

Parallel Implementations of the RCE algorithm

Jean-Didier Legat, Philippe Thissen, Michel Verleysen, Jean-Luc Voz

Microelectronics Laboratory
Université Catholique de Louvain
B-1348 Louvain-La-Neuve
Belgium

Tel : +32.10.47.25.68
Fax : +32.10.47.86.67
E-mail : Legat@dice.ucl.ac.be

Abstract

This paper describes two different parallel architectures for the implementation of the neural network RCE algorithm. The first architecture is a SIMD system based on an ASIC chip including 8 units and all the associated logic which was designed and fabricated in a standard 1.5 μm CMOS technology. Secondly, a bit-serial massively-parallel associative processor is presented. This system will allow the implementation of a full classifier for pattern recognition applications on one single chip.

1. Introduction

The RCE algorithm is a well-known method used for classification tasks [1]. It is based on incremental modifications of a neural network structure. Due to its organisation and as for most of the ANN-based systems where real-time execution is needed, hardware accelerator is required [2]. Parallel implementation is quite well adapted to these classes of algorithms.

In this paper, we present two different approaches. The first system implements a full parallel SIMD architecture. Each unit of the RCE model is associated with a dedicated operative part which contains a simple ALU and its own set of registers. An ASIC chip including 8 units was designed and fabricated. The architecture of this system and the organisation of the VLSI circuit will be described after a short presentation of the RCE algorithm.

The SIMD processor is a high performance system. However, to implement a large number of units (for example, a few hundreds) it requires a huge silicon area. Another disadvantage is that the dimension of the input vector and the algorithm are fixed. For applications where programmability and high density of integration are required, a new bit-serial associated processor has been designed. The architecture of this massively-parallel system will be presented.

2. The RCE Algorithm

The Restricted Coulomb Energy (RCE), proposed by Reilly, Cooper and Elbaum, is one of the first incremental models of neural networks. In this model, decision units are characterised by their influence region, defined by an hypersphere around the unit, whose radius is equal to the threshold of the unit. The state space is then divided into zones, each dominated by different decision units. New units are created with an initial chosen radius if a presented template does not fit into one of the influence regions of the units associated with the correct class; on the other hand, radii associated to units belonging to a wrong class but whose influence regions include the presented pattern are lowered to avoid this situation.

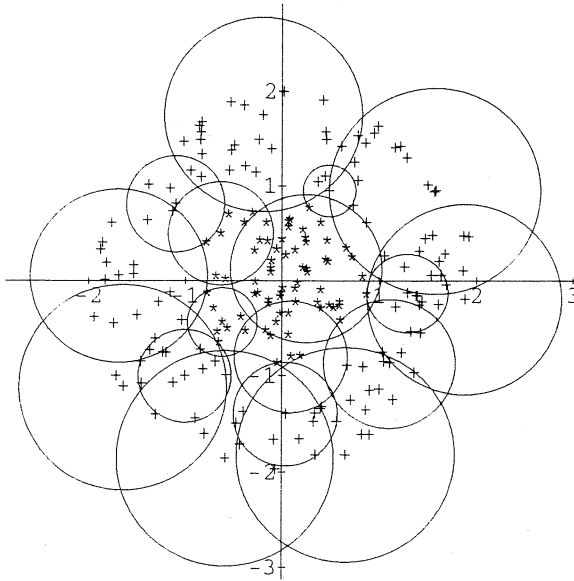
The most important advantages reside in the simplicity of the algorithm and its ability to modify the number of units. The decision regions are built with respect to a threshold associated to each decision unit. The algorithm has only to decide whether a new point belongs to an existing class and, in this case, if it is correctly classified. Then, the decision of creation of a new decision unit is taken, according to the result of this comparison. A new decision unit defines a new hypersphere whose center coordinates are chosen as the actual input prototype.

The creation of class units is also decided, by a simple comparison between the activity of the class units and the desired class given by the supervisor.

