Hamiltonian Neural Nets as a Universal Signal Processor

Wieslaw Sienko, Wieslaw M. Citko
Technical University of Koszalin
Gdynia Maritime Academy
81-225 Gdynia, Morska 83, Poland,
wcitko@am.gdynia.pl

Bogdan M Wilamowski
College of Engineering
University of Idaho
800 Park, Blvd., Boise, ID 83712, USA
wilam@ieee.org

Abstract – **This paper presents how to find an architecture for very large scale lossless neural nets, which can be used as Haar-Walsh spectrum analyzers. This analysis relies on the orthogonality of weight matrices W, where W could be Hurwitz-Radon matrices. The unique feature of these nets is the possibility to treat them either as algorithms or as Hamiltonian physical objects (Haar-Walsh Signal Processors).**

I. INTRODUCTION

Artificial neural nets can bring a revolution to real-time signal processing if they are integrated in silicon as autonomous intelligent microsystems. The design of such systems in silicon is one of the most challenging engineering problems, and requires, besides the technology, an understanding of neurobiology, nonlinear dynamical systems and physics. Many architectures of artificial neural nets in terms of their adaptation and learning, are inspired by suitable prototypes from biological systems. Moreover, biological systems demonstrate another remarkable property, namely, consisting of some billion connected in recurrent loops elements they are stable! To our knowledge, it has not been finally pointed out which architecture and learning algorithms of artificial neural networks are most suitable to be implemented in VLSI technology [1, 2]. Inspired by known results from classical and quantum mechanics we aim at showing that the very large scale artificial neural nets should be implemented as passive or particularly as lossless structures. These technical notions mean that from a mathematical point of view they should be Hamiltonian systems. It is worth noting that:

1. passivity implies BIBO stability of the structure.
2. passivity of the structure can be attained by a compatible connection of elementary passive building blocks i.e. neurons.

II. HAMILTONIAN NEURAL NETS

A general description form of an autonomous Hamiltonian system is given by the following state-space equation:

$$\dot{x} = JH'(x) = v(x) \qquad (1)$$

where: $v(x)$ – a nonlinear vector field
and

$$-J = J^T = J^{-1} \quad (J - \text{skew-symmetric matrix}) \quad (2)$$

There is such a basis in R^{2n}, where matrix J has a form:

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} ; \ 0, 1, -1 \ \text{are (n x n) diagonal matrices} \quad (3)$$

Function $H(x)$ is a Hamiltonian energy of the system. Since in Hamiltonian systems there is no dissipation of energy, their trajectories in the state space can be very complicated for $t \to \pm\infty$. Therefore, the basic method of the movement description is to find periodic solutions, using for example the Maupertuis principle. Equation (1) has constant solutions i.e. every point $x_0 \in R^{2n}$ such that $H'(x_0) = 0$ is the equilibrium and $x(t) \equiv x_0$ is the solution. The neural nets considered in this paper are composed of McCulloch-Pitts' type neurons. A model of such a neuron in the form of signal (information) flow network is given in Fig.1.
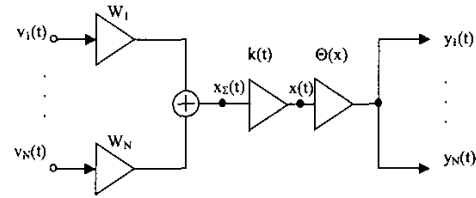

Fig. 1. A model of neuron as a signal-flow network.

Basic features of this neuron are as follows:

1. $x(t) = k(t) * x_\Sigma(t)$ (linear block) (4)

where: $k(t) = L^{-1}\{K(s)\}$, $K(s)$ – transmittance belonging to a class of lossless real functions.

2. The activity function $\Theta(x)$ belongs to class M of nonlinearities i.e.:

$$\mu_1 \le \frac{\Theta(x)}{x} \le \mu_2 \quad \mu_1, \mu_2 \in [0, \infty] \quad (5)$$

The sigmoidal activation function belongs to this class as well.

