Fpga Implementations Of Neural Networks

This book is the first of a two-volume set that constitutes the refereed proceedings of the 17th International Conference on Artificial Neural Networks, ICANN 2007, held in Porto, Portugal, September 2007. Coverage includes advances in neural network learning methods, advances in neural network architectures, neural dynamics and complex systems, data analysis, evolutionary computing, agents learning, as well as temporal synchronization and nonlinear dynamics in neural networks.

This work was to establish whether it was possible to achieve a reasonable speedup by implementing FPGA based Hopfield neural networks for some simple constraint satisfaction problems. The results are significant - our initial implementation using standard Xilinx FPGAs yielded 2-3 orders of magnitude speedup over the Sun Blade 2000 workstation comes with 1.2-GHz version of the 64-bit UltraSPARC III Cu processor. The main problem with the work to date is that the problems are both unrealistically small and simplistic. That is the constraints on the N-Queen problem are simpler than those found in many real world scheduling applications. Thus, it is not clear whether we will be able to optimize the neuron structure for more complex problems since the weights matrix may not contain as many zero elements. Thus a new method for speed improvement of Hopfield neural networks for solving constraint satisfaction problems using Field Programmable Gate Arrays (FPGAs) was proposed and implemented.

This book covers diverse aspects of advanced computer and communication engineering, focusing specifically on industrial and manufacturing theory and applications of electronics, communications, computing and information technology. Experts in research, industry, and academia present the latest developments in technology, describe applications involving cutting-edge communication and computer systems, and explore likely future trends. In addition, a wealth of new algorithms that assist in solving computer and communication engineering problems are presented. The book is based on presentations given at ICOCOE 2015, the 2nd International Conference on Communication and Computer Engineering. It will appeal to a wide range of professionals in the field, including telecommunication engineers, computer engineers and scientists, researchers, academics and students.

This volume constitutes the proceedings of the 8th International Conference on Hybrid Artificial Intelligent Systems, HAIS 2013, held in Salamanca, Spain, in September 2013. The 68 papers published in this volume were carefully reviewed and selected from 218 submissions. They are organized in topical sessions on Agents and Multi Agents Systems; HAIS Applications; Classification and Cluster Analysis; Data Mining and Knowledge Discovery; Video and Image Analysis; Bio-inspired Models and Evolutionary Computation; Learning Algorithms; Systems, MAN, and Cybernetics; Hybrid Intelligent Systems for Data Mining and Applications; Metaheuristics for Combinatorial Optimization and Modelling Complex Systems.

During the 1980s and early 1990s there was signi?cant work in the design and implementation of hardware neurocomputers. Nevertheless, most of these efforts may be judged to have been unsuccessful: at no time have have ha- ware neurocomputers been in wide use. This lack of success may be largely attributed to the fact that earlier work was almost entirely aimed at developing custom neurocomputers, based on ASIC technology, but for such niche - eas this technology was never suf?ciently developed or competitive enough to justify large-scale adoption. On the other hand, gate-arrays of the period m- tioned were never large enough nor fast enough for serious arti?cial-neur- network (ANN) applications. But technology has now improved: the capacity and performance of current FPGAs are such that they present a much more realistic alternative. Consequently neurocomputers based on FPGAs are now a much more practical proposition than they have been in the past. This book summarizes some work towards this goal and consists of 12 papers that were selected, after review, from a number of submissions. The book is nominally divided into three parts: Chapters 1 through 4 deal with foundational issues; Chapters 5 through 11 deal with a variety of implementations; and Chapter 12 looks at the lessons learned from a large-scale project and also reconsiders design issues in light of current and future technology.

