Modeling brain function
Modeling brain function
The world of attractor neural networks

DANIEL J. AMIT

Racah Institute of Physics
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td></td>
<td>xiii</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Philosophy and Methodology</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Reduction to physics and physics modeling analogues</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Methods for mind and matter</td>
<td>3</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Some methodological questions</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>Neurophysiological Background</td>
<td>9</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Building blocks for neural networks</td>
<td>9</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Dynamics of neurons and synapses</td>
<td>12</td>
</tr>
<tr>
<td>1.2.3</td>
<td>More complicated building blocks</td>
<td>15</td>
</tr>
<tr>
<td>1.2.4</td>
<td>From biology to information processing</td>
<td>17</td>
</tr>
<tr>
<td>1.3</td>
<td>Modeling Simplified Neurophysiological Information</td>
<td>18</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Neuron as perceptron and formal neuron</td>
<td>18</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Digression on formal neurons and perceptrons</td>
<td>20</td>
</tr>
<tr>
<td>1.3.3</td>
<td>Beyond the basic perceptron</td>
<td>25</td>
</tr>
<tr>
<td>1.3.4</td>
<td>Building blocks for attractor neural networks (ANN)</td>
<td>27</td>
</tr>
<tr>
<td>1.4</td>
<td>The Network and the World</td>
<td>31</td>
</tr>
<tr>
<td>1.4.1</td>
<td>Neural states, network states and state space</td>
<td>31</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Digression on the relation between measures</td>
<td>33</td>
</tr>
<tr>
<td>1.4.3</td>
<td>Representations on network states</td>
<td>35</td>
</tr>
<tr>
<td>1.4.4</td>
<td>Thinking about output mechanism</td>
<td>38</td>
</tr>
<tr>
<td>1.5</td>
<td>Spontaneous Computation vs. Cognitive Processing</td>
<td>44</td>
</tr>
<tr>
<td>1.5.1</td>
<td>Input systems, transducers, transformers</td>
<td>44</td>
</tr>
<tr>
<td>1.5.2</td>
<td>ANN’s as computing elements — a position</td>
<td>45</td>
</tr>
</tbody>
</table>
## Contents

1.5.3 ANN’s and computation of mental representations ................................ 48  
Bibliography .................................................................................. 53  

2 The Basic Attractor Neural Network ........................................ 58  
2.1 Networks of Analog, Discrete, Noisy Neurons ......................... 58  
  2.1.1 Analog neurons, spike rates, two-state neural models ............. 58  
  2.1.2 Binary representation of single neuron activity .................. 63  
  2.1.3 Noisy dynamics of discrete two-state neurons ................. 65  
2.2 Dynamical Evolution of Network States ................................ 68  
  2.2.1 Network dynamics of discrete-neurons ......................... 68  
  2.2.2 Synchronous dynamics .............................................. 70  
  2.2.3 Asynchronous dynamics ........................................... 72  
  2.2.4 Sample trajectories and lessons about dynamics ............... 74  
  2.2.5 Types of trajectories and possible interpretation – a summary ...... 79  
2.3 On Attractors ........................................................................ 81  
  2.3.1 The landscape metaphor ............................................. 81  
  2.3.2 Perception, recognition and recall ................................ 84  
  2.3.3 Perception errors due to spurious states – possible role of noise ...... 85  
  2.3.4 Psychiatric speculations and images ............................... 87  
  2.3.5 The role of noise and simulated annealing ..................... 89  
  2.3.6 Frustration and diversity of attractors ........................... 91  
Bibliography .................................................................................. 95  

3 General Ideas Concerning Dynamics ........................................ 97  
3.1 The Stochastic Process, Ergodicity and Beyond ...................... 97  
  3.1.1 Stochastic equation and apparent ergodicity .................. 97  
  3.1.2 Two ways of evading ergodicity .................................. 101  
3.2 Cooperativity as an Emergent Property in Magnetic Analog .... 105  
  3.2.1 Ising model for a magnet – spin, field and interaction .......... 105  
  3.2.2 Dynamics and equilibrium properties ......................... 108  
  3.2.3 Noiseless, short range ferromagnet .............................. 112  
  3.2.4 Fully connected Ising model: real non-ergodicity ............. 119
## Contents

