MAMMALIAN RETINA MODELING AND ON-LINE LEARNING ON CNN ARCHITECTURE
FROM WETWARE TO SILICON

Theses of the *Ph.D.* Dissertation

Dávid Bálya

Analogical and Neural Computing Systems Laboratory
Computer and Automation Institute
Hungarian Academy of Sciences

Scientific adviser:

Tamás Roska, D.Sc.
ordinary member of the HAS

Budapest, 2003

"If our brain were so simple that we could understand it,
we would be so simple that we could not."

Jonathan Evans

Introduction

The proven limitations in pure digital computing motivated an intensive research on the field of neural/nonlinear parallel processing and the fabrication of the first hardware prototypes. However, contrary to all expectations, this did not lead to a breakthrough of a new technology. The reason for that was not necessarily the well-known bottleneck, the limited precision of all analog hardware implementations, but other factors. The spread and utilization of these novel computing tools in industrial applications were even more limited by the imperfection of the architectures, to mention the most important: quick reprogramming of these new computing structures was not possible, restricting their use to very specific tasks. Furthermore, the approach assuming a fully connected processing net turned out to be a major obstacle from the realization point of view, since the implementation complexity increases exponentially with the number of processors.

Computing architectures based on the cellular neural/nonlinear network (CNN) paradigm offer an adequate solution for the problems above. CNNs are regular, single or multi-layer, parallel processing structures with analog nonlinear computing units (base cells). The state values of the individual processors are continuous in time and their connectivity is local in space. The program of these networks is completely determined by the pattern of the local interactions, the so-called template, and the local logic and arithmetic (analog) instructions. The time-evolution of the analog transient, driven by the template operator and the processor dynamics, represents the computation in CNN. Results can be defined both in equilibrium and/or non-equilibrium states of the network. Completing the base cells of CNN with local sensors, local data memories, arithmetical and logical units, furthermore with global program memories and control units results in the CNN Universal Machine (CNN-UM) architecture. The CNN-UM is an analogic (analog and logic) supercomputer, it is universal both in Turing sense and as a nonlinear operator, therefore it can be used as a general architectural framework when designing CNN processors. Up to the present there have been various physical implementations of this architecture: mixed-signal VLSI, emulated digital VLSI and optical.

In this dissertation, the focus is put on modelling and implementing different neurobiological phenomenon on the CNN-UM architecture. Parallel algorithmic solutions are shown for some application related problems as well where the demand for an extremely high computing power and the requirement for fast processing make it impossible for a designer to apply a fully sequential approach. The underlying idea of most methods is based on spatio-temporal dynamics of diffusion and wave phenomena that can be well reproduced and controlled on a parallel processing array. The majority of CNN templates used in the detection algorithms are such that their prototypes can be tested on existing VLSI implemented CNN chips. The goal of the research conducted on the field of retina modelling is twofold. Though, understanding the behaviour and functionality of a biological system relying on morphological, pharmacological and neurophysiological observations is a natural motivation of all modelling experiments, in this dissertation the engineering approach is stressed the endeavour to reveal the secrets of robust and fault-tolerant feature detection.

The task of my work was to contribute to the research of retinal modeling and to find solutions for the hardware implementation of the multi-layer analogic cellular nonlinear network universal machine (CNN-UM) in mammalian retina modeling. I also had to prove that it is worthwhile and feasible to deal with CNN-UM analog VLSI chips in neurobiological modeling. This work included theoretical, simulation and experimental evaluation and validation.

During the preparatory work I have found the abstract neuron model based on the receptive field calculus for modeling different sensory modalities and algorithms that might open new ways of integrating neurobiological models into the common engineering practice. Some of them have already been verified experimentally on the ACE4k and on the CACE1k cellular visual microprocessors. These solutions exploit the implicit parallelism and sensor-near properties of the CNN-UM to enhance and speed-up the computation of different algorithms.

A CNN-UM can be applied as a stored programmable analogic cellular computer using analog transient wave-type computing and logic operators for feature extraction. Feature vectors can be used for object recognition after post-processing. The classification of feature vectors is an important task especially if we have a fast feature detector such as the CNN-UM.

I wish to point out that the synergy of the neurobiology, based on retina modeling, and the CNN-UM technology, especially analog chips with optical input, could be fruitfully realized.