Thinking Machines: Machine Learning and Its Hardware Implementation covers the theory and application of machine learning, neuromorphic computing and neural networks. This is the first book that focuses on machine learning accelerators and hardware development for machine learning. It presents not only a summary of the latest trends and examples of machine learning hardware and basic knowledge of machine learning in general, but also the main issues involved in its implementation. Readers will learn what is required for the design of machine learning hardware for neuromorphic computing and/or neural networks. This is a recommended book for those who have basic knowledge of machine learning or those who want to learn more about the current trends of machine learning. Presents a clear understanding of various available machine learning hardware accelerator solutions that can be applied to selected machine learning algorithms Offers key insights into the development of hardware, from algorithms, software, logic circuits, to hardware accelerators Introduces the baseline characteristics of deep neural network models that should be treated by hardware as well Presents readers with a thorough review of past research and products, explaining how to design through ASIC and FPGA approaches for target machine learning models Surveys current trends and models in neuromorphic computing and neural network hardware architectures Outlines the strategy for advanced hardware development through the example of deep learning accelerators

This volume of Advances in Intelligent Systems and Computing highlights key scientific achievements and innovations in all areas of automation, informatization, computer science, and artificial intelligence. It gathers papers presented at the IITI 2017, the Second International Conference on Intelligent Information Technologies for Industry, which

was held in Varna, Bulgaria on September 14–16, 2017. The conference was jointly co-organized by Technical University of Varna (Bulgaria), Technical University of Sofia (Bulgaria), VSB Technical University of Ostrava (Czech Republic) and Rostov State Transport University (Russia). The IITI 2017 brought together international researchers and industrial practitioners interested in the development and implementation of modern technologies for automation, informatization, computer science, artificial intelligence, transport and power electrical engineering. In addition to advancing both fundamental research and innovative applications, the conference is intended to establish a new dissemination platform and an international network of researchers in these fields.

This book constitutes the refereed proceedings of the 13th International Conference on Field-Programmable Logic and Applications, FPL 2003, held in Lisbon, Portugal in September 2003. The 90 revised full papers and 56 revised poster papers presented were carefully reviewed and selected from 216 submissions. The papers are organized in topical sections on technologies and trends, communications applications, high level design tools, reconfigurable architecture, cryptographic applications, multi-context FPGAs, low-power issues, run-time reconfiguration, compilation tools, asynchronous techniques, bio-related applications, codesign, reconfigurable fabrics, image processing applications, SAT techniques, application-specific architectures, DSP applications, dynamic reconfiguration, SoC architectures, emulation, cache design, arithmetic, bio-inspired design, SoC design, cellular applications, fault analysis, and network applications.

FPGA Implementations of Neural NetworksSpringer Science & Business Media

The authoritative reference on NEURON, the simulation environment for modeling biological neurons and neural networks that enjoys wide use in the experimental and computational neuroscience communities. This book shows how to use NEURON to construct and apply empirically based models. Written primarily for neuroscience investigators, teachers, and students, it assumes no previous knowledge of computer programming or numerical methods. Readers with a background in the physical sciences or mathematics, who have some knowledge about brain cells and circuits and are interested in computational modeling, will also find it helpful. The NEURON Book covers material that ranges from the inner workings of this program, to practical considerations involved in specifying the anatomical and biophysical properties that are to be represented in models. It uses a problem-solving approach, with many working examples that readers can try for themselves.

This book provides a structured treatment of the key principles and techniques for enabling efficient processing of deep neural networks (DNNs). DNNs are currently widely used for many artificial intelligence (AI) applications, including computer vision, speech recognition, and robotics. While DNNs deliver state-of-the-art accuracy on many AI tasks, it comes at the cost of high computational complexity. Therefore, techniques that enable efficient processing of deep neural networks to improve metrics—such as energy-efficiency, throughput, and latency—without sacrificing accuracy or increasing hardware costs are critical to enabling the wide deployment of DNNs in AI systems. The book includes background on DNN processing; a description and taxonomy of hardware architectural approaches for designing DNN accelerators; key metrics for evaluating and comparing different designs; features of the DNN processing that are amenable to hardware/algorithm co-design to improve energy efficiency and throughput; and opportunities for applying new technologies. Readers will find a structured introduction to the field as well as a formalization and organization of key concepts from contemporary works that provides insights that may spark new ideas.

This book constitutes the refereed proceedings of the 8th International Workshop on Artificial Neural Networks, IWANN 2005, held in Vilanova i la Geltrú, Barcelona, Spain in June 2005. The 150 revised papers presented - including the contribution of three invited speakers - were carefully reviewed and selected from 240 submissions for inclusion in the book and address the following topics: mathematical and theoretical methods, evolutionary computation, neurocomputational inspired models, learning and adaptation, radial basic functions structures, self-organizing networks and methods, support vector machines, cellular neural networks, hybrid systems, neuroengineering and hardware implementations, pattern recognition, perception and robotics and applications in a broad variety of fields.

