

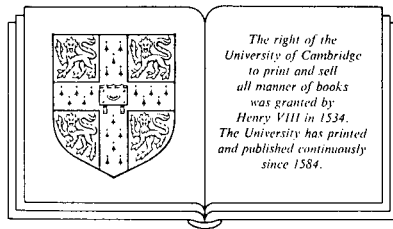
# Modeling brain function

## The world of attractor neural networks

**DANIEL J. AMIT**

*Racah Institute of Physics*

SFB 185  
Nichtlineare Dynamik



**Cambridge University Press**

Cambridge

New York   Port Chester   Melbourne   Sydney

# Contents

---

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

1.5.3	ANN's and computation of mental representations . . . . .	48
	Bibliography . . . . .	53
<b>2</b>	<b>The Basic Attractor Neural Network</b>	<b>58</b>
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
<b>3</b>	<b>General Ideas Concerning Dynamics</b>	<b>97</b>
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

- 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 Synaptic 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
<b>5</b>	<b>Storage and Retrieval of Temporal Sequences</b>	<b>215</b>
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

- 5.3 Temporal Sequences by Delayed Synapses . . . . . 226
  - 5.3.1 A simple generalization and its motivation . . . . . 226
  - 5.3.2 Dynamics with fast and slow synapses . . . . . 229
  - 5.3.3 Simulation examples of sequence recall . . . . . 231
  - 5.3.4 Adiabatically varying energy landscapes . . . . . 235
  - 5.3.5 Bi-phasic oscillations and CPG's . . . . . 238
- 5.4 Tentative Steps into Abstract Computation . . . . . 239
  - 5.4.1 The attempt to reintroduce structured operations . . . . . 239
  - 5.4.2 ANN counting chimes . . . . . 241
  - 5.4.3 Counting network – an exercise in connectionist programming . . . . . 241
  - 5.4.4 The network . . . . . 243
  - 5.4.5 Its dynamics . . . . . 245
  - 5.4.6 Simulations . . . . . 248
  - 5.4.7 Reflections on associated cognitive psychology . . . . . 251
- 5.5 Sequences Without Synaptic Delays . . . . . 253
  - 5.5.1 Basic oscillator - origin of cognitive time scale . . . . . 253
  - 5.5.2 Behavior in the absence of noise . . . . . 255
  - 5.5.3 The role of noise . . . . . 256
  - 5.5.4 Synaptic structure and underlying dynamics . . . . . 259
  - 5.5.5 Network storing sequence with several patterns . . . . . 262
- 5.6 Appendix: Elaborate Temporal Sequences . . . . . 262
  - 5.6.1 Temporal sequences by time averaged synaptic inputs . . . . . 262
  - 5.6.2 Temporal sequences without errors . . . . . 266
- Bibliography . . . . . 267
- 6 Storage Capacity of ANN's . . . . . 271**
  - 6.1 Motivation and general considerations . . . . . 271
    - 6.1.1 Different measures of storage capacity . . . . . 271
    - 6.1.2 Storage capacity of human brains . . . . . 273
    - 6.1.3 Intrinsic interest in high storage . . . . . 275
    - 6.1.4 List of results . . . . . 275
  - 6.2 Statistical Estimates of Storage . . . . . 278
    - 6.2.1 Statistical signal to noise analysis . . . . . 278
    - 6.2.2 Absolute informational bounds on storage capacity . . . . . 283
    - 6.2.3 Coupling (synaptic efficacies) for optimal storage . . . . . 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
<b>7</b>	<b>Robustness - Getting Closer to Biology</b>	<b>345</b>
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

7.4	Effective Cortical Cycle Times . . . . .	375
7.4.1	Slow bursts and relative refractory period . . . .	375
7.4.2	Neuronal memory and expanded scenario . . . .	377
7.4.3	Simplified scenario for relative refractory period . . . . .	378
7.5	Appendix: Technical Details . . . . .	380
7.5.1	Digression - the mean-field equations . . . . .	380
7.5.2	Dilution requirement . . . . .	384
	Bibliography . . . . .	385
<b>8</b>	<b>Memory Data Structures</b>	<b>387</b>
8.1	Biological and Computational Motivation . . . . .	387
8.1.1	Low mean activity level and background- foreground asymmetry . . . . .	387
8.1.2	Hierarchies for biology and for computation . . .	388
8.2	Local Treatment of Low Activity Patterns . . . . .	389
8.2.1	Demise of naive standard model . . . . .	389
8.2.2	Modified ANN and a plague of spurious states .	391
8.2.3	Constrained dynamics – monitoring thresholds .	396
8.2.4	Properties of the constrained biased network . .	398
8.2.5	Quantity of information in an ANN with low activity . . . . .	403
8.2.6	More effective storage of low activity (sparse) patterns . . . . .	405
8.3	Hierarchical Data Structures in a Single Network . . . .	409
8.3.1	Early proposals . . . . .	409
8.3.2	Explicit construction of hierarchy in a single ANN . . . . .	410
8.3.3	Properties of hierarchy in a single network . . . .	412
8.3.4	Prosopagnosia and learning class properties . . .	412
8.3.5	Multy-ancestry with many generations . . . . .	414
8.4	Hierarchies in Multi-ANN: Generalization First . . . .	418
8.4.1	Organization of the data and the networks . . .	418
8.4.2	Hierarchical dynamics . . . . .	420
8.4.3	Hierarchy for image vector quantization . . . . .	422
8.5	Appendix: Technical Details for Biased Patterns . . . .	423
8.5.1	Noise estimates for biased patterns . . . . .	423
8.5.2	Mean-field equations in noiseless biased network	424
8.5.3	Retrieval entropy in biased network . . . . .	424



8.5.4	Mean-square noise in low activity network . . . .	425
	Bibliography . . . . .	426
<b>9</b>	<b>Learning</b>	<b>428</b>
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
<b>10</b>	<b>Hardware Implementations of Neural Networks</b>	<b>461</b>
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
	<b>Glossary</b>	<b>481</b>
	<b>Index</b>	<b>487</b>