3. Energy absorbed by the neuron:

$$E = \sum_{i=1}^{N} \int_{-\infty}^{t} v_i(\tau) y_i(\tau) d\tau = \int_{-\infty}^{t} x_\Sigma(\tau)\Theta(x(\tau)) d\tau \ge 0, \forall t \quad (6)$$

For a first order neuron i.e. under the assumption that the linear block $k(t)$ is an ideal integrator, one obtains:

$$\dot{x}_\Sigma(t) = x(t) \quad (7)$$

and the energy (6) can be seen as a Hamiltonian function of the neuron. Hence:

$$E = H(x) \ \text{and} \ H'(x) = \Theta(x) \quad (8)$$

The structure of the Hamiltonian neural net (HNN) can be obtained as compatible connections of N lossless neurons fulfilling (4), (5), (6). An example of a two neuron net is shown in Fig.2.
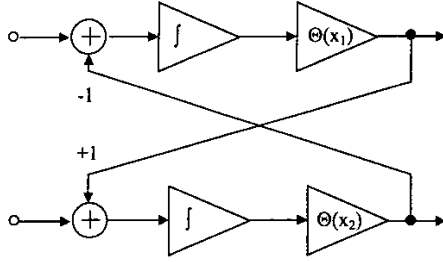
Fig. 2. Two compatibly connected neurons - a two neuron lossless net.

It can be seen that the state-space description of the net from Fig.2 is as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ +1 & 0 \end{bmatrix} \begin{bmatrix} \Theta(x_1) \\ \Theta(x_2) \end{bmatrix} \qquad (9)$$

It is a Hamiltonian system with a Hamiltonian function given by:

$$E = \sum_{i=1}^{N} E_i \qquad (10)$$

where: E_i – energy (6) absorbed by the i-th neuron.

The two neuron lossless net from Fig.2 can be treated as an elementary building block to create very large-scale neural networks. Generally, a lossless neural net composed of N neurons is described by the following state-space equation:

$$\dot{x} = W\Theta(x) \qquad (11)$$

where: W-matrix of information flow connections (weight matrix) and $W = -W^T$ (skew-symmetry).

Thus, a neural net composed of N elementary neuron pairs from Fig.2 with orthogonal weight matrix W i.e.

$$W \cdot W^T = 1 \qquad (12)$$

is a Hamiltonian system, with activation function $\Theta(x) = H'(x)$, Since

$$W^2 = -1 \quad \text{i.e.} \quad W^{-1} = W^T = -W \qquad (13)$$

so the Hamiltonian neural net can be seen as an involutional operator. The weight matrix of Hamiltonian neural net can be formulated as follows:

$$W_{2^n} = \begin{bmatrix} W_{2^{n-1}} & W_C \\ -W_C^T & -W_{2^{n-1}} \end{bmatrix} ; n = 1, 2, \ldots \qquad (14)$$

where:

$$W_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

and

$$W_C \cdot W_C^T = 1 \quad \left(\dim W_C = \dim W_{2^{n-1}} \right) \qquad (15)$$

$$W_{2^{n-1}} \cdot W_C - W_C \cdot W_{2^{n-1}} = 0$$

The simplest solution of (15) is: $W_C = 1$.

Another solution of (15) is: $W_C = W_{2^{n-1}} + 1$

Thus, for example, a weight matrix of 8-neuron Hamiltonian neural net is as follows:

Example 1.

$$W_4 = \frac{1}{\sqrt{7}} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & 0 & -1 & -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 0 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 0 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 0 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 & 0 \end{bmatrix}$$

III. HAMILTONIAN NEURAL NET AS A SPECTRUM ANALYZER

Some basic properties of the HNN can be derived from (11). First, as mentioned above, the structure of such a net creates a nonlinear vector field: $W\Theta(x)=v(x)$ with a single equilibrium point for $x = 0$. Hence HNN determines a type of orthogonal transformation, namely:

$$W \cdot \Theta(x) + I_{in} = 0 \qquad (16)$$

where: I_{in} – input vector (input data or signal)

It is worth noting that (16) gives the steady state solution of the net under constant excitation. Hence, output of the net i.e.

$$\Theta(x) = WI_{in} \qquad (17)$$

where rows and columns of W constitute orthogonal Haar basis (see example 1), can be seen as a Haar spectrum of input vector. Since, however $\Theta(x)$ is output of nonlinear, dynamical Hamiltonian system, (17) is true only for such a bounded input that $|\Theta(x)| \leq 1$ (for sigmoidal (± 1) activation functions).

