High Speed Color Recognition with an analog Neural Network Chip

Gundolf Geske, Frank Stüpmann, Ansgar Wego
Silicann Technologies
Joachim-Jungius-Strasse 9, 18059 Rostock, Germany
stuepmann@silicann.com

Abstract—**The neuro chip introduced is a classificator which is intended for fast classification of signal vectors up to the width of 10. It consists of analog components. The width of the output vector is also 10. Due to the implementation of analog hardware, the chip works fully parallel and needs less than 5 µs to recognize a pattern.**

The analog approach necessitates that capacitive storage elements are used for storing synaptic weights. The storage of analog voltages in a capacitor of only 1 pF with a precision of more than 6 bit is possible for a period of time of up to several minutes by suitable circuit technique. To fulfill vector-matrix multiplications two arrays of 66 and 70 analog multipliers are integrated. The advantage of the analog approach in terms of speed, however, requires a high effort in modelling complex transfer function.

We show that the circuit is able to perform color recognition tasks in combination with an analog sensor. Results show that color recognition can be achieved with a precision sufficient for the demands of the human eye. By segmentation of the color space, the neural network can be trained with a precision beyond the spectral resolution of the human eye.

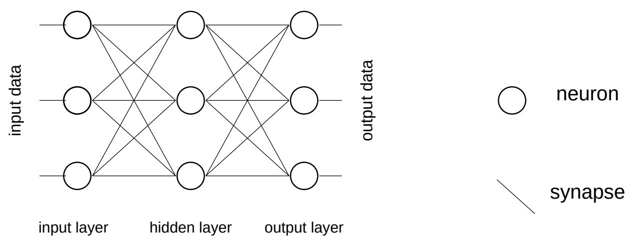

Fig.1: Multilayer perceptron, which is a nonlinear network that consists of an input layer, an output layer, and one hidden layer. One hidden layer is sufficient in order to be a universal approximator for any continuous nonlinear function.

1 Introduction

Neural networks are a very promising computational technology due to their capabilities in modelling and solving complex problems that are hardly approachable with traditional methods such as statistical pattern recognition and conventional artificial intelligence. The most attractive feature of neural network paradigms is their learning capability by which they can solve problems where formalization is not known or where solution is available only in some instances used as training examples [6]. Artificial neural networks (ANNs) can also perform human brain-like tasks such as object and pattern recognition, speech recognition and solving classification problems.

ANNs usually consist of a number of simple processing elements called neurons that are interconnected with each other. In most cases, one or more layers of neurons are considered that are connected in a feedforward or recurrent way [8]. The strength of the interconnections is quantified by means of interconnection weights. These weight values are calculated in a training process, so that a neural net is not programmed but it is trained.

There are many ways of implementing a neural network: many kinds of software algorithms and hardware implementations. In the field of hardware, we distinguish two main approaches: operations of neural networks can be implemented in analog or digital circuit style. Analog circuits are appropriate to the implementation of parallel computing structures. Analog implementations of multiplier matrices are very efficient because these matrices are regular arrays built out of small circuits. The computational speed of these circuits are relatively low (several 100 kHz) but all circuits are working in parallel which leads to a powerful architecture. In contrast, digital circuits can consist of very fast components, but the operations are performed sequentially. Therefore, analog neural networks can be superior to digital implementations which are e.g. implemented in a DSP. Furthermore, analog circuits, which compute mathematical functions like additions, subtractions and multiplications, are smaller and consume less power than corresponding digital circuits. These features of analog circuits make it possible to achieve true parallel computation by placing many circuits on a single chip.

Our current neural network chip consists of 10 input neurons, 6 hidden and 10 output neurons. Furthermore, there are two bias neurons. The number of synapses amounts to 136. The processing speed for the classification of one pattern is less than 5 µs.

2 Problem of color recognition

To verify the neural network chip's ability to function, several benchmark problems, such as breast cancer classification problems [1, 2, 3, 7], known to neural network researchers, are tested with good results.

Another recognition problem to be processed is the color detection. We use a color sensor delivering three analog current signals which are converted by transimpedance converters to voltage signals. These signals are scaled to the input range of the neural network chip and can be evaluated in this way. The outputs of the sensor correspond to the colors red, green and blue. The spectral sensitivities of these outputs are represented in figure 2. This diagram shows overlappings between the color channels what is necessary in order to reflect the whole color space without gaps.