"This thesis describes the Field Programmable Gate Array (FPGA) implementations of two powerful techniques of Computational Intelligence (CI), the Particle Swarm Optimization algorithm (PSO) and the Neural Network (NN). Particle Swarm Optimization (PSO) is a popular population-based optimization algorithm. While PSO has been shown to perform well in a large variety of problems, PSO is typically implemented in software. Population-based optimization algorithms such as PSO are well suited for execution in parallel stages. This allows PSO to be implemented directly in hardware and achieve much faster execution times than possible in software. In this thesis, a pipelined architecture for hardware PSO implementation is presented. Benchmark functions solved by software and FPGA hardware PSO implementations are compared. NNs are inherently parallel, with each layer of neurons processing incoming data independently of each other. While general purpose processors have reached impressive processing speeds, they still cannot fully exploit this inherent parallelism due to their sequential architecture. In order to achieve the high neural network throughput needed for real-time applications, a custom hardware design is needed. In this thesis, a digital implementation of an NN is developed for FPGA implementation. The hardware PSO implementation is designed using only VHDL, while the NN hardware implementation is designed using Xilinx System Generator. Both designs are synthesized using Xilinx ISE and implemented on the Xilinx Virtex-II Pro FPGA Development Kit"--Abstract, leaf iii.

Back-Propagation (BP) Algorithm is one of the efficient learning algorithms for the training of Artificial Neural Networks (ANN). The efficient hardware implementation of the BP Algorithm can find its application in the broad field of applications. The common computing platforms to build the BP algorithm based ANN Systems are Application Specific Integrated Circuits (ASICs) and General-Purpose Processors (GPP) based computers. However, due to a high demand of maintaining a trade-off between performance and flexibility, such computing machines become a bottleneck for further advanced improvements. In the last few decades, there has been significant progress in the field of Field Programmable Gate Arrays (FPGAs), which are based on the reconfigurable hardware platform. One of the main advantages of FPGAs are its flexibility, it is possible to reprogram the same hardware and achieve good performance by allowing parallel computation at the same time. The focus of this thesis is to implement the BP algorithm based ANN system on reconfigurable platform(FPGA). The proposed designs are coded on the software platform, MATLAB and in Verilog Hardware Description Language (Verilog HDL) on FPGA and synthesized on artix-7 FPGA evaluation kit. The validation of the design is verified on two benchmarks and comparisons are observed and discussed between two platforms.

This book gathers selected research papers presented at the First International Conference on Embedded Systems and Artificial Intelligence (ESAI 2019), held at Sidi Mohamed Ben Abdellah University, Fez, Morocco, on 2–3 May 2019. Highlighting the latest innovations in Computer Science, Artificial Intelligence, Information Technologies, and Embedded Systems, the respective papers will encourage and inspire researchers, industry professionals, and policymakers to put these methods into practice.

The three volume set LNCS 5551/5552/5553 constitutes the refereed proceedings of the 6th International Symposium on Neural Networks, ISNN 2009, held in Wuhan, China in May 2009. The 409 revised papers presented were carefully reviewed and selected from a total of 1.235 submissions. The papers are organized in 20 topical sections on theoretical analysis, stability, time-delay neural networks, machine learning, neural modeling, decision making systems, fuzzy systems and fuzzy neural networks, support vector machines and kernel methods, genetic algorithms, clustering and classification, pattern recognition, intelligent control, optimization, robotics, image processing, signal processing, biomedical applications, fault diagnosis,