Since $W^2 = -1$ it is clear that, this Haar analysis sets up the following relationships:

a) $(\Theta, I_{in}) = 0$, where $(.,.)$ denotes scalar product in l_2

b) The components of vector $\Theta(x)$ are Haar coefficients. Thus, the HNN performs a decomposition of the input vector in the sum of orthogonal patterns (columns or rows of weight matrix W). If the input vector consists of discrete samples of a time function, then these patterns can be treated as Haar-like wavelets.

c) If for a given input data set there are large Haar coefficients, then spectrum analysis fixes a number of principal components of the input data.

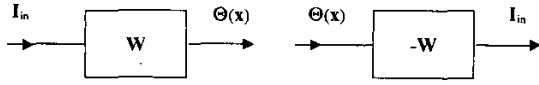The Haar analysis using HNN can be schematically shown as in Fig 3.

Fig. 3. Haar analysis using HNN

Note: Haar analysis illustrated in Fig. 3 means that one assumes a physical neural net (a signal processor) with orthogonal, skew-symmetric weight matrix **W** solving the following, ill-conditioned, differential equations:

$$\dot{x} = W\Theta(x) + I_{in}$$

Such a signal processor cannot be practically realized (in silicon) due to the pure imaginary eigenvalues of matrix **W**. On the contrary, the Haar transformation given by algebraic equation (16) is ready to use as an algorithm. In the following section we show how to solve the problem of physical realizability.

Example 2.
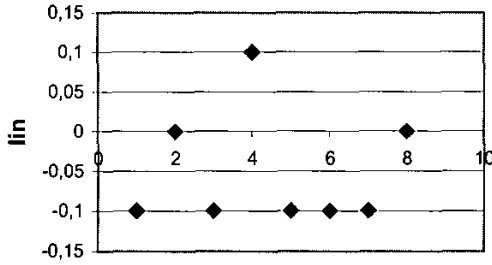
Haar analysis of an input signal I_{in}, using the HNN with weigth matrix W_4 from example 1, is given by the vectors: •

$$I_{in} = [-0.1, 0.0, -0.1, 0.1, -0.1, -0.1, -0.1, 0.0]^T$$

$$\Theta(x) = [-0.113, 0.151, -0.037, 0.037, 0.113, 0.037, 0.037, -0.075]^T$$

It means, that:

$$I_{in} = \Theta_1 w_1 + \Theta_2 w_2 + ... + \Theta_8 w_8$$

where: w_i, i=1,...,8 denotes Haar-basis i.e. rows of W_4

Note: Haar spectrum $\Theta(x)$ in this example is the output of the dynamical system and does not depend on the concrete shape of the activation functions (if $|\Theta_i(x)| < 1$).

IV. HNN AS AN ORTHOGONAL FILTER

Orthogonal filtering is one of the basic operations in modern signal processing. For example, in digital communication transmitted messages are encoded in the form of orthogonal symbols. Thus the transmitted signal consists of the symbols corrupted by additive noise. Hence, the main function of any receiver is to perform such an orthogonal filtering. A basic structure of an orthogonal filter using the structure of the HNN is shown in Fig.4.
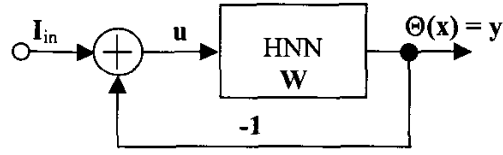

Fig. 4. Structure of an orthogonal filter.

It can be seen that such a filter performs the following decomposition:

$$I_{in} = u + y; \quad y = \Theta(x) \tag{18}$$

where: **u** and **y** are orthogonal i.e. $y = Wu$ and $(u, y) = 0$. At the same time (18) sets up two types of orthogonal transformations:

$$y = 0.5 (1+W) I_{in}; (I_{in}, y) \neq 0 \tag{19}$$

and

$$u = 0.5 (1-W) I_{in} \tag{20}$$

According to (19) the input signal I_{in} is decomposed into Haar or Walsh basis (Walsh functions take only the values 1 and −1) and the output signal **y** constitutes the Haar or Walsh spectrum, respectively. Assuming that the above symbols are encoded by columns or rows of weight matrix $(1 + W)$, the largest Haar coefficient of the spectrum at the output of the orthogonal filter can be used as a measure of presence of information hidden in noise. Indeed, one obtains

$$I_{in} = w_i + n = u + y \tag{21}$$

where: w_i - a symbol (a column or row of matrix $(1 + W)$),
     $i \in [1, ...,N]$,
     n - additive noise, $N = \dim W$

Hence, there is such a minimal S/N ratio, that

$$\Theta_i = (w_i + n, w_i) > \Theta_k = (w_i + n, w_k), \forall k \neq i \tag{22}$$

where: Θ_i − the largest Haar coefficient of the output spectrum, then the orthogonal filtering is performed.