Training data also have revealed this behavior. Figure 3 shows sensor voltages that were sampled by scanning a TFT display. The display has emitted 216 different RGB values (8 bit per channel) which were used for the training of the neural net and are represented by the sensor voltages in the diagram. It is clearly recognizable that e.g. the blue color channel is influenced by green or red spectral quotas and vice versa. Therefore, a linear transformation of the sensor voltage into a RGB color space is not possible straight away. Usually, correction matrices are calculated with high effort to enable a transformation. However, this can be carried out also with a neural net. The advantage of a neural net is that the non idealities of the sensor are compensated by including them into the training process.

Beside the transformation of sensor signals into a color value space (e.g. RGB) the neural network chip can be used for classifying incoming signals in different classes of colors.

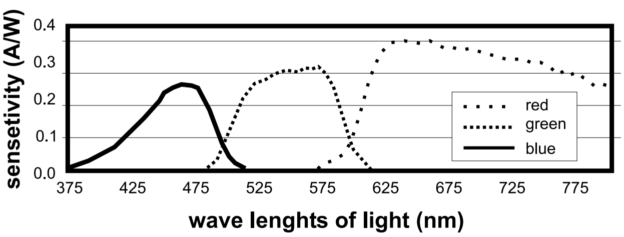

Fig.2: Spectral sensitivity of the sensor outputs [5].

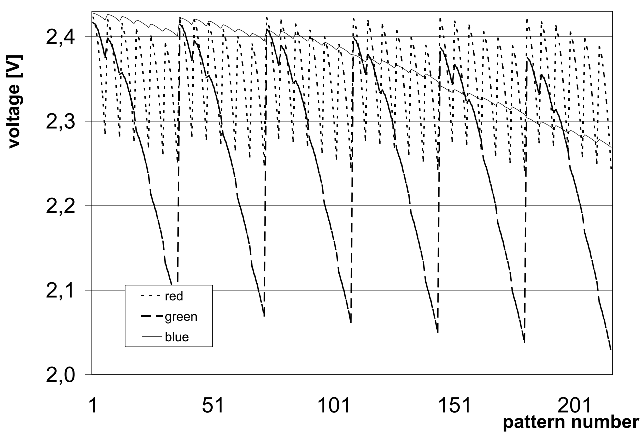

Fig.3: Sensor voltages of TFT-monitor measurement for RGB values between 0 and 250 in steps of 50.

3 Training and Test with a chip model

In order to solve the color recognition problem the neural net is trained with a training software. The software contains a complete model of the chip with its circuit behavior.

Modelling of the neuro chip for off-chip training can be carried out in two ways: analytically and numerically.

1. With the analytic method used circuit blocks are modelled mathematically on basis of composition and sizes of transistors.

2. The numeric method as a second variant has proved itself more effective in practice. The transfer functions of single blocks are investigated by simulation and saved for later use in simulations on system level. The determination of the synaptic weights by training is executed on basis of these transfer functions.

Sensor voltages and the corresponding RGB values are brought together in the training software as a vector pair during the training. The synaptic weights are adapted as the learning cycles are repeated, so that the output error of the network decreases until a predefined cut off criterion is reached. The backpropagation algorithm (BPA) is used for the training. We use as a cut off criterion either a percentage mean squared error (PMSE) as follows[1] [4]

$$E = 100 \cdot \frac{1}{N \cdot P} \sum_{p=1}^{P} \sum_{i=1}^{N} (o_{pi} - t_{pi})^2 \tag{1}$$

[1] o_{min} and o_{max} are the minimum and maximum values of output coefficients in the problem representation (assuming they are the same for all output nodes), N is the number of the network output nodes, and P is the number of patterns (examples) in the data set considered. t_{pi} is the training value.

