Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Concurrent Parallel-Sequential Processing in Gamma Controlled Cortical-Type Networks of Spiking Neurones

Edgar Körner, Ursula Körner

1997

Preprint:

This is an accepted article published in Proc. International Conference on Artificial Neural Networks. The final authenticated version is available online at: https://doi.org/[DOI not available]

14479

Wulfram Gerstner Alain Germond
Martin Hasler Jean-Daniel Nicoud (Eds.)

Artificial Neural Networks – ICANN '97

7th International Conference
Lausanne, Switzerland, October 8-10, 1997
Proceeedings

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany

Juris Hartmanis, Cornell University, NY, USA

Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Wulfram Gerstner
Alain Germond
Martin Hasler
Jean-Daniel Nicoud
Swiss Federal Institute of Technology
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
E-mail: wulfram.gerstner@di.epfl.ch
germond@de.epfl.ch
hasler@de.epfl.ch
nicoud@di.epfl.ch

Cataloging-in-Publication data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Artificial neural networks : 7th international conference ; proceedings / ICANN '97, Lausanne, Switzerland, October 8 - 10, 1997. Wulfram Gerstner ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ; Singapore ; Tokyo : Springer, 1997
(Lecture notes in computer science ; Vol. 1327)
ISBN 3-540-63631-5

CR Subject Classification (1991): F.1.1, C.2.1, C.1.3, I.2, G.1.6, I.5.1, B.7.1, J.1, J.2

ISSN 0302-9743
ISBN 3-540-63631-5 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1997
Printed in Germany

Typesetting: Camera-ready by author
SPIN 10545612 06/3142 – 5 4 3 2 1 0 Printed on acid-free paper

Concurrent Parallel-Sequential Processing in Gamma Controlled Cortical-Type Networks of Spiking Neurones

Edgar Koerner and Ursula Koerner

HONDA R&D Europe (Germany) Ltd., Future Technology Division,
Offenbach/Main, Germany.
email: Edgar.Koerner@hre-ftr.f.rd.honda.co.jp

Abstract. In sensory recognition, rapid activation of an initial hypothesis based on few but reliably detected features, and the discrimination of the object by refined analysis of the sensory input are two conflicting requirements. We show that these two aspects of sensory recognition can be optimally served by a neocortical firmware model, the stereotypically structured columnar architecture, and that the resulting ensemble temporal encoding performed by such a modular unit enables both rapid and robust recognition by utilisation of the dimension of time to encode the reliability of a decision. The model suggests a novel functional interpretation of gamma oscillations in cortical processing.

1. How to trigger and refine a hypothesis in neural systems?

The marvelous performance of vision must be based on an active search for interpretation guided by top-down feedback that integrates the details of sensory analysis into a globally consistent hypothesis. However, bi-directional processing by itself does not resolve the problem of combinatorial explosion. For making a proper prediction, a hypothesis on the sensory input is both a prerequisite and result of processing. The activation of an initial hypothesis in forward processing for a subsequent bi-directional refinement must be made as rapid as possible, and, with a high probability, it has to be correct. First, activation of an initial hypothesis should be based on evaluation of afferent stream by ensemble coincidence detection since the required time interval for observation of the input for either average spike rates or spike sequences would by far exceed the experimentally observed ones: Face selective cells at inferotemporal cortex (IT) respond with a latency of about 100 ms, which is barely several milliseconds more than the time required for pure signal transmission from retina to IT without any additional delay [7]. The principal decision on the input is seemingly done with arrival of the very first spikes at any hierarchic level by an ensemble coincidence detection [8], as it is expected from the short integration interval of cortical neurones. Second, for a both rapid and correct initial hypothesis, the barrage of afferent signals must be reduced to a sparse set of the most reliably signaled features only. Thorpe suggested that cortical neurons emit a spike in response to an afferent input with a latency that is the shorter the better the input matches the tuning characteristics of the neuron. Then, rapidly transmitted best matches trigger the recall,

while partial matches are delayed or even suppressed in a multilevel hierarchy of forward coincidence [1].

But there are two serious flaws in this scheme: 1) Encoding the degree of matching into relative latency of spiking can be evaluated at the next higher level only then if all of the local decision processes (not the spikes!) at the lower level are in synchrony to each other. 2) The coding capacity of that scheme is rather low, and the analog information provided by the partial matches cannot be used effectively.

A model for encoding of analog information of partial matches into relative time advance of spiking that provides the necessary coding capacity has been proposed by Hopfield [3], but it bears the cost of fixed latency for decision making at any level of processing. Furthermore, this model lacks a mechanism to enforce the unitary time relation between local decision processes over the entire system which is actually the necessary precondition for its proper function.