Methods used in the experiments

In the course of my work, theorems and assertions from the field of ordinary and partial differential equations, reported results on robustness and feasibility analysis connected to nonlinear networks, methodologies related to neurobiological modelling were explored. The biological modelling is based on morphological, pharmacological and neurophysiological observations and measurements.

Designed CNN templates and algorithmic procedures were tested on software simulators, such as the CANDY package and ALADDIN PRO 2.x, and on an industrial framework embedding different Cellular Visual Microprocessors: ACE4k and CACE1k. The simulators were developed in the Analogical and Neural Computing Laboratory. The implementation independent description of the developed methods was completed in different "CNN languages" (template, UMF, Alpha, AMC) ensuring their applicability on different hardware platforms. During the design process the VLSI implementation complexity was minimized, therefore, for solving various tasks only nearest neighbour templates were required. Most of these are linear interaction operators (templates) that can be tested on existing VLSI prototype systems or nonlinear operators with simple nonlinear interactions. These operators are expected to be built in the near future. In software simulations, the numerical integration of coupled ordinary differential equations was based on first order adaptive explicit Euler formula. In certain cases, for validation, a higher order (fourth order Runge-Kutta) formula could also be used.

The experiments and research efforts were interdisciplinary. A number of analyses were performed in a close cooperation with neurobiologists. In retina modelling experiments, observations and guidance of biologists supported the synthesis and analysis of various CNN based models.

New scientific results

1. Thesis: Establishing a sophisticated hardware implemented mammalian retina model based on neurobiological data; *Chapter 2*

I have shown that a qualitatively correct typical mammalian retina can be build with the cellular nonlinear network (CNN) technology, which can be implemented on silicon as a programmable analogue very large-scale integrated (VLSI) chip. I have designed different cell-layers of the retina with space-invariant nearest neighbour interactions and monotonic continuous synaptic characteristics. In the verification process, the spatio-temporal patterns of the model output were compared to the experimental data recorded from rabbit retina.
Published in IJBC, CTA, ISCAS, CNNA, ECCTD

1.1 I defined the structure of the multi-layer CNN retina model and described a method for its tuning.

I have worked out the structure of the multi-layer mammalian retina model based on morphology and electro-physiological measurements. The building blocks of the model are the nonlinear abstract neuron layers with special synapses. A method to tune the model parameters and fine-tune the structure is developed and used it for model improvement (tuning), confirmation (falsification), and robustness analysis. I have tuned the model to qualitatively reproduce the electro-physiological measurements of several channels of the rabbit retina. This is the first neuromorphic mammalian retina model that reproduces some of the measured outputs of several parallel channels.

1.2 The feasibility of implementing the developed multi-layer retina model on analogue VLSI chips

I have worked out how to implement the developed multi-layer retina model on VLSI chips using decompositions in three domains: spatial, temporal, and structural. The method to transform the originally multi-layer, temporally continuous retina model to the general purpose, programmable, complex-kernel cellular nonlinear network universal machine (CNN-UM) is given as an analogic (analogue transient and logic) algorithm. This analogic algorithm, the decomposed model blocks, was implemented on a hardware implementation of the complex-kernel CNN-UM, the CACE1k chip. In summary, I defined the steps and designed a method, to establish a qualitatively correct retina model starting from the neurobiological findings through the CNN model until the hardware implementation.

2. Thesis: I implemented an efficient analogic algorithm, based on a measured retinal function, to detect rapid global changes in the visual scene; *Chapter 3*

I designed an algorithm based on the properties of the rapid global change suppression global spatial-temporal event. The steps of the algorithm are the following: difference calculation, spatial integration, influence expansion, and transient inhibition. This algorithm is implemented on a general purpose CNN-UM chip (ACE4k) therefore it became an on-line, real-time tool to detect sudden global changes in the input video-flow. The algorithm can be plugged into the retina model as a special retina function block.

Published in JCSC, ISCAS03

3. Thesis: On-line learning on the CNN-UM architecture, classification with neural networks; *Chapter 4*

The typical implementation of the artificial neural networks is simulation; although some custom designed processors have been proposed. However, these processors should be integrated into the commonly used framework. I have developed a method to transform the general feed-forward neural networks to the CNN-UM platform based on state distribution. Thus the original network can be computed in the general purpose, programmable, parallel CNN-UM environment as an analogic algorithm. The method was used to efficiently implement the adaptive resonance theory (ART) network.