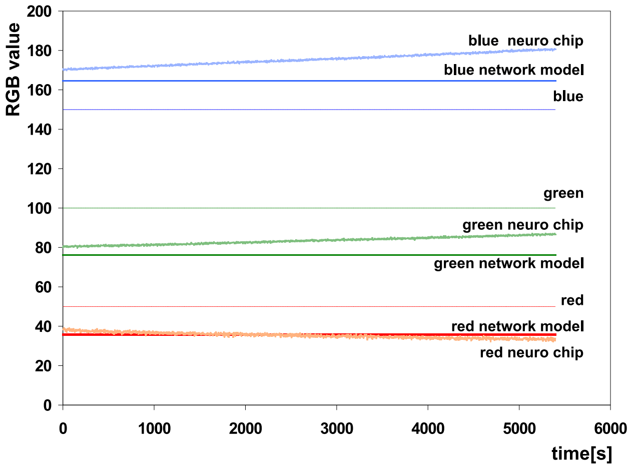

Fig.4: The color curves over a period of time of over 1.5 h for the network model (software) and the neural network chip; the drift appears by unloading of the storage capacitors of the chip.

or the maximum absolute error $E_{max} = \max(\text{abs}(o_{pi} - t_{pi}))$.

If one puts sensor voltage values to the inputs of the neural net after the training, the corresponding RGB values appear at the output[2]. The precision of these output values depend on different factors of influences:

1. on the kind and the quality of the training data sets;
2. on the topology of the net: topology parameters are the number of the hidden neurons or the number of hidden layers;
3. circuit restrictions enhance the difficulties in terms of precision. The limitation of the weight values due to the restricted voltage ranges belongs to this restrictions. In addition, the multipliers's range of values is limited so that the transfer function has a tanh-like appearance.

The first and second factors of influence are typical neural network characteristics and they occur in a similar manner in software as in hardware neural nets. The third factor is a quality of the analog circuitry and is specific to the neural network chip. This characteristic can be simulated when training the net by software and it can be taken into account when calculating the sets of weights. Another quality of the analog circuitry is not ascertainable for the software net and that are coincidental variabilities, such as offsets. They cause a difference between the model and the chip behavior which cannot be reduced by off-chip training.

For the off-chip training 216 RGB values (per color channel the values 0, 50 100, 150, 200 and 250) of a LC-display were taken with the sensor. The off-chip training shows an absolute mean difference of 12.2 RGB values. After the download of the weights into the chip a mean error over all patterns of 17.5 RGB values is reached. The difference of the output pattern between the off-chip model and the real chip is 9.4 RGB values on average.

[2]The outputs conduct analog continuous voltages; to get RGB values a scaling has to be performed.

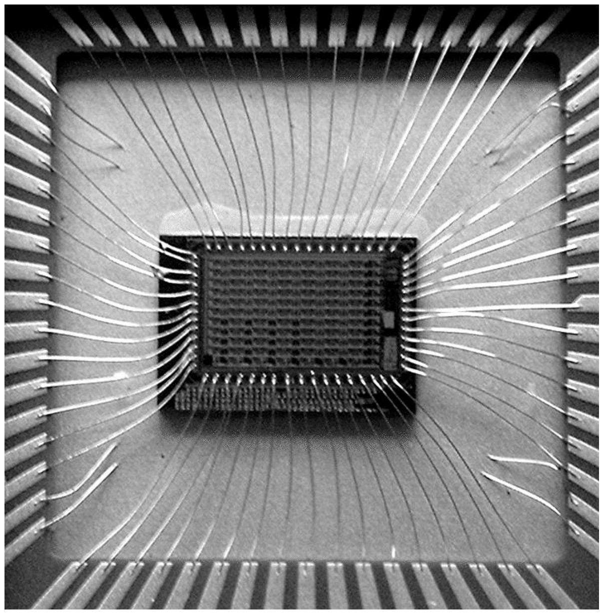

Fig.5: Analog Neural Network Chip with the topology of 10 input neurons, 6 hidden neurons and 10 output neuron. The number of synapses amounts to 136. The size of the die is approx. 10 mm^2.

The behavior in terms of precision is represented for one special example in figure 4 which shows that the predefined training values of 50, 100, 150 (RGB) and the output of the neural network model (software) diverge. Deviations through coincidental influences in the real chip are represented with less than 10 RGB values in the diagram. In the temporal course, a drift of the outputs which is caused by changes of the capacitors charges can be observed. The drift is extremely low if one considers that the weight voltages are stored in capacitors with a capacity of only 1 pF.

4 Chip topology and performance

Figure 5 shows the neuro chip that contains a 10:6:10 topology with 136 synapses. In addition to the 26 neurons two bias neurons are included which provide a constant value. These bias values improve the performance of the net in order to solve more complex problems.