In this paper, we propose a solution to the problem in combining a code that supports fast processing but has low capacity, with a code that has an extremely high encoding capacity but requires comparatively long time for decoding at any processing level. As we will show, neocortical columns could provide the firmware control architecture to embody the proposed concurrent processing utilising the two qualitatively different codes.

2. Encoding patterns and pattern sequences at any columnar module -- Computational hypothesis

2.1. A cortical-type neuron model

The model neurons evaluate the ensemble of spikes at their input in a short integration interval (coincidence detector characteristic). Depending on basic columnar neuron types, the share of afferent, efferent and lateral contribution to the input is different. The threshold is set high to enable the release of a spike only if a sufficient number of inputs provide a spike during the integration interval that constitute the feature the neuron is tuned to. In case of best match, the spike is released with minimum latency. This latency increases with lower but still sufficient degree of matching. Non-specific depolarizing modulation of the membrane potential (MP) of all neurons of the global system enables the signaling of partial matches.

2.2. Columnar modules as elementary processing nodes in neocortical-type neural networks

The neocortex seems to be composed of complex nodes, columnar units, the basic internal organisation of which is rather similar regardless of the information represented there. We propose that this unitary architecture of columnar units represents not the structure of the knowledge stored there, but the control that forces the system to make that representations [4]. Columns are clustered to functional units, the macrocolumns. In the ventral visual pathway, a macrocolumn at the lower, still topographic representation levels, assembles all features the system has acquired to describe the respective local spot at the retinal input (like the set of oriented edges in V1), and at the higher non-topographic levels of processing a macrocolumn seems

to represent a class of features and its variations [10]. We formalise any columnar unit into 3 functional different submodules (Fig. 1): Submodules A define the prototypes encoded in a macrocolumn and perform a parallel categorisation, submodules B decode a sequence of activated categories, and submodules C represent patterns of selected activation trajectories at all submodules B of a macrocolumn.