Published in IJNS, CNNA02, ECCTD03

3.1 I designed a method to transform the general feed-forward artificial neural networks to the CNN-UM platform.

The key of the method is how to represent the weights of the general feed-forward neural networks by the states of cells in the CNN array. I have applied state distribution for this purpose, which is followed by the parallel computation of the CNN-UM as an analogic algorithm. Finally the state aggregation collects the results. The method makes possible to efficiently employ the widely used artificial neural networks on CNN-UM, which can be implemented in a general-purpose hardware. The logic of the transformed analogic algorithms depend on the type of the network, but can be built up from a few simple subroutines, which are given for both binary and greyscale cases: state aggregation, hard-limited output computation, winner selection, and new node creation.

3.2: I implemented the ART classification network on a CNN-UM chip.

I have shown how to implement the ART network on the CNN-UM architecture. The original algorithm is modified to be suitable for real-time computing and I have derived the optimal binary decision functions: a choice function to select the most similar category, a match function to reject the inappropriate categories, and a learning function to update the category prototype. The algorithm is suitable to cluster the input vectors into self-generated categories with adjustable resolution or can be applied for supervised classification without predefining the number of categories. The algorithm is convergent, stable, robust, and its execution time is independent of the number of the existing categories and features. The method is implemented on the ACE4k cellular visual microprocessor, which can handle maximum 32 features in 64 categories.

New practical results

- I designed a model framework to describe cellular networks composed by several common analogue neurons, where only the key parameters should be set, but the system is still neuromorph. I have shown how this framework can be transformed into CNN differential equations.
- I analyzed the previously described model framework and in some typical cases I derived explicit formulas for the overshoot and the time to reach the peak value of the transient. I derived the possible exponential transfer functions for the models.
- I derived an explicit and adaptive-stepsize integration formula, which has only small extra cost plus uses no back-step and speeds up the simulation tremendously.
- I tuned the retina model to qualitatively reproduce the rabbit retina measurements.
- I applied the rapid global change suppression algorithm successfully for video-shot detection.

Application of the results

The results of the mammalian retina research (Thesis 1) can be embedded into several complex algorithms and applications targeting real-life applications, like object classification, recognition, tracking, and alarming. The examined and reproduced retina effects may be useful for image processing tasks such as:

- edge and object corner detection in space and time,
- object level motion detection with size selectivity (beside local interactions),
- speed, size and intensity selective video-flow processing with noise filtering

and for design principles, as well:

- spatial, temporal, and spatio-temporal decomposition of the input flow for sparse signal representation,
- adaptive parallel On-Off channel processing for signal flow normalization,
- vertical interaction of the decomposed channels to decrease the correlation between the processing flows.

Our ambitious goals are building the front-end processing of an artificial vision system and designing the software of retinal implants where the developed model could be directly used.

The presented global spatial-temporal event detector (Thesis 2) can be used in different environments. The algorithm serves as a preprocessing or filtering function indicating sudden changes of the input-flow. The algorithm, due to its low memory requirements and fast processing speed, can be a common function in any on-line, real-time spatial-temporal algorithm, such as:

- Motion compensation and image stabilization,
- Artificial video-shot change detection (e.g. hard cuts, fades and dissolve),
- Key-frame selection for content-based video compression,
- Time instance determination for re-initialization of the history based processing flow,
- Accident detection of the mounted camera (e.g. the tripod overturns),
- Mammalian retina model improvement: rapid global change suppression effect.

The proposed analogic ART-like algorithm (Thesis 3) can be applied wherever is a need for high-speed adaptive classification and it is possible to compute several simple features. A low-cost and high-speed pattern recognition system with low power consumption and limited accuracy can be a viable alternative to existing solutions. Some possible application areas are:

- On-line morphological identification
- Site identification by texture classification, e.g. the road ahead is sandy, rocky or wet.
- Biometric security systems, e.g. fingerprint verification or face recognition
- Biomedical signal (e.g. EEG) analysis
- Pattern recognition
- Target tracing: object – model binding

Acknowledgments

In the first place, I would like to thank my supervisor, Professor Tamás Roska, for his unbroken enthusiasm, consistent support, and also his outstanding personal qualities. Under his careful guidance, I was introduced to the wealth of material that ultimately led to this dissertation. I consider it a great privilege to be part of his highly qualified research group.