Because of the analog circuit technique inputs and outputs have a continuous range of values. Since the neural net is not clocked the time domain is also continuous. The performance of a neural net is indicated in connection per second (CPS). Accordingly the meaning

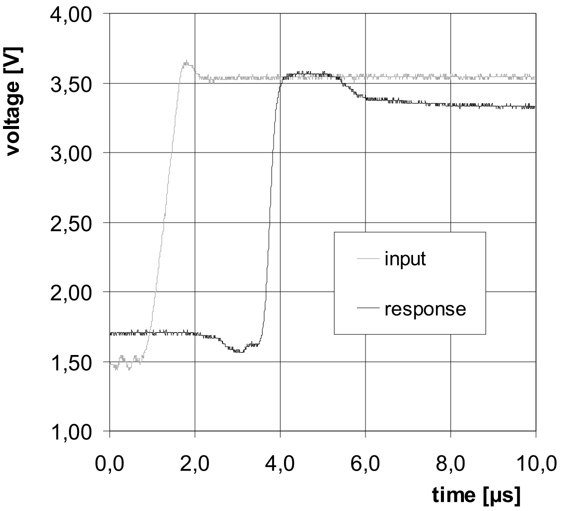

Fig.6: Measuring results shows a pin to pin response time of approx. 5 µs. The output currents of ±500 µA are transferred by a transimpedance converter to an output range of 1.5 . . . 3.5 V.

of 1 CPS is one multiplication addition operation per second. Because of the relatively high parallelism the chip achieves a processing performance of 27.2 MCPS (mega connections by second). The response time of the chip is within 5 µs as depicted in figure 6.

In combination with the described sensor, color processing tasks (transformations into color value spaces or classifications) with a color alteration frequency <200 kHz can be performed.

5 Conclusions & Outlook

The study has been shown that color detection can be carried out with an analog neural hardware net. An essential advantage is the speed of the analog chip which enables applications with high real time requirements. The restricted precision of the analog circuits is adverse. We have found, however, that the appearing limitations to a large extent are caused by the network topology and to a small degree by the circuit inaccuracies.

To improve the results in future, another chip with a changed topology is in preparation. This contains a broader hidden layer. To reach the appropriate accuracy the fault influence of the analog circuitry must be reduced. The attempt of producing more precise circuits is not made, since the occupied silicon area would increase too strongly.

1. Instead the coincidental fluctuations are included in the training model. This presupposes, however, that every single circuit must be accessible for measuring. A design for testability which enables access on every internal circuit was developed for this purpose.
2. Another way of increasing precision is to train the chip in an "'in the loop strategy"'. Therefore, a statistical approach was chosen (not described in this paper) in which the synaptic weights are perturbed randomly an weight sets that reduce the error are saved. Circuit inaccuracies can be eliminated through this approach.
3. The partitioning of the complete color space in sub color spaces for each of which a training is performed is another action. Thus, several sets of weights have to be stored and the attainable precision is only dependent on the qualities of the sensor.

References

[1] K. P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34, 1992.

[2] O. L. Mangasarian, R. Setiono, and W. H. Wolberg. Pattern recognition via linear programming: Theory and application to medical diagnosis. *SIAM Publications, Philadelphia, "Large scale numerical optimization"*, pages 22–30, 1990.

[3] O. L. Mangasarian and W. H. Wolberg. Cancer diagnosis via linear programming. *SIAM News*, 23(5):1 & 18, September 1990.

[4] Lutz Prechelt. Proben1 — A Set of Neural Network Benchmark Problems and Benchmarking Rules. *Universitt Karlsruhe, Fakultt fr Informatik*, 21/94, September 1994.

[5] Data Sheet. MCS3AT/BT 3-elementiger Farbsensor. *MAZeT GmbH, Gschwitzer Str. 32, 07745 Jena*, (V1.8), Februar 2001.

[6] Bing J. Sheu and Joongho Choi. *Neural Information Processing and VLSI.* Kluwer Academic Publishers, Boston, 1995.

[7] W. H. Wolberg and O. L. Mangasarian. Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the National Academy of Sciences, USA*, 87:9193–9196, 1990.

[8] Jacek M. Zurada. *Introduction to artificial neural systems.* West publishing Company, St. Paul, 1992.