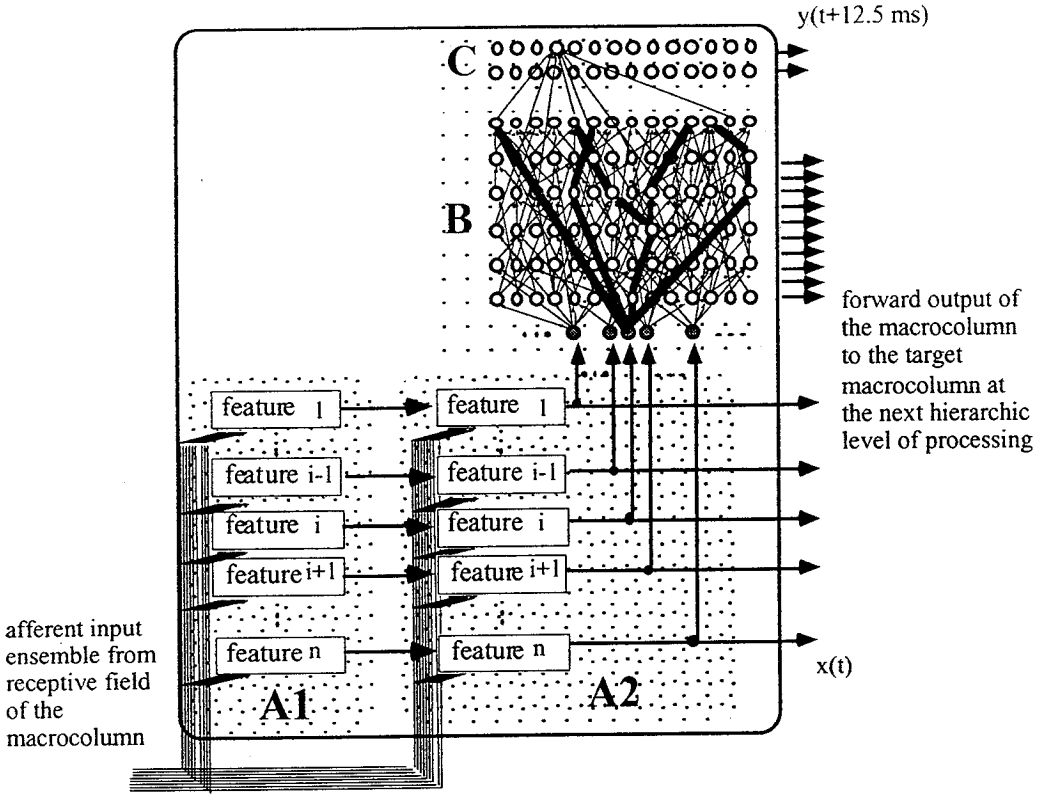

Fig. 1: Model macrocolumn composed of n minicolumnar units.
Submodule A1 generates the initial hypothesis that is transmitted by A2, and during MP modulation, A2 generates a sequence of partial matches. Submodule B consist of a population of variable tapped delay lines that re-code any spike sequence into a localised object representation at submodules C. A1 is related to the spiny stellate of layer 4 (SS4). SS4 make multisynaptic contacts to selected pyramids of layer 3 (Py3), which we interpret as a strong driving input that elicits a spike at a Py3 if the respective SS4 fires. SS4 make contacts within the local processing module only. Py3 forward its output both to the next hierarchic level of processing and (via its ascending axon collateral's) to the supragranular pyramids of the same macrocolumn. Afferent inputs contact not only SS4 (A1), but to a weaker degree also lower Py3 (A2). We relate B to the upper neocortical layer 3 because the system of ascending axon collateral's of supragranular pyramids that make en-passant-type connections to other supragranular pyramids within the macrocolumn could serve as such tapped delay lines. Neurones in C (which we relate to layer 2) represent those traces as sparse and very specific parallel object descriptions which are signalled with a fixed delay after the start of MP modulation.

2.3. Rapid activation of initial hypothesis

Neuronal feature detectors give a graded response to many input configurations. For the next level decision making unit, only a perfect matching of the input to the feature the respective neuron is tuned to delivers by itself a definite decision on the sensory input. The lower the analog activation of a neuronal feature detector, the more the brain needs an already established context to properly interpret the decision of that feature detector. Hence, only the response of strongly activated feature detectors should be used to select the initial hypothesis. We assume the threshold of the feature detectors in A is set high: only rather perfect hits release a spike at all. A fast forward inhibition pool for all A1 submodules of a macrocolumn (Fig. 1) permits the signaling of one best match per macrocolumn only. Then, that very limited number of reliably detected features trigger an initial hypothesis at the respective next higher level, steering clear of the combinatorial explosion problem. The activation of an initial hypothesis in this system is very rapid, since no time for decoding is required at any level of processing. The very first spikes define the decision while less good matches are discarded from the afferent stream. The disadvantage of this procedure is the loss of information on any less salient features of the input because of the missing capacity to encode analogue scale evaluations.

2.4. Refinement of initial hypothesis by analog similarity measures

Once the system has managed to select a proper and reliable hypothesis, it has to include the less salient features into its analysis of the input to verify and refine the hypothesis. A monotone increasing depolarisation of MP would enable the release of spikes by feature detectors having a less salient feature at its input. The ranked time sequence of spikes across the ensemble of afferent channels could precisely encode their difference of the degree-of-fit of feature to input [3]. Since the ensemble coincidence detection in our A submodule system cannot encode this ranked sequence of partial matches, an additional module has to be provided at any local processing node. We propose that the sequence of partial matches generated at any macrocolumn in submodules A is decoded by submodules B and C (Fig. 1).

2.5. Clocking of decision intervals in cortical processing

To make such a dual encoding scheme work in a large-scale neural system, three necessary conditions have to be guaranteed: 1) To rectify local decisions according to a globally consistent hypothesis and to follow rapid fluctuations of the input, the decision process has to be repeated several times with a sufficient high frequency. 2) To enable the evaluation of the sequence of partial matches, a time interval must be prepared in-between any recurrence of the decision process since decoding of spike sequences bears the cost of a fixed delay at any stage of processing. 3) To encode analogue values into relative latency in a large-scale neural system, the modulation must be synchronous over all the system. Moreover, this definition of a globally synchronous time interval for refined analysis must be in a definite relation to the initial hypothesis which has to be refined.

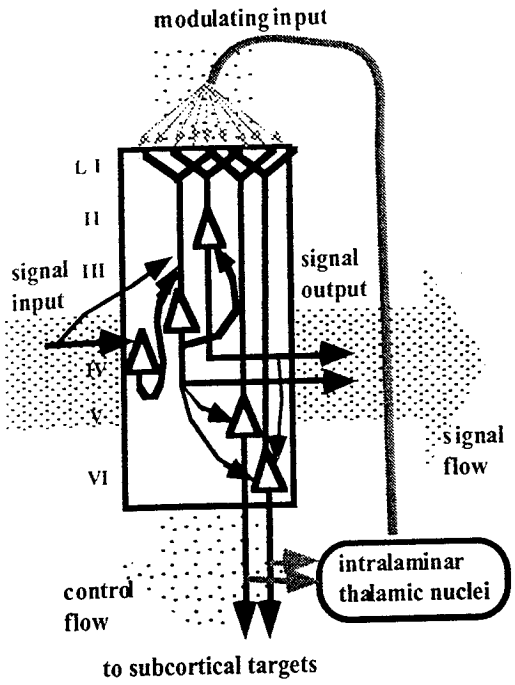

Fig. 2: Global synchronisation of decision intervals by gamma modulation

We propose that the initial wave of most reliably detected features in the A-submodule system triggers gamma activation of the thalamic intralaminar nuclei (ILN) and that the ILN modulate the MP of the cortical pyramids to enforce a synchronous decision interval for analogue encoding onto the cortex (Fig. 2). The ILN receive inputs from infragranular pyramids from all over the cortex, and return a non-specific modulating feedback to apical dendrites of pyramids in layer 1 [5]. All principal neurones of cortical columns extent apical dendrites to layer 1 (and may undergo the here proposed modulation by ILN), except those that are the major recipients of afferent input (like spiny stellate of layer 4, SS4).