I am very grateful to Prof. Frank Werblin for two wonderful semesters that I could spend in Berkeley, where as a fresh graduate student I was already posed to big challenges. Cooperation with his research group, especially with Botond Roska, was a great gift, where I learned a lot and not just about neurobiology.

It is hard to make a complete list of all those who made the past few years especially colorful. I recall Erik Nemeth and Thomas Munch, Shelly Fried from the Werblin Lab in Berkeley. And of course, there is a long list from the Analogical Lab. Special thanks to Csaba Rekeczky for the continuous interaction and to Ákos Zarándy for a number of helpful discussions including the pre-review of this dissertation; to István Szatmári and Péter Földesy for assistance during chip tests, and to the younger ones Gergely Tímár, Viktor Gál, István Petrás, Zoltán Szlávik, György Cserey and the others. I am equally indebted to Katalin Keserű and Péter Szolgay for their practical and official aid.

I had some really great teachers who should be mentioned here. Prof. Pál Rózsa and József Hámori from the graduate years, and other unforgettable classes by Endre Selényi, György Fodor, and Gábor Horváth. I am grateful to Lajos Rónyai, who was my supervisor for a TDK paper during my undergraduate studies and introduced me to the algorithmic thinking. I cannot leave out my non-technological scientific advisors: László Karvalics, József Mészáros and my multi-faceted high school teacher Mrs. Gábor Mód.

I am very grateful to my mother and to my whole family who always believed me and in my goals and support me in all possible ways. And last, but not least the support of the Hungarian Academy of Sciences is acknowledged.

The author's journal publications

In print

D. Bálya, I. Petrás, T. Roska, R. Carmona, A. Rodríguez Vázquez: "Implementing the Multi-Layer Retinal Model on the Complex-Cell CNN-UM Chip Prototype", *International Journal of Bifurcation and Chaos (IJBC)*, Vol. 14(2), February 2004.

V. Gál, J. Hámori, T. Roska, **D. Bálya**, Zs. Borostyánkői, M. Brendel, K. Lotz, L. Négyessy, L. Orzó, I. Petrás, Cs. Rekeczky, J. Takács, P. Venetiáner, Z. Vidnyánszky, Á Zarándy: "Receptive Field Atlas and Related CNN Models", *International Journal of Bifurcation and Chaos (IJBC)*, Vol. 14(2), February 2004.

D. Bálya: "Sudden Global Spatial-Temporal Change Detection and its Applications", *Journal of Circuits, Systems, and Computers (JCSC)*, Vol. 12(6), December 2003.

D. Bálya: "CNN Universal Machine as Classification Platform: an ART-like Clustering Algorithm", *International Journal of Neural Systems (IJNS)*, Vol. 13(6), December 2003.

2002

D. Bálya, B. Roska, T. Roska, F. S. Werblin: "A CNN framework for modeling parallel processing in a mammalian retina", *International Journal on Circuit Theory and Applications (CTA)*, Vol. 30, 2002; pp. 363-393.

2001

F.S. Werblin, B. Roska, **D. Balya**: "Parallel processing in the mammalian retina: lateral and vertical interactions across stacked representations", *Progress in Brain Research* 2001, "Concepts and Challenges in Retinal Biology" Kolb, H., Ripps, H, & Wu, S. (Eds), Elsevier Science; 131:229-38

L. Fortuna, P. Arena, **D. Bálya** and Á. Zarándy: "Cellular Neural Networks: A Paradigm for Nonlinear Spatio-Temporal Processing", *IEEE Circuits and Systems Magazine*, Vol. 1-4, 2001; pp. 6-21.