3.3 From Dynamics to Landscapes – The Free Energy ..... 125
  3.3.1 Energy as Lyapunov function for noiseless dynamics ..... 125
  3.3.2 Parametrized attractor distributions with noise ..... 126
  3.3.3 Free-energy landscapes – a noisy Lyapunov function ..... 127
  3.3.4 Free-energy minima, non-ergodicity, order-parameters ..... 129

3.4 Free-Energy of Fully Connected Ising Model ..... 131
  3.4.1 From minimization equation to the free-energy ..... 131
  3.4.2 The analytic way to the free-energy ..... 133
  3.4.3 Attractors at metastable states ..... 140

3.5 Symaptic Symmetry and Landscapes ..... 141
  3.5.1 Noiseless asynchronous dynamics – energy ..... 141
  3.5.2 Detailed balance for noisy asynchronous dynamics ..... 142
  3.5.3 Noiseless synchronous dynamics – Lyapunov function ..... 143
  3.5.4 Detailed balance for noisy synchronous dynamics ..... 145

3.6 Appendix: Technical Details for Stochastic Equations ..... 146
  3.6.1 The maximal eigen-value and the associated vector ..... 146
  3.6.2 Differential equation for mean magnetization ..... 147
  3.6.3 The minimization of the dynamical free-energy ..... 150
  3.6.4 Legendre transform for the free-energy ..... 152

Bibliography ..... 153

4 Symmetric Neural Networks at Low Memory Loading ..... 155
  4.1 Motivations and List of Results ..... 155
    4.1.1 Simplifying assumptions and specific questions ..... 155
    4.1.2 Specific answers for low loading of random memories ..... 158
    4.1.3 Properties of the noiseless network ..... 162
    4.1.4 Properties of the network in the presence of fast noise ..... 166
  4.2 Explicit Construction of Synaptic Efficacies ..... 169
    4.2.1 Choice of memorized patterns ..... 169
    4.2.2 Storage prescription – “Hebb’s rule” ..... 170
4.2.3 A decorrelating (but nonlocal) storage prescription ........................................... 172
4.3 Stability Considerations at Low Storage ............................................................... 174
  4.3.1 Signal to noise analysis – memories, spurious states ....................................... 174
  4.3.2 Basins of attraction and retrieval times ......................................................... 178
  4.3.3 Neurophysiological interpretation ..................................................................... 180
4.4 Mean Field Approach to Attractors ......................................................................... 181
  4.4.1 Self-consistency and equations for attractors .................................................. 181
  4.4.2 Self-averaging and the final equations .............................................................. 187
  4.4.3 Free-energy, extrema, stability ........................................................................ 189
  4.4.4 Mean-field and free-energy – synchronous dynamics ....................................... 191
4.5 Retrieval States, Spurious States – Noiseless ......................................................... 192
  4.5.1 Perfect retrieval of memorized patterns ............................................................ 192
  4.5.2 Noiseless, symmetric spurious memories ......................................................... 194
  4.5.3 Non-symmetric spurious states ...................................................................... 198
  4.5.4 Are spurious states a free lunch? .................................................................... 199
4.6 Role of Noise at Low Loading ................................................................................. 200
  4.6.1 Ergodicity at high noise levels - asynchronous ............................................... 200
  4.6.2 Just below the critical noise level .................................................................... 201
  4.6.3 Positive role of noise and retrieval with no fixed points ................................. 206
4.7 Appendix: Technical Details for Low Storage ....................................................... 208
  4.7.1 Free-energy at finite $p$ – asynchronous ......................................................... 208
  4.7.2 Free-energy and solutions – synchronous dynamics ....................................... 209
  4.7.3 Bound on magnitude of overlaps ................................................................... 211
  4.7.4 Asymmetric spurious solution ....................................................................... 212
Bibliography .................................................................................................................. 213

