I have been using allow the potential for this type of behaviour to be studied in networks where individual cells are modelled and so could complement Freeman’s work.

7.4.4 Coherent responses to stimulation in networks with spatially localised connectivity

In this section I describe the results of some simulations using the model network described above. In all cases the strengths of connections were random, chosen from a uniform distribution between 0 and 1. The parameters of the network were average firing threshold, average membrane time-constant, standard deviation of neurons’ connectivity distribution and average connectivity of each neuron. All of these variables were selected from gaussian distributions. The aim of these studies was to determine whether or not networks of spiking neurons with spatially localised connectivity could produce coherent spatio-temporal patterns of activity in response to stimulation. The simplest response of this type is the production of a travelling pulse of activity in a one-dimensional system stimulated by an initial discrete
Figure 2. Pulse propagation in one-dimensional networks. In each of the three panels the network is aligned vertically and is shown at successive time-steps from left to right. Firing units are shown as filled pixels.

external excitation. A range of one-dimensional configurations of the model involving networks containing both excitatory and inhibitory neurons and pure excitatory networks was studied. In all cases the networks were initially quiescent, with all membrane potentials set to zero. A group of adjacent neurons at the bottom of the one-dimensional network were then fired and the subsequent behaviour was studied. Figure 2 shows the results of three simulations which capture the three types of behaviour possible in this configuration. In these examples the networks consisted of 1005 neurons, of which 75% were excitatory and 25% inhibitory. Both types of units made 75 synapses on average, and the standard deviation of their connectivity distributions was a range of 75 locations in the network. The membrane leakage parameter $d_j$ was 0.95 for both types. The inhibitory units fired with a strength $ap_j$ of −10 while the excitatory units fired with a strength of 1. The threshold $th_j$ was always 50 for the inhibitory units. In the panel on the left the threshold $th_j$ of the excitatory units was 10.8, in the centre it was 10.6 and on the right 10.4. One hundred additional excitatory neurons fired together at the bottom of the net provided the initial stimulus.

Space, in other words the position of neurons in the network, is represented on the y-axis, and time proceeds rightwards along the x-axis. The initial externally applied stimulus can be seen as the solid bar at the bottom left of the three panels. In the left panel the stimulus begins to produce a travelling pulse in the network but this soon dies away. In the rightmost panel the stimulus also begins to produce a pulse, but this soon loses its coherence and activity
Figure 3. A series of images of a two-dimensional network show the development and propagation of a travelling non-linear wavefront of activity. Firing units are shown as filled pixels.
reverberates chaotically throughout the network. In the centre panel the stimulus-induced pulse travels across the entire 10,005 neuron network without losing coherence. Initially this appears to be a promising result. However, the networks that produced the three different behaviours shown differed only in the average value of the membrane time-constant of their constituent neurons. This difference was less than 5% across the three behaviours. Moreover, the boundary between those behaviours were not abrupt, but rather “islands” of each behaviour appear and disappear as parameter values such as average time-constant are varied. Even worse, a set of parameter values which produced one behaviour with a particular stimulus intensity might produce a different behaviour if the stimulus intensity was varied. This extreme sensitivity to parameter values was not cured when networks composed of both inhibitory and excitatory units were tested. Clearly such sensitivity, particularly if it is dependent on stimulus intensity, makes a functional role for dissipative structures extremely unlikely in biological analogues of the modelled system. The problem of parameter sensitivity in models of this type is well known from the classic work of Beurle (1956, 1962). We will now consider its solution.

If these coherent travelling pulse responses are interpreted as being dissipative structures then, in the light of our understanding of dissipative structure formation, part of the problem may be that the stimulus has two separate roles to play. Not only must it provide the coherent seed for a larger scale structure, but also it must be of appropriate intensity to put the system into a state near a phase transition (at least locally). The choice of parameter values may be less critical if the network is maintained near a phase transition by diffuse background activity which is distinct from the coherent stimulus. This notion was pursued in studies of two-dimensional networks in which each neuron in the network had a small random amount of potential added to it at each time-step. One reason for the choice of two-dimensional networks was for comparison of the types of structures generated with those commonly produced in chemical and physical systems.