2000

T. Roska, **D. Bálya**, M. Csapodi, Á. Zarándy: "Analogikai celluláris számítógépek. Egy új számítógépelv" *Természet Világa, Informatika különszám*, Természettudományi Közlöny Vol. 131, 2000; pp. 71-74 (Analogic Cellular Computers - A new paradigm)

1999

D. Bálya and T. Roska, "Face and eye detection by CNN algorithms", *Journal of VLSI Signal Processing Special Issue: Spatiotemporal Signal Processing with Analogic CNN Visual Microprocessors, (JVSP Special Issue)*, Vol. 23. No.2/3. pp. 497-511, guest editors: T. Roska and Á. Rodríguez-Vázquez, Kluwer, ISSN 0922-5773, 1999

The author’s international conference publications

2003

D. Bálya, T. Gergely, I. Szatmari, and Cs. Rekeczky: "Classification of Spatio-Temporal Features: the Nearest Neighbor Family", *IEEE European Conference on Circuit Theory and Design ECCTD 2003*, Krakow Sept., 2003

Z. Szlávik, **D. Bálya**, T. Roska: "Properties of the adaptive integration formula to compute the CNN dynamic equations", *IEEE European Conference on Circuit Theory and Design ECCTD 2003*, Krakow Sept., 2003

Cs. Rekeczky, I. Szatmári, G. Tímár, **D. Bálya**: "Adaptive Multi-rate, Multi-grid and Multi-scale Algorithms Running on Analogic Architecture”, *IEEE European Conference on Circuit Theory and Design ECCTD 2003*, Krakow Sept., 2003

G. Tímár, **D. Bálya**, I. Szatmári, and Cs. Rekeczky: "Feature Guided Visual Attention With Topographic Array Processing and Neural Network-Based Classification', *International Joint Conference on Neural Networks IJCNN 2003*, Portland July, 2003

D. Bálya: "Detecting Global Spatial-Temporal Events: Saccadic Suppression", *IEEE International Symposium on Circuits and Systems ISCAS 2003*, Bangkok May, pp. III-838-841

Cs. Rekeczky, **D. Bálya**, G. Timár, and I. Szatmári: "Bio-Inspired Flight Control and Visual Search with CNN Technology", *IEEE International Symposium on Circuits and Systems ISCAS 2003*, Bangkok May, pp.III-774-777

I. Szatmári, **D. Bálya**, T. Gergely, Cs. Rekeczky, and T. Roska: "Multi-Channel Spatio-Temporal Topographic Processing for Visual Search and Navigation", *SPIE Microtechnologies for the New Millennium 2003*, Gran Canaria May, pp. 297-306

2002

D. Bálya, Cs. Rekeczky, T. Roska: “Basic Mammalian Retinal Effects on the Prototype Complex Cell CNN Universal Machine”, *Proceedings of 7th IEEE International Workshop on Cellular Neural Networks and their Applications CNNA 2002*, Frankfurt/Main July, 2002; pp. 251-258

D. Bálya and T. Roska: “A Supervised and unsupervised ART-like Classification of Binary Vectors Implemented on CNN Universal Machine”, *Proceedings of 7th IEEE International Workshop on Cellular Neural Networks and their Applications CNNA 2002*, Frankfurt/Main July, 2002; pp. 616-623

D. Bálya, Cs. Rekeczky, T. Roska: “The Real Mammalian Retinal Model Implemented on Complex Cell CNN Universal Machine”, *IEEE International Symposium on Circuits and Systems ISCAS 2002*, Scottsdale May, pp. IV-161-164.

2001

D. Bálya, B. Roska, T. Roska, F. S. Werblin: “A CNN model framework and simulator for biologic sensory systems*”, 15th IEEE European Conference on Circuit Theory and Design ECCTD 2001*, Helsinki Sept, 2001; pp. I-357-360.

F. S. Werblin, B. Roska, **D. Bálya**, Cs. Rekeczky and T. Roska: “Implementing a retinal visual language in CNN: a neuromorphic study”, *IEEE International Symposium on Circuits and Systems ISCAS 2001*, Sydney May, 2001, pp. III-333-336.

2000

D. Bálya, B. Roska, T. Roska, F. S. Werblin: “A Qualitative Model-Framework for Spatio-temporal Effects in Vertebrate Retinas”, *Proc. of 6th IEEE International Workshop on Cellular Neural Networks and their Applications CNNA*, Catania April, 2000; pp. 165-170.