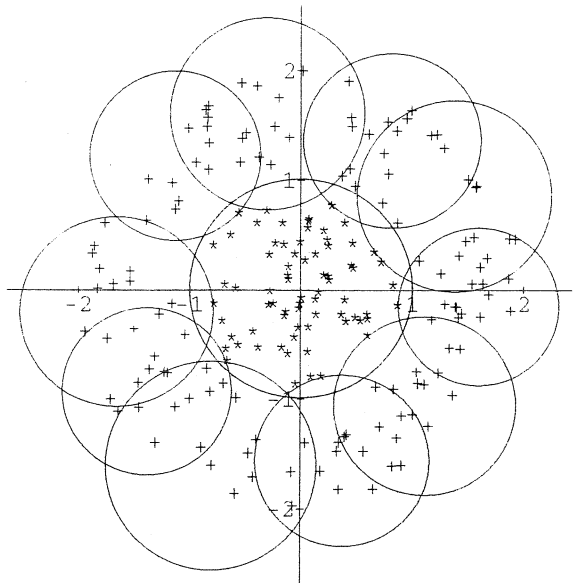
The whole set of templates is presented as many times as necessary to obtain a stable configuration (no more units creation, no more radii decrease). Such a procedure guarantees the convergence. The worst case would be the creation of as many decision units as there are prototypes used for the learning. As the set of prototypes is finite, and also because the radius associated to each decision unit always decreases, the algorithm converges in a finite number of steps.

At the end of this procedure, the ratio of correct classification (number of correctly classified templates

versus the total number of templates) for the learning set is 100%.



a : without initial conditions



b : with initial conditions

Figure 1 : Example of circular and ring distributions

One of the main drawbacks of this algorithm is its sensitivity to initial conditions. This includes the initial radius associated to each decision unit, and the choice of the presentation order of the prototypes. A simple procedure to choose the initial values of the decision unit vectors has been proposed in [3]. This procedure reduces the number of decision units created during the learning phase, and increases the conver-

gence of the algorithm. It also decreases the size of the overlapping regions.

As an example, figure 1 illustrates a 2-class problem where one class is a circular distribution, which is a very simple problem for the RCE algorithm, and the other is a ring distribution, which is more difficult to cover. The RCE algorithm, without preprocessing (adequate choice of initial conditions), has allocated 4 units to cover the circular distribution, and 13 units for the ring. The variance of the circle radii is important. With preprocessing, the circular distribution is covered with only one unit, and the other 10 units have almost identical radii. In that sense, the covering is much better.

3. SIMD Implementations

The only difference between the original RCE algorithm and our implementation resides in the metric used to compute the influence field. To allow a simple and compact VLSI implementation, the Manhattan distance (the sum of the absolute differences) has been chosen instead of the standard Euclidean distance. Without a priori information about the distributions, it has been verified experimentally that the choice of the distance metric does not affect the performances of the algorithm

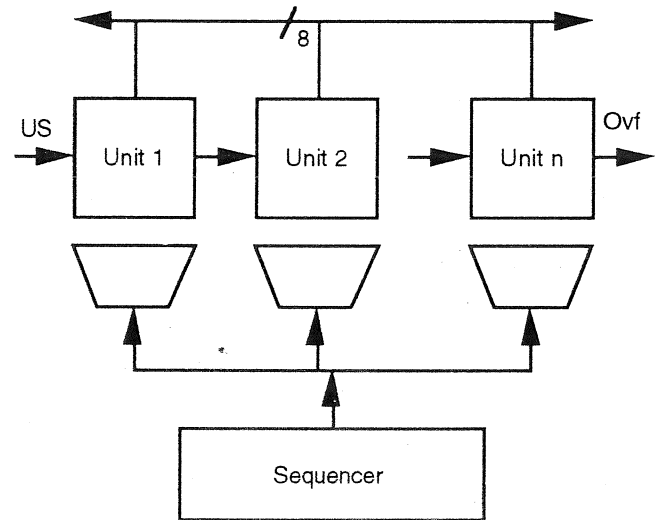


Figure 2 : SIMD General Architecture

The general architecture of the SIMD processor is shown in figure 2. Each unit is connected to a common 8-bit bus used to initialise the registers and transfer the input vector of the pattern to classify. A 1-bit signal US (Unit Select) enables to add a new active unit when no classification occurs.

Figure 3 illustrates the block diagram of the operative unit : it includes a simple ALU to perform integer addition or subtraction and a set of six 8-bit registers. Four are used to store the coordinates, the radius of the hypersphere and the class associated with this unit; while the two last registers are working registers used to memorise intermediate data.

A cascadable ASIC circuit which implements 8 units and their decoding logic was designed. The chip was fabricated in a standard 1.5 μm CMOS technology thanks to Eurochip [4]. The die size is 67 mm^2 . A microphotograph of a part of the circuit is shown in figure 4.