Results from two separate simulations are shown in Figures 3 and 4. The network which produced the behaviour in Figure 3 consisted of 10,000 excitatory neurons, firing with a strength $aP_i$ of 1, each of which made an average of 15 connections with a distribution whose standard deviation was 5 locations, a threshold $th_j$ of 10, a leakage $d_j$ of 0.9, and a refractory period of 1 step. At each time-step 1.1 units of potential were injected into each unit.

In Figure 3 a series of images of activity in a two-dimensional network shows how an initially random activity distribution nearly dies completely
Figure 4. A two-dimensional network showing "target" waves — a series of non-linear wavefronts emanating from the same centre. Firing units are shown as small circles.

away, but then begins to organise and amplify around a point in the network from which a spreading coherent front of activity emerges. It is also apparent
that this organised activity is not directly due to an injected signal (although one was injected into the network earlier in the simulation). Apparently, also in keeping with the behaviour of physical systems, this front is centred around some sort of "defect" in the connectivity distribution of the network. This conjecture is supported by the observation that networks with identical parameter values but built using different seeds for the random number generator which determines the connectivities and weights in the network can produce quite different behaviour — often failing to produce coherent responses at all, or producing waves emanating from a number of centres simultaneously. In this second situation waves annihilate each other upon meeting, consistent with the behaviour of non-linear wave-fronts formed in dissipative structures in a range of other systems such as the unstirred Belousov–Zhabotinski reaction or in slime-mould aggregation (see e.g., Nicolis, 1989, pp. 323–327). Further evidence of the similarity between the behaviour of this type of neural network and the examples mentioned previously is shown in Figure 4. Here a much smaller network of 2304 neurons can be seen producing "target" waves — a series of successive non-linear waves emanating from the same centre.

The behaviours described above indicate that neural networks with spatially localised connectivity can support dissipative structure formation. The types of behaviours obtained were similar in many ways to those found by Destexhe and Babloyantz (1991a,b) in studies of networks with spatially localised connectivity driven by oscillatory inputs in which spiral waves of coherent activity, another typical form of simple dissipative structures, could be generated. The production of these behaviours in the networks I have studied remained quite sensitive to network parameter values even when these networks were driven far from equilibrium by diffuse background potential injection. The effect that background stimulation should have on the system is to maintain it near the phase transition between a stable quiescent state and a high activity state which may be chaotic or may involve bulk oscillations of activity throughout the network. Past this critical state the behaviour of groups of neurons is bistable — one or other of these two states can be induced by suitable perturbations. In the experiments just described, structures only arose some time into the simulation, well after any coherent activity due to the stimulus injected at the start of simulation has dispersed. The emergent structures must therefore have been seeded by small inhomogeneities in the networks. If such small inhomogeneities are to produce structure then the system must be in a supercritical state well into a bistable regime. This requirement for supercriticality may well account for the sensitivity of
behaviour to network parameters. In the framework I have been proposing for dynamic information processing, structure formation would be seeded by coherent stimulus patterns which could be expected to have much larger effects than inhomogeneities in the structures of the network alone. (In fact, we can conjecture that the conjunction of stimulus patterns and inhomogeneities in the structure of the network could act together to selectively produce particular behaviours in response to particular stimulus sequences.) Given the presence of stronger "seeds" to structure formation the network need not be in a supercritical regime as strong perturbations could push it locally into a particular stable state beyond the critical point from a critical (or even marginally subcritical) regime. These conjectures suggest a series of studies which could be undertaken to examine the mechanisms of spatio-temporal pattern formation in neural networks in more detail. First, we need to establish whether critical dynamics are indeed being maintained by background stimulation of these networks. The range of conditions over which criticality can be maintained must also be determined. Given a critical regime we can then attempt to examine the induction of structure in more detail, both in terms of the properties of localised and diffusive processes in the network and in terms of the effects of inhomogeneities in the connectivity of networks and coherent perturbing stimuli. One possibility of particular interest is that systematic organisation of inhibitory and excitatory neurons such as that found in cortical microcolumns may extend the range of background activities which produce bistability. The series of studies I report here examines the first of these issues: the maintenance of the network in a critical regime which is a prerequisite for dissipative structure formation.