telecommunication, sensor network and transportation systems, as well as applications.
This book discusses reliability applications for power systems, renewable energy and smart grids and highlights trends in reliable communication, fault-tolerant systems, VLSI system design and embedded systems. Further, it includes chapters on software reliability and other computer engineering and software management-related disciplines, and also examines areas such as big data analytics and ubiquitous computing. Outlining novel, innovative concepts in applied areas of reliability in electrical, electronics and computer engineering disciplines, it is a valuable resource for researchers and practitioners of reliability theory in circuit-based engineering domains.
This is Volume III of a three volume set constituting the refereed proceedings of the Third International Symposium on Neural Networks, ISNN 2006. 616 revised papers are organized in topical sections on neurobiological analysis, theoretical analysis, neurodynamic optimization, learning algorithms, model design, kernel methods, data preprocessing, pattern classification, computer vision, image and signal processing, system modeling, robotic systems, transportation systems, communication networks, information security, fault detection, financial analysis, bioinformatics, biomedical and industrial applications, and more.
This book lies at the interface of machine learning – a subfield of computer science that develops algorithms for challenging tasks such as shape or image recognition, where traditional algorithms fail – and photonics – the physical science of light, which underlies many of the optical communications technologies used in our information society. It provides a thorough introduction to reservoir computing and field-programmable gate arrays (FPGAs). Recently, photonic implementations of reservoir computing (a machine learning algorithm based on artificial neural networks) have made a breakthrough in optical computing possible. In this book, the author pushes the performance of these systems significantly beyond what was achieved before. By interfacing a photonic reservoir computer with a high-speed electronic device (an FPGA), the author successfully interacts with the reservoir computer in real time, allowing him to considerably expand its capabilities and range of possible applications. Furthermore, the author draws on his expertise in machine learning and FPGA programming to make progress on a very different problem, namely the real-time image analysis of optical coherence tomography for atherosclerotic arteries.
This book contains the papers presented at the 14th International Conference on Field Programmable Logic and Applications (FPL) held during August 30th- September 1st 2004. The conference was hosted by the Interuniversity Micro- Electronics Center (IMEC) in Leuven, Belgium. The FPL series of conferences was founded in 1991 at Oxford University (UK), and has been held annually since: in Oxford (3 times), Vienna, Prague, Darmstadt, London, Tallinn, Glasgow, Villach, Belfast, Montpellier and Lisbon. It is the largest and oldest conference in reconfigurable computing and brings together academic researchers, industry experts, users and newcomers in an informal, welcoming atmosphere that encourages productive exchange of ideas and knowledge between the delegates. The fast and exciting advances in field programmable logic are increasing steadily with more and more application potential and need. New ground has been broken in architectures, design techniques, (partial) run-time reconfiguration and applications of field programmable devices in several different areas. Many of these recent innovations are reported in this volume. The size of the FPL conferences has grown significantly over the years. FPL in 2003 saw 216 papers submitted. The interest and support for FPL in the programmable logic community continued this year with 285 scientific papers submitted, demonstrating a 32% increase when compared to the year before. The technical program was assembled from 78 selected regular papers, 45 additional short papers and 29 posters, resulting in this volume of proceedings. The program also included three invited plenary keynote presentations from Xilinx, Gilder Technology Report and Altera, and three embedded tutorials from Xilinx, the Universit ? at Karlsruhe (TH) and the University of Oslo.
This book constitutes the thoroughly refereed proceedings of the First IEEE Colombian Conference, ColCACI 2018, held in Medellin, Colombia, in May 2018. The 17 full papers presented were carefully reviewed and selected from 60 submissions. The papers are organized in topical sections on artificial neural networks; computational intelligence; computer science.
In this book, a global team of experts from academia, research institutes and industry presents their vision on how new nano-chip architectures will enable the performance and energy efficiency needed for AI-driven advancements in autonomous mobility, healthcare, and man-machine cooperation. Recent reviews of the status quo, as presented in CHIPS 2020 (Springer), have prompted the need for an urgent reassessment of opportunities in nanoelectronic information technology. As such, this book explores the foundations of a new era in nanoelectronics that will drive progress in intelligent chip systems for energy-efficient information technology, on-chip deep learning for data analytics, and quantum computing. Given its scope, this book provides a timely compendium that hopes to inspire and shape the future of nanoelectronics in the decades to come.
This book constitutes the refereed proceedings of the sixth International Conference on Artificial Neural Networks - ICANN 96, held in Bochum, Germany in July 1996. The 145 papers included were carefully selected from numerous submissions on the basis of at least three reviews; also included are abstracts of the six invited plenary talks. All in all, the set of papers presented reflects the state of the art in the field of ANNs. Among the topics and areas covered are a broad spectrum of theoretical aspects, applications in various fields, sensory processing, cognitive science and AI, implementations, and neurobiology.
This book features selected research papers presented at the International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN 2020), held at the Sir M. Visvesvaraya Institute of Technology on 20–21 February 2020. Discussing advances in evolutionary computing technologies, including swarm intelligence algorithms and other evolutionary algorithm paradigms which are emerging as widely accepted descriptors for mobile sustainable networks virtualization, optimization and automation, this book is a valuable resource for researchers in the field of evolutionary computing and mobile sustainable networks.
This book highlights recent research on bio-inspired computing and its various innovative applications in information and communication technologies. It presents 38 high-quality papers from the 10th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA 2019) and 9th World Congress on Information and Communication Technologies (WICT 2019), which was held at GIET University, Gunupur, India, on December 16-18, 2019. As a premier conference, IBICA-WICT brings together researchers, engineers and practitioners whose work involves bio-inspired computing, computational intelligence and their applications in information security, real-world contexts, etc. Including contributions by authors from 18 countries, the book offers a valuable reference guide for all researchers, students and practitioners in the fields of Computer Science and Engineering.