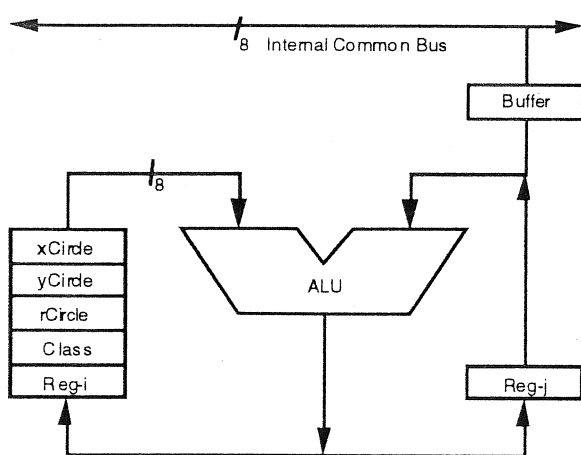


Figure 3 : Unit Block Diagram

4. Associative Processor Architecture

An associative memory can be defined as a memory system in which stored data words are identified according to their contents [5]. Figure 5 shows a basic classical associative memory block with an array of content-addressable cells, a comparand register and a mask register. Results of the comparisons are stored in the Tag register.

By increasing the complexity of the basic cell, it has been shown that parallel logic and arithmetic operations can be performed on data stored in the associative array [6][7].

The RCE algorithm can be easily implemented in this kind of architecture. Each unit is associated with a specific word of the array. A word is selected (the Tag register being set to one) when the n-dimensional input vector is included into the centroid of the influence region. This is simply done by computing the Manhattan distance as for the SIMD system. Some additional logic has to be added to inform the

control unit when more than one unit is selected and to be able to activate a new unit when an unclassified input vector is found.

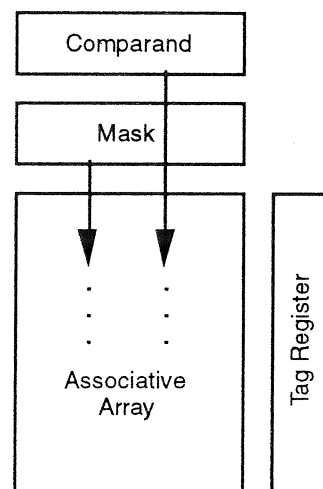


Figure 5 : Classical Associative Memory Model

Because of the expensive logic in each memory bit and the communication problem in fully parallel associative processors, a bit-serial word-parallel architecture using the concept of parallel processing with vertical data has been chosen. One bit-column of each word being processed simultaneously, only one ALU is implemented by word.

The block diagram of a word is shown in figure 6. Each storage cell can be accessed in read or write mode with a 1-bit bus for each mode. This allows to complete an addition of two n-bit operands in n+k clock cycles (k being the number of cycles needed to initialise the operation).

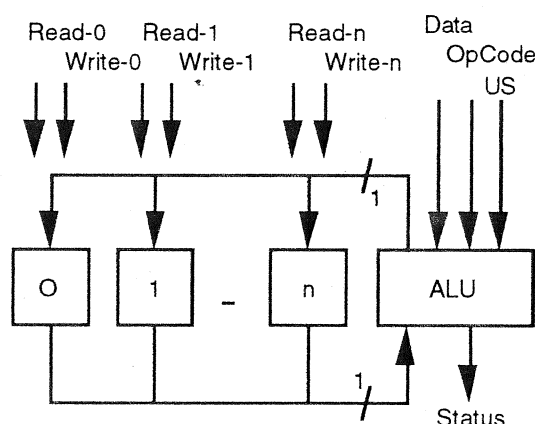


Figure 6 : Word Block Diagram of the Bit-Serial Word-Parallel Associative Processor

A CMOS full custom integrated circuit which implements an associative processor of 512 prototypes is being designed. Each prototype consists of a bit-serial associative word of 512 storage cells which can be organised, for example, in 64 dimensions of 8-bit resolution. A great effort will be done on the speed optimisation in order to obtain a clock frequency as high as possible with the technology. The chip can be cascadable if the application needs more than 512 prototypes.

5. Conclusions

Two parallel implementations of the RCE algorithm have been presented. The first system is based on a ASIC circuit which is fully operational. For applications with high numbers of prototypes, a bit-serial associative processor has been designed. Its full custom VLSI implementation is being developed.