7.4.5 Self-organising criticality in neural networks

In the previous experiments it was found that networks with spatially localised connectivity could support dissipative structure formation. However, this behaviour was very sensitive to network parameter values. One of the prerequisites for dissipative structure formation is approximately critical dynamics. In the previous section it was assumed that diffuse low-intensity stimulation maintained networks in this state. In this section the maintenance of critical dynamics by diffuse stimulation is examined in more detail.

There is a striking similarity between diffusely driven neural networks with spatially localised connectivity and forest fire models which are often used to study phase transitions and critical phenomena in percolation theory (see e.g., Stauffer, 1985). Both types of model consist of large numbers of elements
Figure 5. Membrane potential distribution in a driven two-dimensional network. The membrane potential of each unit is proportional to the diameter of the circle representing each unit. Firing units are shown by asterisks.

(neurons or groups of trees) which only interact with their close neighbours (defined either by localised connectivity or by adjacency on a grid) through a process (spiking or burning) in which their stored "energy" (membrane potential or tree-density) is dissipated. In both cases stored "energy" is gradually replenished (through diffuse background stimulation or tree-growth). The major differences between them are the relative diffuseness of the interaction neighbourhood and the exponential decay of membrane potential in the neural network. One very interesting property of some forest fire models which has recently been discovered (Bak, Chen & Tang, 1990; Chen, Bak & Jensen, 1990) is that a critical state can be maintained over a relatively wide
range of conditions though active relocation processes in the system — this is known as self-organising criticality (Bak & Chen, 1991; Bak, Tang & Wiesenfeld, 1987). Production of criticality over a similar wide parameter range would clearly increase the biological plausibility of a role for dissipative structures in cortical information processing.

Given the similarity of the network to the forest fire models studied by Chen et al. (1990), their procedures were adopted for examining critical behaviour in the neural network model described previously. Two measures were used to assess criticality — the first is a measure of spatial structure, whilst the second measures behaviour over time.

The spatial hallmark of criticality is the wide range of scales over which components in a system interact. These scales do not merely reflect the distances over which direct connections in the system exist, they are the ranges over which groups of components produce systematic effects via intermediate connections. In forest fire models these influences manifest themselves as clusters of sites, all of which have similar tree densities. The equivalents in the neural network model are clusters of neurons, all of which have similar membrane potentials. The size of these clusters indicates the range of influence. In a critical system the probability of finding clusters of a given size is proportional to some power of that size. This relationship should extend well beyond the scale over which direct interaction can occur and should extend up to the size of the system if boundary effects could be ignored. The relationship between cluster size and probability can be measured objectively using the structure function (see e.g., Paladin & Vulpiani, 1987):

$$S_q(r) = \left\langle \left| v_i - v_{i+r} \right|^q \right\rangle$$

(2)

where $i$ denotes a neuron in the network, $v_i$ is the membrane potential of that neuron, $v_{i+r}$ is the potential of another neuron at a distance $r$ from neuron $i$, square brackets indicate averaging over a shell which includes all neurons at the distance $r$, angle brackets indicate time-averaging. A power law relationship between $S_q(r)$ and $r$ indicates criticality. If desired, a series of structure functions using different moments $q$ can be used to determine the critical exponents of the system.

