Doctoral Thesis ETH No. 16896

Spike-based Winner-Take-All Computation in a Multi-Chip Vision System

A dissertation submitted to the

Swiss Federal Institute of Technology Zurich

for the degree of

Doctor of Technical Sciences

presented by

Matthias Ferdinand Oster

Diplom-Ingenieur Univ., Technische Universität München

born 13. Juni 1977
citizen of Germany

accepted on the recomendation of

Prof. Dr. Rodney J. Douglas, examiner
Dr. Shih-Chii Liu, co-examiner
Prof. Dr. Bernabe Linares-Barranco, co-examiner

2006

Summary

The asynchronous and time-continuous computation that takes place in biological systems is of great interest because of the capability of these systems to interact with the real-world. In this work we explore such computation in the spike-based winner-take-all network, by developing a theoretical model of its performance and describing its implementation in a large-scale multi-chip vision system.

The winner-take-all is a neuronal network that amplifies the strongest set of inputs and suppresses output from the others. In various neuroscience models, this function is used to make a selection out of a possible set of decisions dependent on the input. As we develop artificial spike-based systems for different applications we need a theoretical understanding of the parameter settings in the spike-based neuronal networks where we can obtain this behavior. Previous analyses of winner-take-all behavior have considered analog or spike-rate coded inputs. Here we are interested in the computation in the transition between single-spike and spike rate coding.

In the theoretical part of this thesis (Chapter 2), we show for both regular and Poisson spike trains under which parameters the decision of the winner is optimal, that is it makes use of all information available in the input spikes. For inputs of regular rates the winner can be selected with only one inter-spike interval. For inputs of Poisson rates, the performance of the network depends on the number of spikes the neurons need to reach threshold.

In biology, the winner-take-all network is believed to be part of the cortical microcircuit, in which it both cooperates and competes with other areas to reach a consistent interpretation of the input. We model such context information with self-excitation if the input is stationary, with permitted sets to embed syntactic constraints, and with competition across winner-take-all networks to select the strongest of a set of feature maps.

We then describe how these theoretical analyses are applied in Very-Large-Scale-Integration technology (Chapter 3). Neurons and synapses are implemented using analog transistor circuits, while spikes are represented as digital address-events. We focus on two issues in the implementation: mismatch and a hybrid hardware/software framework. First, mismatch in the analog transistors limits the performance of the winner-take-all network. We characterize the

mismatch in the input synapses and discuss different schemes of compensating the synaptic weights. Second, we present a software framework to analyze the spike trains in the system and to add learning and adaptation capabilities by embedding software agents in the spiking communication.

Finally, in Chapter 4, we analyze the hardware winner-take-all in a multi-chip vision system (CAVIAR), consisting of an artificial retina, a bank of spike-based convolution filters, the winner-take-all network and a learning module that classifies trajectories of moving objects. We show that the inputs to the winner-take-all follow Poisson statistics, although retina and convolution filters are deterministic and exhibit only a small amount of variation. When the system is stimulated with a moving object, the input to the winner-take-all is a spike wave of Gaussian shape that is traveling across the neurons. With this input we can compare the winner-take-all performance predicted by our theoretical model with the performance of our implementation when applied to this large-scale system that performs real-time asynchronous computation.

Zusammenfassung

Die asynchrone und zeitkontinuierliche Informationsverarbeitung biologischer Systeme ist von großem Interesse, da sie diesen Systemen ermöglicht, in Echtzeit mit der Welt zu interagieren. In dieser Arbeit erforschen wir diese Rechenprinzipien in einem pulsbasierten neuronalen Netz zur Maximumselektion. Wir entwickeln ein theoretisches Modell der Leistungsfähigkeit und beschreiben die Implementierung in einem hochintegrierten Bildverarbeitungssystem.

Neuronale Netze zur Maximumselektion verstärken die stärkste Gruppe von Eingangssignalen und unterdrücken die Aktivität der anderen Neurone. Diese Funktion wird häufig in neurowissenschaftlichen Modellen verwendet, um Entscheidungen in Abhängigkeit des Eingangs auszuwählen. Um solche pulsbasierten Netzwerke in technische Applikationen umzusetzen, benötigen wir ein theoretisches Verständnis ihrer Wirkungsweise und Parameterbereiche. Vorhergehende Untersuchungen von Netzwerken mit Maximumselektion haben die Eingänge entweder als analoge Signale oder aber frequenzkodiert angesehen. Unser Interesse gilt den Prinzipien der Berechnung im Übergang der Kodierung als einzelne Pulse und als Pulsrate.

Im Theorieteil dieser Dissertation (Kapitel 2) untersuchen wir die Parameterbereiche des Netzwerks für die optimale Erkennung von Eingangssignalen mit konstanter Rate und mit Poisson-Statistik. Wir sprechen von optimaler Erkennung, wenn alle in den Eingangssignalen enthaltene Informationen ausgenutzt werden. Bei konstanter Rate kann das Netzwerk das stärkste Eingangssignal mit nur zwei Pulsen detektieren; bei Kodierung mit Poisson-Statistik hängt die Erkennungsleistung von der Anzahl der Pulse ab, die die Nervenzellen zum Erreichen ihres Schwellwertes benötigen.

Selektionsnetzwerke werden in der Modellierung von kortikalen Standardnetzwerken eingesetzt, um Informationen verschiedener Gehirnareale gleichzeitig kompetitiv und kooperativ zu integrieren. Wir berücksichtigen solche Kontextinformationen in unserem Netzwerk durch Selbsterregung der Neurone für stationäre Eingangssignale, durch Koaktivierungsmuster für die Integration syntaktischer Nebenbedingungen, und durch Selektion auf Netzwerkebene in der Musterkennung mit Filterbänken.

Das theoretische Verständnis des Netzwerks ermöglicht die Umsetzung als hochintegrierte Halbleiterschaltung (Kapitel 3). Neurone und Synapsen werden

durch analoge Transistorschaltkreise implementiert; digitale Pulse modellieren die Aktionspotentiale biologischer Nervenzellen. Wir konzentrieren uns auf zwei Implementierungsfragen: auf Parameterschwankungen und die Integration von Algorithmen. Parameterschwankungen in den analogen Schaltkreisen mindern die Leistungsfähigkeit des Selektionsnetzwerks. Wir untersuchen die Parameterschwankungen in den Eingangssynapsen und diskutieren verschiedene Verfahren zur Kompensation der synaptischen Gewichte. Als zweites entwickeln wir eine hybride Hardware/Softwarearchitektur zur Analyse der Pulsfolgen im System, und um Lern- und Adaptionsalgorithmen in die Pulskommunikation zu integrieren.

In Kapitel 4 diskutieren wir die Leistungsfähigkeit des Netzwerkes eingebettet in ein hochintegrierte Bildverarbeitungssystem (CAVIAR), bestehend aus einer künstlichen Retina, einer pulsbasierten Filterbank, dem Selektionsnetzwerk und einem Lernmodul, das die Trajektorien sich bewegender Objekte klassifiziert. Wir zeigen, daß die Signale im System der Poissonverteilung folgen, obwohl sowohl Sensor als auch die Vorverarbeitung deterministisch sind und nur geringe Parameterschwankungen aufweisen. Sich bewegende Objekte lösen eine Welle von Aktivität im Profil einer Normalverteilung aus, die die Neuronen des Netzes entlangläuft. Mit dieser Modellierung der Eingangssignale kann die Leistungsfähigkeit des Netzwerkes in der Vorhersage des theoretischen Modells mit der Anwendung in einem hochintegrierten System verglichen werden, um echtzeitfähige asynchrone Rechenarchitekturen zu verstehen.