5 Storage and Retrieval of Temporal Sequences ......................................................... 215
  5.1 Motivations: Introspective, Biological, Philosophical ......................................... 215
    5.1.1 The introspective motivation ....................................................................... 215
    5.1.2 The biological motivation ........................................................................... 216
    5.1.3 Philosophical motivations ........................................................................... 218
  5.2 Storing and Retrieving Temporal Sequences ....................................................... 221
    5.2.1 Functional asymmetry ................................................................................. 221
    5.2.2 Early ideas for instant temporal sequences ................................................. 221
# Contents

5.3 Temporal Sequences by Delayed Synapses \hspace{1cm} 226  
- 5.3.1 A simple generalization and its motivation \hspace{1cm} 226  
- 5.3.2 Dynamics with fast and slow synapses \hspace{1cm} 229  
- 5.3.3 Simulation examples of sequence recall \hspace{1cm} 231  
- 5.3.4 Adiabatically varying energy landscapes \hspace{1cm} 235  
- 5.3.5 Bi-phasic oscillations and CPG’s \hspace{1cm} 238  
5.4 Tentative Steps into Abstract Computation \hspace{1cm} 239  
- 5.4.1 The attempt to reintroduce structured operations \hspace{1cm} 239  
- 5.4.2 ANN counting chimes \hspace{1cm} 241  
- 5.4.3 Counting network – an exercise in connectionist programming \hspace{1cm} 241  
- 5.4.4 The network \hspace{1cm} 243  
- 5.4.5 Its dynamics \hspace{1cm} 245  
- 5.4.6 Simulations \hspace{1cm} 248  
- 5.4.7 Reflections on associated cognitive psychology \hspace{1cm} 251  
5.5 Sequences Without Synaptic Delays \hspace{1cm} 253  
- 5.5.1 Basic oscillator - origin of cognitive time scale \hspace{1cm} 253  
- 5.5.2 Behavior in the absence of noise \hspace{1cm} 255  
- 5.5.3 The role of noise \hspace{1cm} 256  
- 5.5.4 Synaptic structure and underlying dynamics \hspace{1cm} 259  
- 5.5.5 Network storing sequence with several patterns \hspace{1cm} 262  
5.6 Appendix: Elaborate Temporal Sequences \hspace{1cm} 262  
- 5.6.1 Temporal sequences by time averaged synaptic inputs \hspace{1cm} 262  
- 5.6.2 Temporal sequences without errors \hspace{1cm} 266  
Bibliography \hspace{1cm} 267  

6 Storage Capacity of ANN’s \hspace{1cm} 271  
6.1 Motivation and general considerations \hspace{1cm} 271  
- 6.1.1 Different measures of storage capacity \hspace{1cm} 271  
- 6.1.2 Storage capacity of human brains \hspace{1cm} 273  
- 6.1.3 Intrinsic interest in high storage \hspace{1cm} 275  
- 6.1.4 List of results \hspace{1cm} 275  
6.2 Statistical Estimates of Storage \hspace{1cm} 278  
- 6.2.1 Statistical signal to noise analysis \hspace{1cm} 278  
- 6.2.2 Absolute informational bounds on storage capacity \hspace{1cm} 283  
- 6.2.3 Coupling (synaptic efficacies) for optimal storage \hspace{1cm} 285
6.3 Theory Near Memory Saturation ............................ 289
   6.3.1 Mean-field equations with replica symmetry ....... 289
   6.3.2 Retrieval in the absence of fast noise ............ 294
   6.3.3 Analysis of the $T = 0$ equations ............... 299
6.4 Memory Saturation with Noise and Fields .................. 304
   6.4.1 A tour in the $T$-$\alpha$ phase diagram .......... 304
   6.4.2 Effect of external fields – thresholds and PSP’s 308
   6.4.3 Fields coupled to several patterns ............... 311
   6.4.4 Some technical details related to phase diagrams 312
6.5 Balance Sheet for Standard ANN ........................... 315
   6.5.1 Limiting framework and analytic consequences .... 315
   6.5.2 Finite-size effects and basins of attraction:
       simulations ........................................... 318
6.6 Beyond the Memory Blackout Catastrophe .................. 324
   6.6.1 Bounded synapses and palimpsest memory ......... 324
   6.6.2 The $7 \pm 2$ rule and palimpsest memories ........ 328
6.7 Appendix: Replica Symmetric Theory ....................... 330
   6.7.1 The replica method ............................... 330
   6.7.2 The free-energy and the mean-field equations . . 332
   6.7.3 Marginal storage and palimpsests ................ 339
Bibliography .................................................. 342