The structure function of a network of 2304 excitatory $(\alpha p=1)$ units was measured over 50,000 time-steps. Each of the units made an average of 15 connections with a connectivity distribution whose standard deviation was a range of three locations. The average thresholds were 7.5, the leakage parameter $d_j$ was 0.999 and the refractory period was 1 time-step. At each time-
Critical dynamics of neural networks

step, 0.05 units of potential were injected into each unit. Figure 5 shows the
distribution of membrane potentials in a similar network at one point during a
structure function experiment. Informally this figure does appear to show
clusters of neurons over a range of sizes, just as we expect in a system at
criticality. Subjectively it is very similar to the corresponding figure in Chen et
al. (1990). The results of the more objective structure function measurement are
shown in Figure 6. Because of the statistical nature of the structure function and
the very low expected probabilities of large clusters in critical systems it is
necessary to calculate the structure function over very long time-series. As the
method used to calculate the structure function in the present studies was
extremely computationally intensive, more conclusive results might have been
obtained had a longer run been possible (the data as presented required two
weeks of CPU time on a twelve processor Encore Multimax computer!). As it
stands, the structure function obtained does appear to show a power law
relationship which will appear as a straight line in a log–log graph such as
Figure 6. However, the radius of scaling only extends to between 5 and 10 units
of length, well below system size. The average range of connections in the
network was approximately 2 units and 90% of all connections were over less
than 5 units so the scaling relationship cannot be directly attributed to local
connectivity (which should have an exponential relationship to distance based
on the underlying gaussian rather than power law distribution anyway).
However, the limited range of scaling means we cannot be strongly confident
that the structure function obtained is best modelled as a truncated power-law
rather than some other relationship.

In addition to the measurement of the structure function Chen et al. (1990)
also measured the growth of perturbations over time. This type of
measurement is fundamental to the characterisation of chaos. The famous
"butterfly effect" (Lorenz, 1963) depends on the fact that in a chaotic system
orbits diverge exponentially. If one of a pair of otherwise identical systems is
subject to a very small perturbation then the difference between the states of the
two systems will grow exponentially in time. On the other hand, if the systems
were governed by a non-chaotic attractor then the perturbation would either
disappear altogether in time (for point attractors) or the difference between the
two systems would settle to a constant (limit cycle attractors). In the critical
regime perturbation are expected to grow. However, this growth is not
exponential but is weaker, being some power of time. These dynamics are
sometime referred to as "weak chaos."
Chen et al. (1990) measured the growth of perturbation by running a pair of identical forest fire simulations in parallel for some time to allow them to reach criticality. They then made very small random changes to the tree density at all sites in one of the models and measured the average difference between tree densities in corresponding sites in the two networks as time progressed. The equivalent procedure in neural networks is to make small changes to the membrane potentials of neurons in one of a pair of networks and then to measure the growth of this difference between corresponding units. In experiments of this type the main practical problem is not the length of run required, but rather limitations to the size of network that can be simulated. Once again, we wish to test for a power-law relationship, in this instance between time and a measure of distance between the states of two simulations. In order for results to be conclusive, reasonably long regions which exhibit either clear power-law or clear exponential relationships are required. There are two limitations to the length of scaling region which it is practical to produce from an experiment. First, as the simulations must be of finite size there is an upper limit to the maximum difference between network states. Second, for a time after the initial perturbation is made its effects will remain localised in individual neurons. It is only after neurons fire and the effects of perturbations become distributed that their effects on the collective behaviour of the network become interpretable. As the state of the networks is expressed through two complementary variables, membrane potential and neural firing, the effect of a perturbation may be expressed in one or other of these variables at any time.

Figure 6. A log–log plot of the structure function of a two-dimensional network.
Both of these measures can be used to derive a measure of state-difference between network. The measure is based on:

$$\delta = \sum_{i} |x_i^a - x_i^b|$$

that is, the sum of the absolute differences between the firing variables $x_i$ between the two networks $a$ and $b$ over all sites $i$. The measure based on membrane potential is also the sum of absolute difference in the membrane potential variable $v_i$ over all sites between the two networks. However, to reduce noise all pairs of neurons in which one or both are firing or refractory are excluded from the sum. The average potential difference over the number of pairs included in the sum is then taken.