References

- [1] D.L. Reilly, L.N. Cooper, C. Elbaum. "A Neural Model for Category Learning". *Bio. Cybern.*, vol 45, 1982, pp. 35-41.
- [2] Intel Neural Network Group. "Design and Implementation of a Recognition Accelerator". *Proceedings of the Canadian Conference on Very Large Scale Integration*, 1993.
- [3] M. Verleysen, F. Blayo, J.D. Legat. "LVQ-like procedure for the initialization of the RCE model". *Congrès européen de mathématiques*, Paris, July 1992.
- [4] P. Giot. "VLSI Implementation of an Evolutive Neural Network". *Final Work*, Université Catholique de Louvain, Louvain-la-Neuve, June 1993.
- [5] S.S. Yau, H.S. Fung. "Associative Processor Architecture". *Computing Surveys*, vol. 9, n° 1, March 1977, pp. 3-27.
- [6] I.D. Scherson, S. Ilgen. "A Reconfigurable fully parallel associative processor". *Journal of Parallel and Distributed Computing*, vol. 6, 1989, pp 69-89.
- [7] I.D. Scherson, D.A. Kramer, B.D. Alleyne. "Bit-Parallel Arithmetic in a Massively-Parallel Associative Processor". *IEEE Transaction on Computers*, vol. 41, n° 10, Oct. 1992, pp 1201-1210.

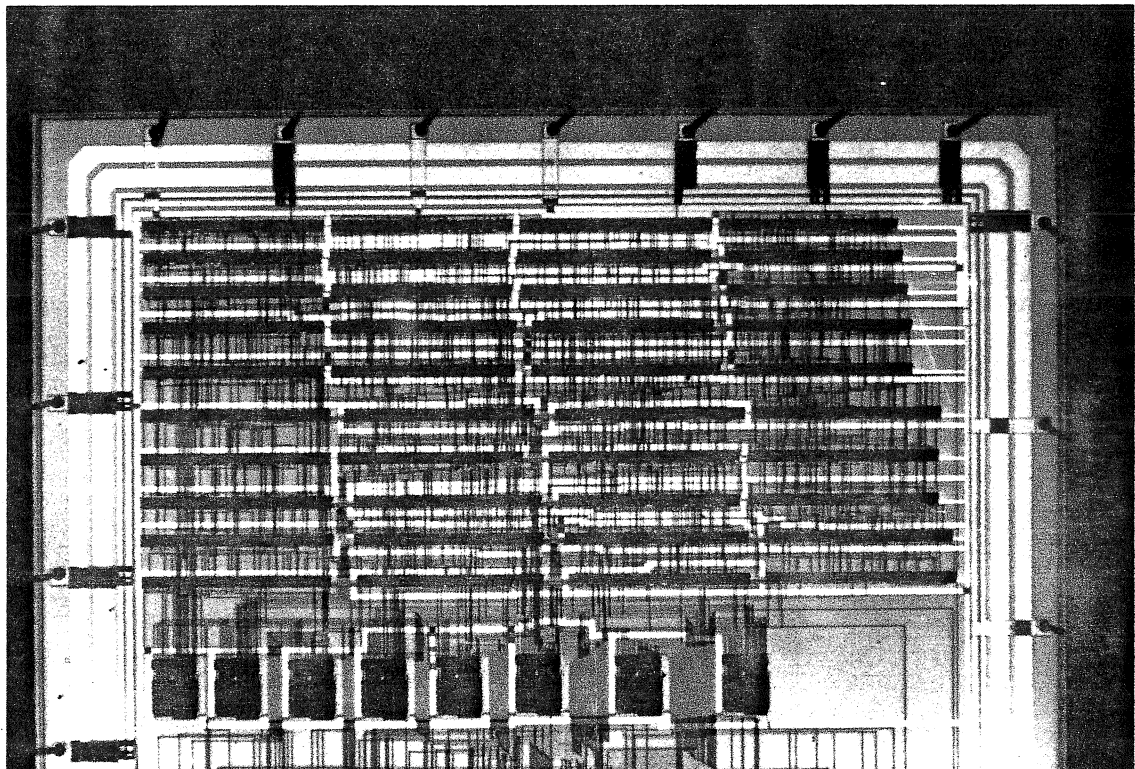


Figure 4 : Chip Microphotograph