Contents

Preface  xi
Acknowledgment  xix

1  Overview  1
1.1 History of Neural Networks  1
1.2 Neural Networks in Software  2
1.2.1 Artificial Neural Network  2
1.2.2 Spiking Neural Network  3
1.3 Need for Neuromorphic Hardware  3
1.4 Objectives and Outlines of the Book  5
References  8

2  Fundamentals and Learning of Artificial Neural Networks  11
2.1 Operational Principles of Artificial Neural Networks  11
2.1.1 Inference  11
2.1.2 Learning  13
2.2 Neural Network Based Machine Learning  16
2.2.1 Supervised Learning  17
2.2.2 Reinforcement Learning  20
2.2.3 Unsupervised Learning  22
2.2.4 Case Study: Action-Dependent Heuristic Dynamic Programming  23
2.2.4.1 Actor-Critic Networks  24
2.2.4.2 On-Line Learning Algorithm  25
2.2.4.3 Virtual Update Technique  27
2.3 Network Topologies  31
2.3.1 Fully Connected Neural Networks  31
2.3.2 Convolutional Neural Networks  32
2.3.3 Recurrent Neural Networks  35
2.4 Dataset and Benchmarks  38
2.5 Deep Learning  41
2.5.1 Pre-Deep-Learning Era  41
2.5.2 The Rise of Deep Learning  41
2.5.3 Deep Learning Techniques  42
2.5.3.1 Performance-Improving Techniques 42
2.5.3.2 Energy-Efficiency-Improving Techniques 46
2.5.4 Deep Neural Network Examples 50
References 53

3 Artificial Neural Networks in Hardware 61
3.1 Overview 61
3.2 General-Purpose Processors 62
3.3 Digital Accelerators 63
3.3.1 A Digital ASIC Approach 63
3.3.1.1 Optimization on Data Movement and Memory Access 63
3.3.1.2 Scaling Precision 71
3.3.1.3 Leveraging Sparsity 76
3.3.2 FPGA-Based Accelerators 80
3.4 Analog/Mixed-Signal Accelerators 82
3.4.1 Neural Networks in Conventional Integrated Technology 82
3.4.1.1 In/Near-Memory Computing 82
3.4.1.2 Near-Sensor Computing 85
3.4.2 Neural Network Based on Emerging Non-volatile Memory 88
3.4.2.1 Crossbar as a Massively Parallel Engine 89
3.4.2.2 Learning in a Crossbar 91
3.4.3 Optical Accelerator 93
3.5 Case Study: An Energy-Efficient Accelerator for Adaptive Dynamic Programming 94
3.5.1 Hardware Architecture 95
3.5.1.1 On-Chip Memory 95
3.5.1.2 Datapath 97
3.5.1.3 Controller 99
3.5.2 Design Examples 101
References 108

4 Operational Principles and Learning in Spiking Neural Networks 119
4.1 Spiking Neural Networks 119
4.1.1 Popular Spiking Neuron Models 120
4.1.1.1 Hodgkin-Huxley Model 120
4.1.1.2 Leaky Integrate-and-Fire Model 121
4.1.1.3 Izhikevich Model 121
4.1.2 Information Encoding 122
4.1.3 Spiking Neuron versus Non-Spiking Neuron 123
4.2 Learning in Shallow SNNs 124
4.2.1 ReSuMe 124
4.2.2 Tempotron 125
4.2.3 Spike-Timing-Dependent Plasticity 127
4.2.4 Learning Through Modulating Weight-Dependent STDP in Two-Layer Neural Networks 131
4.2.4.1 Motivations 131
4.2.4.2 Estimating Gradients with Spike Timings 131
4.2.4.3 Reinforcement Learning Example 135
4.3 Learning in Deep SNNs 146
4.3.1 SpikeProp 146
4.3.2 Stack of Shallow Networks 147
4.3.3 Conversion from ANNs 148
4.3.4 Recent Advances in Backpropagation for Deep SNNs 150
4.3.5 Learning Through Modulating Weight-Dependent STDP in Multilayer
   Neural Networks 151
   4.3.5.1 Motivations 151
   4.3.5.2 Learning Through Modulating Weight-Dependent STDP 151
   4.3.5.3 Simulation Results 158
5 Hardware Implementations of Spiking Neural Networks 173
   5.1 The Need for Specialized Hardware 173
   5.1.1 Address-Event Representation 173
   5.1.2 Event-Driven Computation 174
   5.1.3 Inference with a Progressive Precision 175
   5.1.4 Hardware Considerations for Implementing the Weight-Dependent STDP
       Learning Rule 181
   5.1.4.1 Centralized Memory Architecture 182
   5.1.4.2 Distributed Memory Architecture 183
   5.2 Digital SNNs 186
   5.2.1 Large-Scale SNN ASICs 186
   5.2.1.1 SpiNNaker 186
   5.2.1.2 TrueNorth 187
   5.2.1.3 Loihi 191
   5.2.2 Small/Moderate-Scale Digital SNNs 192
   5.2.2.1 Bottom-Up Approach 192
   5.2.2.2 Top-Down Approach 193
   5.2.3 Hardware-Friendly Reinforcement Learning in SNNs 194
   5.2.4 Hardware-Friendly Supervised Learning in Multilayer SNNs 199
   5.2.4.1 Hardware Architecture 199
   5.2.4.2 CMOS Implementation Results 205
   5.3 Analog/Mixed-Signal SNNs 210
   5.3.1 Basic Building Blocks 210
   5.3.2 Large-Scale Analog/Mixed-Signal CMOS SNNs 211
   5.3.2.1 CAVIAR 211
   5.3.2.2 BrainScaleS 214
   5.3.2.3 Neurogrid 215
   5.3.3 Other Analog/Mixed-Signal CMOS SNN ASICs 216
   5.3.4 SNNs Based on Emerging Nanotechnologies 216
   5.3.4.1 Energy-Efficient Solutions 217
   5.3.4.2 Synaptic Plasticity 218
   5.3.5 Case Study: Memristor Crossbar Based Learning in SNNs 220
   5.3.5.1 Motivations 220
   5.3.5.2 Algorithm Adaptations 222