As noted above, any systematic relationship between time and perturbation divergence produced by experiment is likely to be clipped at the top and bottom of its range. There are three possible types of results to a perturbation growth experiment: divergence becomes constant in time, or it is either governed by a power or an exponential relationship. The first of these results would be clear even if there was an initial "slow start" to divergence. The
second two possibilities may, however, be difficult to distinguish in noisy data with a limited scaling region. For this reason, the design of the following simulation experiments included a control condition where perturbation divergence was in measured networks whose connectivity was not spatially localised. These networks should only produce limit cycle or chaotic behaviour (see e.g., Kürten & Clark, 1988). The control networks were identical to their corresponding “experimental” systems in all details apart from the spatial dependence of connectivity in the experimental networks. Otherwise corresponding neurons had the same time constants, threshold, and number of connections with the same weights (although of course the targets of these connections differed). As these experiments were run on toroidally connected networks (i.e., there were no edges to these networks) the total number of connections in the control and experimental nets was identical even though the connections in the control networks were potentially of “infinite” range. In all cases networks of 8100 excitatory \((ap_j = 1)\) units were used. They all had average thresholds \(th_j\) of 15 and decays \(d_j\) of 0.999. On each step each unit was injected with 0.15 units of potential in one series of experiments and 0.075 units in a second series. In each experiment after 500 steps the potential in each unit was perturbed in one net of each pair by an average of 0.0015 units. In the experimental condition the connectivity distribution had a standard deviation of three locations; in the control condition the standard deviation was 900 locations.

The results of these studies into perturbation growth are shown in Figures 7 – 9. In all cases these are log–log plots of a difference measure against time. Power-law divergence should produce a straight line in these graphs whereas exponential growth will produce an exponential curve. Figure 7 shows the growth of the difference in firing measure between the two networks in the control and experimental pairs on the left and right respectively. Both sets of data are quite noisy and it is not clear whether these data are best fitted by straight lines or exponential curves.

Since the noise is of relatively high frequency compared to the underlying trend we are interested in the underlying trend which may be made clearer by high-frequency filtering. In Figure 8 the data from Figure 7 is shown after noise was removed with an exponentially decaying travelling window filter. The exponential signature of chaos is readily apparent in the control condition while there is a clear linear trend in the central region of the data from the experimental condition. Similar clear data can also be obtained by averaging over a number of simulations where connection strengths and thresholds are generated with different random number sequences.
Clearer results are also obtained when the potential difference measure is used instead of the action potential spike measure. Figure 9 shows results using this measure. In addition to showing clear differences between the linear and exponential trends in the experimental and control conditions these data also begin to show the robustness of self-organising criticality in neural networks. They were obtained from networks with parameters identical to those used in the spike difference simulations described above, yet the rate of background stimulation in this experiment was half that used in the spike-difference experiment. In both spike and potential difference measures, good evidence of criticality could be obtained at both rates of background stimulation.

The results of these studies indicate that the role of the low-level background stimulation in the studies of dissipative structure formation I described above may well have been to induce criticality in networks. This is an interesting result in terms of the general model of time-dependent information processing I outlined previously. However, if that model itself is to be of any value we need to assess not only whether it is possible for its underlying processes to occur in the brain, but also whether they actually do. It is usually the case that the validity of models of biological networks is hard to test because the behaviour of sufficient numbers of specific neurons and the structure of their architectural relationship cannot be determined in living.
Figure 9. A log–log plot of the growth of a perturbation measured in terms of differences between membrane potentials in corresponding units in two networks. In the experimental condition on the right the networks have spatially localised connectivity; in the control condition on the left they do not. The rate of background activity injected into the net is half that of the networks used in Figure 7.

tissue in adequate detail. The model I have been proposing has some advantages in this respect since self-organising criticality produces an easily identifiable and measurable collective behaviour, a $1/f$ power spectrum, in a system (Bak et al., 1987; Christensen, Olami & Bak, 1992). The power spectrum is a measure of the relationship between the relative frequency and intensity of energy dissipation in a system — in a neural network intensity corresponds to the number of neurons firing within a given period, and it is therefore easy to produce reasonable measures of such bulk activity with extracellular recording or EEG techniques. Freeman and van Dijk (1987) found that EEG activity recorded from the visual cortex of a rhesus monkey performing a task requiring visual attention had a $1/f$ power spectrum. Similar results have been reported by Young et al. (1992) who found $1/f$ power spectra in local field potential recordings from the visual and medial-temporal cortices of both behaving and anaesthetised macaque monkeys. It is hard to explain the production of these $1/f$ spectra by any means other than dissipation of diffuse non-coherent background activity through a process of self-organising criticality, although
the nature of processing arising from coherent activity produced by stimuli is much more.