7 Robustness - Getting Closer to Biology 345
   7.1 Synaptic Noise and Synaptic Dilution .................. 345
   7.1.1 Two meanings of robustness ...................... 345
   7.1.2 Noise in synaptic efficacies ..................... 347
   7.1.3 Random symmetric dilution of synapses .......... 352
   7.2 Non-Linear Synapses & Limited Analog Depth .......... 355
   7.2.1 Place and role of non-linear synapses .......... 355
   7.2.2 Properties of networks with clipped synapses . . 357
   7.2.3 Non-linear storage and the noisy equivalent .... 359
   7.2.4 Clipping at low storage level ................. 362
   7.3 Random vs. Functional Synaptic Asymmetry ............ 363
   7.3.1 Random asymmetry and performance quality ....... 363
   7.3.2 Asymmetry, noise and spin-glass suppression ... 366
   7.3.3 Neuronal specificity of synapses - Dale’s law ... 368
   7.3.4 Extreme asymmetric dilution ..................... 370
   7.3.5 Functional asymmetry ............................ 375
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4 Effective Cortical Cycle Times</td>
<td>375</td>
</tr>
<tr>
<td>7.4.1 Slow bursts and relative refractory period</td>
<td>375</td>
</tr>
<tr>
<td>7.4.2 Neuronal memory and expanded scenario</td>
<td>377</td>
</tr>
<tr>
<td>7.4.3 Simplified scenario for relative refractory period</td>
<td>378</td>
</tr>
<tr>
<td>7.5 Appendix: Technical Details</td>
<td>380</td>
</tr>
<tr>
<td>7.5.1 Digression - the mean-field equations</td>
<td>380</td>
</tr>
<tr>
<td>7.5.2 Dilution requirement</td>
<td>384</td>
</tr>
<tr>
<td>Bibliography</td>
<td>385</td>
</tr>
</tbody>
</table>

