Contents

1 Introduction **8**

1.1 Introduction . . . 8
1.2 Neuronal network approaches versus connectionism . . . 8
1.3 Neurons in the brain, and their representation in neuronal networks . . . 8
1.4 A formalism for approaching the operation of single neurons in a network . . . 8
1.5 Synaptic modification . . . 8
1.6 Long-term potentiation and long-term depression as biological models of synaptic modification that occur in the brain . . . 8
1.7 Distributed representations . . . 8
1.8 Introduction to three simple neuronal network architecture . . . 8
1.9 System-level analysis of brain functions . . . 8

2 Pattern association memory **23**

2.1 Architecture and operation . . . 23
2.2 A simple model . . . 23
2.3 The vector interpretation . . . 23
2.4 Properties . . . 23
2.4.1 Generalization . . . 23
2.4.2 Graceful degradation or fault tolerance . . . 23
2.4.3 The importance of distributed representation for pattern associators . . . 23
2.4.4 Prototype extraction, extraction of central tendency, and noise reduction . . . 23
2.4.5 Speed . . . 23
2.4.6 Local learning rule . . . 23
2.4.7 Capacity
– Linear associative neuronal networks
– Associative neuronal networks with non-linear neurons . . . 23
2.4.8 Interference . . . 23
2.4.9 Expansion recoding . . . 23
2.4.10 Implications of different types of coding for storage in pattern associators . . . 23

3 Autoassociation memory **42**

3.1 Architecture and operation . . . 42
3.1.1 Learning . . . 42
3.2 Recall . . . 42
3.3 Introduction to the analysis of the operation of autoassociative networks . . . 42
3.4 Properties . . . 42
3.4.1 Completion . . . 42
3.4.2 Generalization . . . 42
3.4.3 Graceful degradation or fault tolerance . . . 42
3.4.4 Prototype extraction, extraction of central tendency, and noise reduction . . . 42
3.4.5 Speed . . . 42

3.4.6 Local learning rule . . . 42
3.4.7 Capacity . . . 42
3.4.8 Context . . . 42
3.4.9 Mixture states . . . 42
3.4.10 Memory for sequences . . . 42
3.5 Use of the autoassociation networks in the brain . . . 42

4 Competitive networks, including self-organizing maps 54
4.1 Function . . . 54
4.2 Architecture and algorithm . . . 54
4.2.1 Architecture . . . 54
4.2.2 Algorithm . . . 54
4.3 Properties . . . 54
4.3.1 Feature discovery by self-organization . . . 54
4.3.2 Removal of redundancy . . . 54
4.3.3 Orthogonalization and categorization . . . 54
4.3.4 Sparsification . . . 54
4.3.5 Capacity . . . 54
4.3.6 Separation of non-linearly separable patterns . . . 54
4.3.7 Stability . . . 54
4.3.8 Frequency of presentation . . . 54
4.3.9 Comparison to principal component analysis (PCA) and cluster analysis . . . 54
4.4 Utility of competitive networks in information processing by the brain . . . 54
4.4.1 Orthogonalization . . . 54
4.4.2 Sparsification . . . 54
4.4.3 Brain systems in which competitive networks may be used for orthogonalization and sparsification . . . 54
4.4.4 Removal of redundancy . . . 54
4.4.5 Feature analysis and preprocessing . . . 54
4.5 Guidance of competitive learning . . . 54
4.6 Topographic map formation . . . 54
4.7 Radial basis function networks . . . 54
4.8 Appendix on the algorithms used in competitive networks . . . 54
4.8.1 Normalization of the inputs . . . 54
4.8.2 Normalization of the length of the synaptic weight vector on each dendrite . . . 54
4.8.3 Non-linearity in the learning rule . . . 54
4.8.4 Competition . . . 54
4.8.5 Soft competition . . . 54
4.8.6 Untrained neurons . . . 54
4.8.7 Large competitive nets: further aspects . . . 54