7.5 Conclusion: Information processing and criticality

I will end this chapter by considering a selection of work from other disciplines which shares the theme of information processing at phase transitions. This brief survey not only shows the important role that criticality plays in spatio-temporal information processing in a variety of non-linear systems, but may also provide some evidence of the value of critical dynamics in adapting to externally applied constraints — an important area for neural networks which has not been addressed as yet here.

In the previous section self-organising criticality was examined as part of a specific mechanism of dissipative structure formation in the cortex. Evidence that networks which are broadly similar to cortex can exhibit self-organising criticality together with l/f power spectra recorded in vivo suggest that the cortex may indeed have critical dynamics. Regardless of dissipative structure formation, what implications does this have for the information processing properties of the cortex? A critically organised system is at the boundary between order and chaos. Information processing at such boundaries or phase transitions has been considered by a number of authors in a variety of contexts. These studies suggest critical dynamics may be fundamental to the emergence of optimal computation abilities in distributed systems.

Information processing at phase transitions in non-linear dynamical systems in general has been studied by Crutchfield and Young (1990). Crutchfield and Young's technique results in the construction of a linguistic symbolic description of a system's behaviour, and it is therefore slightly more complex than techniques which aim to produce minimal Markov models (e.g., Destexhe, Nicolis & Nicolis, 1989). The aim of symbolic dynamics is to find the most compact descriptions of the behaviour of a system. For example, if a series of states always follow one another they can be collapsed into a single state for the purposes of symbolic description. If the behaviour of a non-linear dynamical system is known to infinite precision and the system is noise-free, then chaos requires the lengthiest descriptions. If, however, the system is noisy or its state can only be measured to finite precision so that symbols have a minimum "grain size" then the maximally complex symbolic description for a given system lies in the region of the phase transition from order to chaos. The description of highly disordered behaviour becomes simple if system states are only known to finite precision because it can only be described in simple
probabilistic terms rather than complex but completely precise terms. For example, in complete disorder the symbolic description of behaviour reduces to a single symbolic state since the probability of transitions from any partition of state space to any other are equally likely, hence all the sets of transitions associated with all these partitions are equivalent and the partitions can therefore be reduced to a single symbolic state from which there can be no transition. In finding the most compact symbolic descriptions, Crutchfield and Young (1990) also discovered that the behaviour of the system they investigated near a phase transition required a qualitatively different type of grammatical description to that required away from the phase transition. I have recently argued (Kentridge, submitted) that Crutchfield and Young’s work can provide a link between neural networks and more traditional symbolic approaches to artificial intelligence. The qualitative change in the linguistic complexity of symbolic dynamics at phase transitions to a grammar beyond Chomsky’s (1963) class 1 suggests that in order to be capable of natural language understanding a system such as a brain or neural network needs to be operating at a phase transition.