Explains current co-design and co-optimization methodologies for building hardware neural networks and algorithms for machine learning applications This book focuses on how to build energy-efficient hardware for neural networks with learning capabilities—and provides co-design and co-optimization methodologies for building hardware neural networks that can learn. Presenting a complete picture from high-level algorithm to low-level implementation details, Learning in Energy-Efficient Neuromorphic Computing: Algorithm and Architecture Co-Design also covers many fundamentals and essentials in neural networks (e.g., deep learning), as well as hardware implementation of neural networks. The book begins with an overview of neural networks. It then discusses algorithms for utilizing and training rate-based artificial neural networks. Next comes an introduction to various options for executing neural networks, ranging from general-purpose processors to specialized hardware, from digital accelerator to analog accelerator. A design example on building energy-efficient accelerator for adaptive dynamic programming with neural networks is also presented. An examination of fundamental concepts and popular learning algorithms for spiking neural networks follows that, along with a look at the hardware for spiking neural networks. Then comes a chapter offering readers three design examples (two of which are based on conventional CMOS, and one on emerging nanotechnology) to implement the learning algorithm found in the previous chapter. The book concludes with an outlook on the future of neural network hardware. Includes cross-layer survey of hardware accelerators for neuromorphic algorithms Covers the co-design of architecture and algorithms with emerging devices for much-improved computing efficiency Focuses on the co-design of algorithms and hardware, which is especially critical for using emerging devices, such as traditional memristors or diffusive memristors, for neuromorphic computing Learning in Energy-Efficient Neuromorphic Computing: Algorithm and Architecture Co-Design is an ideal resource for researchers, scientists, software engineers, and hardware engineers dealing with the ever-increasing requirement on power consumption and response time. It is also excellent for teaching and training undergraduate and graduate students about the latest generation neural networks with powerful learning capabilities.

This book features research presented at the 1st International Conference on Artificial Intelligence and Applied Mathematics in Engineering, held on 20–22 April 2019 at Antalya, Manavgat (Turkey). In today’s world, various engineering areas are essential components of technological innovations and effective real-world solutions for a better future. In this context, the book focuses on problems in engineering and discusses research using artificial intelligence and applied mathematics. Intended for scientists, experts, M.Sc. and Ph.D. students, postdocs and anyone interested in the subjects covered, the book can also be used as a reference resource for courses related to artificial intelligence and applied mathematics.

This book constitutes the refereed proceedings of the Third International Workshop on Applied Reconfigurable Computing, ARC 2007, held in Mangaratiba, Brazil, in March 2007. The 27 full papers and 10 short papers presented together with a late-comer contribution from ARC 2006 are organized in topical sections on architectures, mapping techniques and tools, arithmetic, and applications.