1999

P. Szolgay, Á. Zarándy, S. Zöld, T. Roska, P. Földesy, L. Kék, T. Kozek, K. László, I. Petrás, Cs. Rekeczky, I. Szatmári and **D. Bálya**, "The Computational Infrastructure for Cellular Visual Microprocessors", *Proc. 7th Intl.Conf. on Microelectronics for Neural, Fuzzy, and Bio-Inspired Systems, (MicroNeuro'99)*, Granada, IEEE Computer Society, 1999; pp. 54-60.

Publications on topics connected to the dissertation

L. O. Chua and L. Yang, "Cellular Neural Networks: Theory", *IEEE Trans. on Circuits and Systems*, Vol. 35, pp. 1257-1272, 1988.

L. O. Chua, and T. Roska, "The CNN Paradigm", *IEEE Trans. on Circuits and Systems*, Vol. 40, pp.147-156, March 1993.

T. Roska and L. O. Chua, "The CNN Universal Machine: an Analogic Array Computer", *IEEE Trans. on Circuits and Systems*, Vol. 40, pp. 163-173, March 1993.

L. O. Chua, T. Roska, and P. L. Venetianer, "The CNN is as Universal as the Turing Machine", *IEEE Trans. on Circuits and Systems*, Vol. 40, pp. 289-291, April 1993.

G. Liñan, S. Espejo, R. Domínguez-Castro, E. Roca and A. Rodríguez-Vázquez "The CNNUC3: An Analog I/O 64x64 CNN Universal Machine Chip Prototype with 7-Bit Analog Accuracy", *Proc. Cellular Neural Networks and their Applications*, (CNNA'2000), Catania, pp. 201-206.

R. Carmona, F. Jiménez-Garrido, R. Domínguez-Castro, S. Espejo, Á. Rodríguez-Vázquez, Cs. Rekeczky and T. Roska "A Bio-Inspired 2-Layer Mixed-Signal Flexible Programmable Chip for Early Vision", *IEEE Transactions on Neural Networks*, Vol. 14(5), pp.1313-1336, 2003

I. Petrás, Cs. Rekeczky, R. Carmona, F. Jiménez-Garrido "Exploration of complex dynamic phenomena in a 32×32-cells stored program 2-layerCNN Universal Machine Chip Prototype with 7-bits Analog Accuracy," *Journal of Circuits, Systems, and Computers*, Vol. 13(6), 2003.

Á. Zarándy, Cs. Rekeczky, P. Földesy, I. Szatmári "The New Framework of Applications - the Aladdin System," *Journal of Circuits, Systems, and Computers*, Vol. 13(6), 2003.

B. Roska and F. S. Werblin "Vertical Interactions across Ten Parallel Stacked Representations in Mammalian Retina," *Nature* 2001. vol. 410 pp. 583-587.

Cs. Rekeczky, B. Roska, E. Nemeth and F. S. Werblin "The network behind spatio-temporal patterns," *Int J Circuit Theory and Applications* 2001. vol. 29, pp. 197-239.

J. E. Dowling, "The retina: an approachable part of the brain", The Belknap Press of Harvard University Press, Cambridge 1987.

F. Werblin, T. Roska, and L. O. Chua, "The Analogic CNN Universal Machine as a Bionic Eye", *International Journal of Circuit Theory and Applications*, Vol. 23, pp. 541-569, 1995.

F. Werblin, A. Jacobs, and J. Teeters, "The Computational Eye", *IEEE Spectrum*, pp. 30-37, 1996.

B. Roska and F. S. Werblin "Rapid global shifts in natural scenes block spiking in specific ganglion cell types" *Nature Neuroscience* 2003, Vol. 6(6), pp. 600-608.

Lienhart Rainer "Reliable Transition Detection in Videos: A Survey and Practitioner's Guide", *International Journal of Image and Graphics (IJIG)*, 2001. vol. 1, No. 3, pp. 469-486.

F. Rosenblatt F. "The Perceptron: A Probabilistic Model for Information Storage and Organization of the Brain," *Psycol. Rev.*, 1958, vol. 65, pp. 386-408.

G. A. Carpenter and S. Grossberg "A massively parallel architecture for a self-organizing neural pattern recognition machine", *Computer Vision, Graphics, and Image Processing*, 1987. Vol. 37 pp. 54-115.

T. Serrano, B. Linares, and A. Andreou "Adaptive resonance Theory Microchips" *Norwell MA: Kluwer Academic,* 1998.