5 Error-correcting networks: perceptrons, backpropagation of error in multilayer networks, and reinforcement learning algorithms **75**
5.1 Perceptrons and one-layer error correcting networks 75
5.1.1 Architecture and general description 75
5.1.2 Generic algorithm (for one-layer network taught by error correction) 75
5.1.3 Variations on the single-layet errorcorrecting network:
– Rosenblat's (1961) perceptron
– Adaline
– Adaptive filter
– Capability and limitations of single-layer error-correcting networks
– Output neurons with continuous values, random patterns
– Output neurons with binary threshold activation functions
– Gradient descent for neurons with continuous-valued outputs 75
5.1.4 Properties
– Generalization
– Graceful degradation or fault tolerance
– Prototype extraction, extraction of central tendency, and noise reduction
– Speed
– Non-local learning rule
– Interference
– expansion recoding 75
5.1.5 Utility of a single-layer error correcting networks in information processing by the brain 75
5.2 Multilayer perceptrons: backpropagation of error networks 75
5.2.1 Introduction 75
5.2.2 Architecture and algorithm 75
5.2.3 Properties of multilayer networks trained by error backpropagation
– Arbitrary mapping
– Fast operation
– Learning speed
– Number of hidden neurons and generalization
– Non-local learning rule 75
5.3 Reinforcement learning 75
5.3.1 Associative reward-penalty algorithm of Barto and Sutton
– Architecture
– Operation 75
5.4 Contrastive Hebbian learning: the Boltzmann machine 75
5.5 Conclusion 75

6 The hippocampus and memory **95**
6.1 What functions are performed by the hippocampus? Evidence from the effect of damage to the brain 95

6.2 Relation between spatial and non-spatial aspects of hippocampal function, episodic memory, and long-term semantic memory 95
6.3 Neurophysiology of the hippocampus 95
6.3.1 Primate hippocampal cells which respond to a combination of spatial ('where') and object ('what') information 95
6.3.2 Spatial views are encoded by primate hippocampal neurons 95
6.4 Architecture of the hippocampus 95
6.5 Plasticity 95
6.6 Outline of the computational hypothesis of hippocampal operation 95
6.7 The CA3 network as an autoassociative memory 95
6.7.1 Analysis of storage capacity 95
6.7.2 The requirement of the input systems for efficient storage of new information 95
6.7.3 The role of the perforant path input in initiating retrieval 95
6.7.4 Predictions arising from the analysis of CA3 95
6.7.5 Dynamics and the speed of operation of autoassociation in the brain 95
6.8 The dentate granule cells 95
6.9 The CA1 network 95
6.9.1 Recoding in CA1 95
6.9.2 Preserving the information content of CA3 patterns 95
6.9.3 The perforant path projection to CA1 95
6.10 Backprojection to the neocortex, and recall from the hippocampus 95
6.10.1 Capacity limitations of the hippocampus, and its implications for the possibility that the hippocampus acts as a buffer store, and for the gradient of retrograde amnesia 95
6.10.2 The recall of information from the hippocampus to the cerebral cortex 95
6.10.3 Quantitative constraints on the backprojection connectivity 95
6.10.4 Simulation of hippocampal operation 95
6.11 Other suggestions concerning hippocampal operation and function 95

7 Pattern association in the brain: amygdala and orbitofrontal cortex 136
7.1 Pattern association involved in learning associations between sensory representations and reward or punishment 136
7.1.1 Emotion and motivation
– Functions of emotion 136
7.1.2 Processing in the cortical visual areas
– Why reward and punishment association of stimuli are not represented early in information processing in the primate brain 136
7.1.3 The amygdala
– Connections
– Effect of amygdala lesions
– Neuronal activity in the primate amygdala to reinforcing stimuli
– Responses of these amygdala neurons to novel stimuli which are reinforcing 136

7.1.4 Orbitofrontal cortex
– Effect of orbitofrontal cortex lesions
– Connections
– Neuronal activity in the primate orbitofrontal cortex to reinforcing stimuli 136
7.1.5 Basal forebrain cholinergic systems, and the noradrenergic system
– Basal forebrain cholinergic neurons
– Noradrenergic neurons 136
7.2 Backprojections in the cerebral cortex — the role of pattern association 136
7.3 Pattern association within the hippocampus 136

8 Cortical networks for invariant pattern recognition 165
8.1 Introduction 165
8.2 Preprocessing in the visual pathways as far as the primary visual cortex 165
8.3 Processing to the inferior temporal cortex in the primate visual system 165
8.3.1 Hierarchical feedforward organization 165
8.3.2 Receptive field size and translation invariance 165
8.3.3 Size and spatial frequency invariant 165
8.3.4 Combinations of features in the correct spatial configuration 165
8.3.5 A view-independent representation 165
8.3.6 Distributed encoding 165
8.3.7 Speed of processing 165
8.3.8 Rapid learning 165
8.4 VisNet — an approach to biologically plausible visual object identification 165
8.5 The cognitron and neocognitron 165
8.6 An alternative approach for object-based representations 165
8.7 Syntactic bindings of separate neuronal ensembles by synchronization 165
8.8 Different processes involved in different types of object identification 165