### 8 Memory Data Structures

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Biological and Computational Motivation</td>
<td>387</td>
</tr>
<tr>
<td>8.1.1 Low mean activity level and background-foreground asymmetry</td>
<td>387</td>
</tr>
<tr>
<td>8.1.2 Hierarchies for biology and for computation</td>
<td>388</td>
</tr>
<tr>
<td>8.2 Local Treatment of Low Activity Patterns</td>
<td>389</td>
</tr>
<tr>
<td>8.2.1 Demise of naive standard model</td>
<td>389</td>
</tr>
<tr>
<td>8.2.2 Modified ANN and a plague of spurious states</td>
<td>391</td>
</tr>
<tr>
<td>8.2.3 Constrained dynamics – monitoring thresholds</td>
<td>396</td>
</tr>
<tr>
<td>8.2.4 Properties of the constrained biased network</td>
<td>398</td>
</tr>
<tr>
<td>8.2.5 Quantity of information in an ANN with low activity</td>
<td>403</td>
</tr>
<tr>
<td>8.2.6 More effective storage of low activity (sparse) patterns</td>
<td>405</td>
</tr>
<tr>
<td>8.3 Hierarchical Data Structures in a Single Network</td>
<td>409</td>
</tr>
<tr>
<td>8.3.1 Early proposals</td>
<td>409</td>
</tr>
<tr>
<td>8.3.2 Explicit construction of hierarchy in a single ANN</td>
<td>410</td>
</tr>
<tr>
<td>8.3.3 Properties of hierarchy in a single network</td>
<td>412</td>
</tr>
<tr>
<td>8.3.4 Prosopagnosia and learning class properties</td>
<td>412</td>
</tr>
<tr>
<td>8.3.5 Multy-ancestry with many generations</td>
<td>414</td>
</tr>
<tr>
<td>8.4 Hierarchies in Multi-ANN: Generalization First</td>
<td>418</td>
</tr>
<tr>
<td>8.4.1 Organization of the data and the networks</td>
<td>418</td>
</tr>
<tr>
<td>8.4.2 Hierarchical dynamics</td>
<td>420</td>
</tr>
<tr>
<td>8.4.3 Hierarchy for image vector quantization</td>
<td>422</td>
</tr>
<tr>
<td>8.5 Appendix: Technical Details for Biased Patterns</td>
<td>423</td>
</tr>
<tr>
<td>8.5.1 Noise estimates for biased patterns</td>
<td>423</td>
</tr>
<tr>
<td>8.5.2 Mean-field equations in noiseless biased network</td>
<td>424</td>
</tr>
<tr>
<td>8.5.3 Retrieval entropy in biased network</td>
<td>424</td>
</tr>
</tbody>
</table>
Contents

8.5.4 Mean-square noise in low activity network . . . . 425
Bibliography ................................................ 426

9 Learning ..................................................... 428
9.1 The Context of Learning ................................. 428
9.1.1 General comments and a limited scope ............ 428
9.1.2 Modes, time scales and other constraints .......... 430
9.1.3 The need for learning modes ........................ 432
9.1.4 Results for learning in learning modes ............. 433
9.2 Learning in Modes ...................................... 434
9.2.1 Perceptron learning .................................. 434
9.2.2 ANN learning by perceptron algorithm ............. 438
9.2.3 Local learning of the Kohonen synaptic matrix ...... 441
9.3 Natural Learning - Double Dynamics ................. 443
9.3.1 General features ...................................... 443
9.3.2 Learning in a network of physiological neurons .... 444
9.3.3 Learning to form associations ........................ 447
9.3.4 Memory generation and maintenance ............... 450
9.4 Technical Details in Learning Models .................. 455
9.4.1 Local Iterative Construction of Projector Matrix .... 455
9.4.2 The free energy and the correlation function ....... 458
Bibliography ................................................ 458

10 Hardware Implementations of Neural Networks .... 461
10.1 Situating Artificial Neural Networks ................. 461
10.1.1 The role of hardware implementations ............ 461
10.1.2 Motivations for different designs .................. 462
10.2 The VLSI Neural Network ............................. 465
10.2.1 High density high speed integrated chip .......... 465
10.2.2 Smaller, more flexible electronic ANN's .......... 469
10.3 The Electro-Optical ANN .............................. 474
10.4 Shift Register (CCD) Implementation ................. 477
Bibliography ................................................ 479

Glossary ................................................... 481

Index ......................................................... 487
Preface

This book summarizes in some detail the ideas, techniques and results developed in the last 5-6 years in the physics community about the collective properties of large assemblies of neurons. The subject has been, and still is, a source of great excitement among physicists the world over and new original ideas are generated incessantly. This enthusiasm has produced a wealth of new concepts and new detailed results which has not gone unnoticed outside physics departments. Biologists have begun to ask themselves whether the properties that physics anticipates in neural networks can indeed be observed and whether they provide useful theoretical guides for the empirical investigation of brain activity; computer scientists would not rule out these ideas as candidates for coherent parallel processing; psychologists and neurologists have been expecting some new useful metaphors for interpreting behavioral dysfunction; cognitive scientists study the new concepts in their continued struggle with the elusiveness of processes of mind, even on the most elementary levels; and technologists have added, of course, Attractor Neural Networks to the list of future industries for sale.