3. Ensemble temporal encoding by columnar modules and a novel interpretation of gamma oscillations

We hypothesise that feature detector neurones of layer 4 in a macrocolumn perform a competitive recall of a best matching feature, the winners of which activate a lower Py3 that rapidly transmits the "express" feature to the next hierarchic level of processing. Fast forward activation of inhibitory neurones in layer 4 prevents any further activation of SS4, and is assumed to cause there a gamma period hyperpolarization [11]. The highly coherent wave of spikes signalling "express" features is assumed to trigger additionally via ILN a gamma period depolarisation of the MP of principal neurones of neocortical columns except SS4. Without modulation, lower Py3 can fire only by driving input from SS4, since the share of afferent inputs is too low to trigger a spike by itself. Because of the weak inputs by en-passant contacts of ascending axon collaterals of Py3, submodules B and C cannot be activated without MP modulation. Only the coherent wave of "express" features propagates upwards the hierarchy of the submodule A system. After this initial hypothesis at any local processing node of each level of processing, the A submodules are blocked for the remainder of the gamma period, while with a certain delay caused by the conduction time from cortex to ILN and back, the ranked sequence of partially matching features are released by the lower Py3 (A2) within the 12.5 ms of raising depolarisation of MP. The global modulation of MP enables the analysis and re-coding of this ranked sequence in a refinement process at all local processing units as proposed above. Since supragranular pyramids send their axons to the next hierarchic

level and to more distant targets of the same level of processing, and since infragranular pyramids feedback the processing state of a columnar unit to lower levels of processing, the refinement process in B and C submodule system is a recurrence of an indeed parallel optimisation guided and controlled by the preceding initial hypothesis in the A submodule system.
Resulting activity patterns in the proposed system resemble experimentally observed ensemble temporal coding [9]. The time course of fast transmission of spikes to next hierarchic level by lower pyramids of layer 3 [1], the concurrent course of activation of neighboured hierarchic levels [6], and the subsequent activation of the supragranular pyramids in an interval of 12...15 ms after layer 4 [2], support our scheme.

4. Conclusion

We claim that the proposed architecture and the described type of ensemble temporal encoding enables a novel type of fast but robust recognition, that is refined in an analysis-by-synthesis manner [4]. Furthermore, we suggest that not synchronisation of spikes, but synchronisation of decision intervals by a global modulation may be a biologically plausible explanation for the experimentally observed gamma oscillations.

References

1. S. Celebrini, S. Thorpe, Y. Trotter & M. Imbert. Dynamics of orientation coding in area V1 of the awake primate. Visual Neuroscience 10 (1993): 811-825
2. R.J. Douglas & K.A.C. Martin. A functional microcircuit for cat visual cortex. Journal of Physiology 440 (1991): 735-769
3. J.J. Hopfield. Pattern recognition computation using action potential timing for stimulus representation. Nature 376 (1995): 33-36
4. E. Koerner, H. Tsujino & T. Masutani. A cortical-type modular neural network for hypothetical reasoning. Neural Networks (in press)
5. R. Llinas & U. Ribary. Coherent 40-Hz oscillation characterizes dream state in humans. Proc.Natl.Acad.Sci.USA, 90 (March 1993): 2078-2081
6. L.G. Nowak, M.H.J. Munk, P. Girard & J. Bullier. Visual latencies in areas V1 and V2 of the macaque monkey. Visual Neuroscience 12 (1995): 371-384
7. M.W. Oram & D.I. Perret. Time course of neural responses discriminating different views of the face and the head. Journal of Neurophysiology 68 (1992): 70-84
8. S. Thorpe, D. Fize & C. Marlot. Speed of processing in the human visual system. Nature 381 (1996): 520-522
9. E. Vaadia, I. Haalman, M. Abeles, H. Bergman,Y. Prut, H. Slovin& A.Aertsen. Dynamics of neuronal interactions in monkey cortex in relation to behavioral events. Nature 373 (1995): 515-518
10. G. Wang, K. Tanaka & M. Tanifuji. Optical imaging of functional organization in the monkey temporal cortex. Science 272 (1996): 1665-1668
11. M.A. Whittington, R. D. Traub & J.G.R. Jeffreys. Synchronized oscillations in interneuron networks. Nature 373 (1995): 612-615