The results of the simulations reported in this chapter suggest that cortex-like neural networks may be able to respond selectively to different spatio-temporal patterns at phase transitions. Crutchfield and Young’s (1990) work emphasises the importance of phase transitions in non-linear dynamical systems in general. There is, however, a considerable gulf between these lines of work. There is no evidence as yet that neural networks with spatially localised connections require non-trivial grammatical descriptions of the behaviour of the dissipative structures they can support. Likewise, although Crutchfield and Young’s (1990) conclusions apply to dynamical systems in general, they put no emphasis on the spatial properties of systems which has been shown to be so important in determining the types of behaviour found in neural networks. A number of studies of information processing at phase transitions have, however, focused on cellular automata, which are broadly similar to neural networks with spatially localised connections in that they involve a temporally extended interaction between many simple components with spatially localised influence. They differ somewhat in that the nature of interactions between components in cellular automata can be much more arbitrary than those in neural nets. Li, Packard and Langton (1990) suggest that the production of “complex” behaviour in cellular automata occurs at a phase transition in the rule-space governing the behaviour of cellular automata. This is interesting from the present standpoint since “complex” behaviour in cellular automata, that is Wolfram’s (1983) class IV behaviour, shares a number of
features with behaviour which produces propagating dissipative structures. It is broadly defined as non-chaotic behaviour involving long transients, propagating structures, a lack of statistical convergence, and relatively low entropy. Other studies have addressed information processing more directly by studying the ability of cellular automata to produce particular types of behaviour. Packard (1988) has studied the selection of cellular automata rules by a genetic algorithm (see, e.g., Holland, 1992) whose fitness criterion was the production of a particular computation. The target involves producing some simple asymptotic behaviour (e.g., total quiescence) on the basis of some simple feature of the system's initial state (e.g., a threshold on the density of initially active cells). Once again, the rules selected tend to be those near the transition to chaos that produces type IV behaviour. Packard conjectures that this selection is largely independent of the target of adaptation; rather, subchaotic rules communicate too little information about the initial state of the system over time (they result in either frozen states or simple oscillations) while the behaviour of automata with chaotic rules involves the transmission of too much information. In the chaotic automata behaviour is persistently dependent on the consequences of all the details of the initial state, and selection of the consequences of particular features is impossible.

These studies of cellular automata show that phase transitions are crucial to the production of complex adaptive behaviour in a variety of spatially distributed non-linear dynamical systems. At least in part this is because these systems can produce coherent spatio-temporal structures which can respond to minor changes in state but whose behaviour is not so sensitive that selection processes cannot operate. Similar selective processes may allow dissipative structures in neural networks to take part in adaptive information processing.

One final line of evidence showing how phase transitions and adaptive behaviour are linked is modelling of gene adaptation and expression using random Boolean networks (see, e.g., Kauffmann, 1989a,b, 1991, for recent reviews). Random Boolean networks are related both to neural networks and cellular automata. They are automata with rules limited to Boolean functions but whose components have non-local interactions similar to many neural networks. Kaufmann's work on these systems suggests once again that adaptation may be most successful at the edge of chaos. Random Boolean networks in which each element receives only two inputs exhibit very ordered behaviour — large proportions of the net become frozen. However, when the number of inputs is increased to three the behaviour of the network becomes quite chaotic. This transition can be approached less abruptly by using probabilistically biased Boolean functions or by increasing the proportion of
rules in which one input dominates all others (e.g., ones which involve a logical
OR with the states of one input — known as canalising functions). In studies of
adaptation the proportion of these canalising rules which produce behaviour at
the edge of chaos increases as selection proceeds. These studies make certain
predictions about the eventual overall size of effect in a system produced by
initial perturbations. These predictions are borne out when the number of genes
whose expression is altered when other genes are mutated is examined in
_Drosophila._

The work I have described above suggests quite strongly that critical
dynamics may be a common requirement in efficient distributed information-
processing systems. Although time-dependent information processing can be
achieved in non-critical neural networks, studies of adaptation in critical
systems described above suggests that evolutionary optimisation may well act
to produce a solution based on critical dynamics. As I have noted, critical
dynamics may be a necessity in dynamical systems which implement the
higher symbolic processes of language and thought. The fact that the critical
conditions necessary for time-dependent responses to stimuli to occur can be
controlled simply by modulating diffuse background activity raises interesting
possibilities for the neural implementation of attention. In this chapter I have
reported a few early steps in the investigation of information processing in
neural nets with critical dynamics. The forms of representation and adaptation
in these systems are as yet unknown. This puts this approach at a disadvantage
compared to some of the artificial neural networks described at the start of this
paper. Nevertheless, I would argue that in the long run networks with critical
dynamics will provide better models of the neural basis of psychological
processes than traditional artificial neural network systems.