9 Motor systems: cerebellum and basal ganglia 189
9.1 The cerebellum 189
9.1.1 Architecture of the cerebellum
– The connections of the parallel fibres onto the Purkinje cells
– The climbing fibre input to the Purkinje cell
– The mossy fibre to granule cell connectivity 189
9.1.2 Modifiable synapses of parallel fibres onto Purkinje cell dendrides 189
9.1.3 The cerebellum as a perceptron 189
9.1.4 Systems-level analysis of cerebellar function
– The vestibulo-ocular reflex
– The control of limb movement
– Classical conditioning of skeletal muscle responses 189
9.2 The basal ganglia 189
9.2.1 Systems-level architecture of the basal ganglia 189
9.2.2 Systems-level analysis of the basal ganglia: effect of striatal lesions 189

9.2.3 Systems-level analysis of the basal ganglia: neuronal activities in different parts of the striatum
– Tail of the caudate nucleus, and posteroventral putamen
– Posterior putamen
– Head of the caudate nucleus
– Anterior putamen
– Ventral striatum . . . 189
9.2.4 What computations are performed by the basal ganglia? . . . 189
9.2.5 How do the basal ganglia perform their computations?
– Interaction between neurons and selection of output
– Convergent mapping within the basal ganglia
– Discussion . . . 189

10 Cerebral neocortex **227**
10.1 The fine structure and connectivity of the neocortex . . . 227
10.1.1 Excitatory cells and connections . . . 227
10.1.2 Inhibitory cells and connections . . . 227
10.1.3 Quantitative aspects of cortical architecture . . . 227
10.1.4 Functional pathways through the cortical layers . . . 227
10.1.5 The scale of lateral excitatory and inhibitory effects, and the concept of modules 227
10.2 Theoretical significance of backprojection in the neocortex . . . 227
10.2.1 Architecture . . . 227
10.2.2 Learning . . . 227
10.2.3 Recall . . . 227
10.2.4 Semantic priming . . . 227
10.2.5 Attention . . . 227
10.2.6 Autoassociative storage, and constraint satisfaction . . . 227
10.2.7 Backprojection from the primary visual cortex, V1, to the lateral geniculate nucleus . . . 227
10.3 Cortical short term memory systems and attractor networks . . . 227
10.4 Coding in the cortex . . . 227
10.4.1 Distributed representations evident in the firing rate distributions . . . 227
10.4.2 The representation of information in the responses of single neurons to a set of stimuli . . . 227
10.4.3 The representation of information in the responses of a population of cortical visual neurons . . . 227

10.4.4 Advantages of the distributed representations found of objects for brain processing
– Exponentially high coding capacity
– Ease with which the code can be read by receiving neurons: the compactness of the distributed representation
– Higher resistance to noise
– Generalization
– Completion
– Gradual degradation or fault tolerance
– Speed of readout of the information . . . 227
10.4.5 Should one neuron be as discriminative as the whole organism, in object encoding systems? . . . 227
10.4.6 Temporal encoding in a spike train of a single neuron . . . 227
10.4.7 Temporal synchronization of the responses of different cortical neurons . . . 227
10.4.8 Conclusions on cortical encoding . . . 227
10.5 Functions of different cortical areas . . . 227
10.5.1 Motor and premotor cortices, and somatosensory cortical areas . . . 227
10.5.2 Parietal cortex . . . 227
10.5.3 Prefrontal cortex . . . 227
10.5.4 Temporal lobe: the organization of representations in higher parts of the cortical visual system, and the representation of semantic knowledge . . . 227
10.6 Principles of cortical computation . . . 227

A1 Introduction to linear algebra for neural networks 283
A1.1 Vectors . . . 283
A1.1.1 The inner or dot product of two vectors . . . 283
A1.1.2 The length of a vector . . . 283
A1.1.3 Normalizing the length of a vector . . . 283
A1.1.4 The angle between two vectors: the normalized dot product . . . 283
A1.1.5 The outer product of two vectors . . . 283
A1.2 Linear and non-linear systems . . . 283
A1.2.1 Linear combinations of vectors, linear independence, and linear separability . 283
A1.2.2 Application to understanding simple neural networks . . . 283

A2 Information theory 292

A3 Pattern associators 318

A4 Autoassociators 337

A5 Recurrent dynamics 351