One explanation for this impact of the study of neural networks seems to be in the type of new concepts that have been generated. They appear plausible upon introspection and they are based on elements with biological flavor. Another attraction is the clarity, the wealth and the detail provided by the quantitative analysis of the properties of such networks. It would have been easy and uncontroversial to write a book restricted to the technical details and the results. Physicists
Preface

would have liked it better and others would have ignored it without regret or complaint. I have set myself a more challenging task, of combining the presentation of the quantitative results, with their full technical beauty, with an attempt to communicate across disciplines. It has created multiple tensions. First, it has significantly expanded the exposition of every subject, adding a non-technical description of each result as well as an attempt to connect the result to topics in other disciplines, even if by way of speculation or metaphor. Second, it has required whole chapters, e.g., Chapter 3, to summarize for the diligent non-physicist, who may like to confront the full theoretical apparatus, an entire culture of physics. Thirdly, it implies that no uniform language could be used throughout the monograph and often the text is in English. I have tried to make it manifestly clear when words attain restricted formal meanings.

I have tried to inform every directly concerned discipline, i.e., biology and cognitive science, that we are conscious of some of our limitations. I had to inform the biologist that we know we have not done justice to his real neurons, and this had to be done without embarking on a full course in neuro-physiology. I had to admit that we have not solved any important outstanding problem in cognitive science, yet insist on the fact that we are introducing novel relevant concepts. This position crosses boundaries of theoretical arguments in cognitive science. I had to indicate that I am conscious of them, react to them, yet I could not expose the full background. The need to address so many concerned audiences and remain brief was perhaps the hardest part of the task, leaving behind many unresolved tensions.

These issues have led to the writing of the first chapter. It might have more appropriately been a chapter of conclusion. Had I had my ideal reader, he would have read this chapter superficially on first reading. Then, if the rest of the book would have kept up his interest, he would have come back to reread the introduction. A slightly less ideal reader would skip the introduction on first reading and at best read it at the end. This should partially explain why quite a few of the concepts mentioned in the Introduction are not fully clarified within this chapter, i.e., Chapter 1, and preserve some of their colloquial sense. The main part of the book is dedicated to the clarification of the new concepts which these models put forth.

The introduction is intended for three classes of imaginary readers: the biologist who is convinced that physicists do not know that the
Preface

world is complicated and therefore cannot possibly bring out of neurons anything of interest; the expert in artificial intelligence who has to be convinced that the Attractor Neural Networks are at least something new (if not something important); and the cognitive scientist-philosopher to whom I feel I owe an explicit statement of my commitments on several basic issues.

Chapter 1 is, therefore, an essay on the rest of the book, trying to identify the commitments implied by the theory. The course of events has, of course, been the reverse. During the whole period of the development of the subject, no commitments were being made while results were being derived. Success breeds responsibility, and if physics intends its results to be taken seriously, it has to commit itself. The stands expressed in this essay may be at odds with some accepted theories. This may be the result either of the fact that new concepts can be unambiguously introduced via the new approach, or because the epistemological approach is inconsistent, which is the only way, I can conceive, that it could be wrong. Such an adverse eventuality does not, of course, reflect on the correctness of the technical developments, which have withstood many theoretical and practical tests. Inasmuch as the outlook argued for in the Introduction may appear controversial, it should be considered as an opening of a discussion, rather than a conclusive state of affairs.

In order to facilitate the navigation in the maze, I have provided a very detailed table of contents which may indicate how to exercise the hopscotch (à la Julio Cortazar). Chapter 2 should be read by all non-physicists since it provides a detailed explanation of what attractors are. Physicists may choose to read only the first two sections which describe how network dynamics follows from simplified neural interaction. Chapter 3 is a pedagogical chapter (I hope), with little direct connection to neural networks. It has the following objectives:

- To the non-physicist, it should convey the idea of non-ergodicity in a noisy system, which makes cooperative or collective behavior non-trivial.