With the improvement in processing systems, machine learning applications are finding widespread use in almost all sectors of technology. Image recognition is one application of machine learning which has become widely popular with various architectures and systems aimed at improving recognition performance. With classification accuracy now approaching saturation point, many researchers are now focusing on resource and energy efficiency. With the increased demand for learning applications in embedded devices, it is of paramount importance to optimize power and energy consumption to increase utility in these low power embedded systems. In recent months, reduced precision neural networks have caught the attention of some researchers. Reduced data width deep nets offer the potential of saving valuable resources on hardware platforms. In turn, these hardware platforms such as Field Programmable Gate Arrays (FPGAs) offer the potential of a low power system with massive parallelism increasing throughput and performance. In this research, we explore the implementations of a deep learning architecture on FPGA in the presence of resource and energy constraints. We study reduced precision neural networks and implement one such architecture as a proof of concept. We focus on binarized convolutional neural network and its implementation on FPGAs. Binarized convolutional nets have displayed a classification accuracy of up to 88% with some smaller image sets such as CIFAR-10. This number is on the rise with some of the new architectures. We study the tradeoff between architecture depth and its impact on accuracy to get a better understanding of the convolutional layers and their impact on the overall performance. This is done from a hardware perspective giving us better insight enabling better resource allocation on FPGA fabric. Zynq ZCU-102 has been used for accelerator implementation. High level synthesis tool (Vivado HLS) from Xilinx is used for CNN definition on FPGA fabric.

This book describes new theories and applications of artificial neural networks, with a special focus on neural computation, cognitive science and machine learning. It discusses cutting-edge research at the intersection between different fields, from topics such as cognition and behavior, motivation and emotions, to neurocomputing, deep learning, classification and clustering. Further topics include signal processing methods, robotics and neurobionics, and computer vision alike. The book includes selected papers from the XIX International Conference on Neuroinformatics, held on October 2-6, 2017, in Moscow, Russia.

The three volume set LNCS 4232, LNCS 4233, and LNCS 4234 constitutes the refereed proceedings of the 13th International Conference on Neural Information Processing, ICONIP 2006, held in Hong Kong, China in October 2006. The 386 revised full papers presented were carefully reviewed and selected from 1175 submissions.

This book constitutes the thoroughly refereed post-proceedings of the Second International Workshop on Reconfigurable Computing, ARC 2006, held in Delft, The Netherlands, in March 2006. The 22 revised full papers and 35 revised short papers presented were thoroughly reviewed and selected from 95 submissions. The papers are organized in topical sections on applications, power, image processing, organization and architecture, networks and communication, security, and tools.

This book constitutes, together with its compagnion LNCS 1606, the refereed proceedings of the International Work-Conference on Artificial and Neural Networks, IWANN'99, held in Alicante, Spain in June 1999. The 91 revised papers presented were carefully reviewed and selected for inclusion in the book. This volume is devoted to applications of

biologically inspired artificial neural networks in various engineering disciplines. The papers are organized in parts on artificial neural nets simulation and implementation, image processing, and engineering applications.

Low-power, high-speed neural networks are critical for providing deployable embedded AI applications at the edge. We describe a Xilinx FPGA implementation of Neural Engineering Framework (NEF) networks with online learning that outperforms mobile Nvidia GPU implementations by an order of magnitude or more. Specifically, we provide an embedded Python-capable PYNQ FPGA implementation supported with a Xilinx Vivado High-Level Synthesis (HLS) workflow that allows sub-millisecond implementation of adaptive neural networks with low-latency, direct I/O access to the physical world. The outcome of this work is NengoFPGA, a seamless and user-friendly extension to the neural compiler Python package Nengo. To reduce memory requirements and improve performance we tune the precision of the different intermediate variables in the code to achieve competitive absolute accuracy against slower and larger floating-point reference designs. The online learning component of the neural network exploits immediate feedback to adjust the network weights to best support a given arithmetic precision. As the space of possible design configurations of such quantized networks is vast and is subject to a target accuracy constraint, we use the Hyperopt hyper-parameter tuning tool instead of manual search to find Pareto optimal designs. Specifically, we are able to generate the optimized designs in under 500 short iterations of Vivado HLS C synthesis before running the complete Vivado place-and-route phase on that subset, a much longer process not conducive to rapid exploration. For neural network populations of 64-4096 neurons and 1-8 representational dimensions our optimized FPGA implementation generated by Hyperopt has a speedup of 10-484× over a competing cuBLAS implementation on the Jetson TX1 GPU while using 2.4-9.5× less power. Our speedups are a result of HLS-specific reformulation (15× improvement), precision adaptation (3× improvement), and low-latency direct I/O access (1000× improvement).