- It should make clear the connection between the treatment of free-energies and that of dynamical phenomena such as neural networks.
xvi

Preface

• It should give a hint of the context from which physicists have been drawing their intuitions.

• Some physicists may find Sections 3 and 4 useful as a different perspective on the connection between dynamics, order-parameters, free-energies etc.

Chapter 3, which has been the hardest to write, can be skipped by everyone.

Chapter 4 is part of the main story. This is a good point to start reading the book, looking back at previous chapters when certain arguments become mysterious. The first three sections are intended for general consumption, giving simple technical rule-of-thumb tools for the identification of attractors. Sections 4 and 5 are a gradual introduction to the wonderful world of mean-field theories. They can be skipped at no risk, unless one intends to analyze one’s own new model.

Chapter 5 is a step beyond simple single pattern attractors, on one possible road toward structured cognitive processes. It is almost all carried out on the technical level of the rule-of-thumb tools of Chapter 4, and could and should be followed by all readers who could master Section 3 in Chapter 4. Readers with no interest in any technicalities should read only Sections 5.1, 5.3.1, 5.4.1–5.4.4. Chapter 6 is about limits of performance. It again can be read on three levels. The totally non-technical reader can do with Sections 6.1, and 6.5 and the prose in Section 6.6. The rule-of-thumb reader should add Sections 6.2 and the full 6.6. Sections 6.3, 6.4, together with the appendix, are the pinnacle of the technical side of the book. One should have a very good motivation before embarking on these sections.

Chapter 7 describes the extent of robustness of the results of the preceding three chapters. This chapter is mostly non-technical and the few formulae spread around are for purposes of definition of the type of variation under which robustness is investigated, rather than for purposes of derivation. Yet one may get the main points by skipping over Sections 7.2.3, 7.2.4 and 7.3.4 and the appendix. Chapter 8 deals with the storage of hierarchically organized data structures. Its main points are summarized in Sections 8.1, 8.2.1, 8.2.4, 8.3.3–8.3.4, 8.4.

Chapter 9 is a description of the nascent state of a theory of learning, in the context of ANN’s. The main ideas can be found in Sections 9.1 and 9.3.1. One level of technical material deals with proofs of the convergence of learning algorithms – Sections 9.2.1, 9.2.3. Another level
Preface

is involved in the definition of some learning scenarios – Sections 9.3.2–9.3.4. Finally, Chapter 10, about hardware, is purely descriptive.

In conclusion, I should stress that this book does not represent an effort toward impartiality. It is the outgrowth of a special intense experience that I was fortunate to have in the collaboration with H. Gutfreund and H. Sompolinsky. They should have coauthored this book with me, since much of what it contains, that is solid, was born in our collective effort. Without them this book would never have come into existence. Other priorities have left me alone in the field, and I can only pray that my precious collaborators would not object too much to the context into which I have inserted our joint technical work. They should not share in the blame for speculation, metaphor and polemic.

There is a long list of credits. First to students – I. Kanter, A. Crisanti, A. Treves, M. Aharoni, Y. Stein – who worked with us at various stages. There is a major intellectual debt I owe John Hopfield and Gerard Toulouse, who have also been a source of encouragement in the process of writing the book. My ideas have also been much affected by M. Abeles, V. Braitenberg, G. Parisi, M. Virasoro, the late E. Gardner, B. Shanon, D. Andler and H. Atlan. I owe a special debt to Profs. R. Gavison, of the Law School, and D. Lehmann, of the department of Computer Science, who have read substantial parts of the manuscript and have helped me with comments and criticism. Finally, special gratitude is due to A. Herz and R. Kuhn, who carefully read the first printing of this manuscript and, with their comments, contributed so much to the improvement of the second printing.

Rome
November, 1988
If it live in your memory, begin at this line —
let me see, let me see