Example 3.
Let a useful signal encoded by the second row w_2 of matrix $1+W_4$ (example1) i.e.

$$w_2 = [-0.377, 1.000, -0.377, 0.377, -0.377, 0.377, -0.377, 0.377]^T$$

is noised by vector **n**

i.e. $I_{in} = w_2 + n$. The Haar-analysis for different S/N ratio is as follows:
for $n = [0.268, 0.217, 0.460, -0.255, -0.291, -0.466, 0.325, 0.026]^T$
i.e S/N= 1.98 [dB]
$\Theta(x) = [0.137, 0.833, 0.292, 0.077, -0.095, -0.458, 0.152 - 0.148]^T$
$\Theta_2 = 0.833 > \Theta_i$; i ≠2 - vector w_2 is recognized.

for $n = [-0.68, -0.528, -0.452, -0.383, 0.432, 0.637, 0.381, -0.428]^T$

i.e S/N= -0.02 [dB]

$\Theta(x)=[-0.40,0.763,0.067,-0.259,-0.024,0.795, 0.257, 0.131]^T$

$\Theta_2 \sim \Theta_6$ - vectors w_2 or w_6 are recognized.

V. IMPLEMENTABILITY OF HNN

As mentioned above the HNN described by (11) cannot be realized as physical object. On the contrary, the orthogonal filter shown in Fig. 4 could be implemented in silicon, even if the weight matrix **W** is not exactly skew-symmetric. It is clear, that this implementation is guaranteed by the stabilizing action of negative feedback loops. Moreover, by cascading two such orthogonal filters one obtains a Haar spectrum analyzer equivalent to that shown in Fig. 3. In Fig. 5 such a cascade is schematically presented.
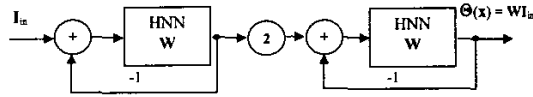

Fig. 5. Haar spectrum analyzer by using two orthogonal filters.

Thus, without going into the details of technological issues, one can state that very large scale artificial neural networks can be realized as physical objects by the structure of orthogonal filters.

VI. HNN AS A UNIVERSAL SIGNAL PROCESSOR

It is not the main goal of this paper to present signal processing by using the HNN. Nevertheless, it seems to be clear that having such nets implemented in the form of orthogonal filters, one could realize in real-time most of the signal processing known from advanced wavelet analysis [3], for example:

- signal denoising
- coherent structure extraction
- pattern recognition and classification
- multiscale and multiresolution signal analysis
- image processing
- data compression

It is worth noting, that known in mathematics problem of orthogonal Hurwitz-Radon matrices and solution given by Hurwitz-Radon theorem [4] can be used to formulate two essential issues in signal processing performed by HNN, namely:

- finding the best-adapted bases for given class of signals
- decomposition of given signal (pattern/image) into orthogonal components.

Indeed, let $W_1, ..., W_s$ be e set of orthogonal skew-symmetric Hurwitz-Radon matrices i.e.

$$W_j W_k + W_k W_j = 0 \text{ for } j \neq k; j, k = 1, ... , s$$

Let $\alpha_1, ..., \alpha_s$ be real numbers with $\Sigma \alpha_j^2 = 1$. Then:

$$W(\alpha) = \sum_{j=0}^{s} \alpha_j W_j \quad ; W_0 = \pm 1 \qquad (23)$$

is orthogonal, where $s_{max} = \rho(n) -1$; $\rho(n)$-Radon number of n. Hence the following adaptation rule:

Find such a vector of parameters α that the weight matrix $W(\alpha)$ of HNN sets up the best-adapted basis.

VII. CONCLUDING REMARKS

In this paper we have presented how to find the most suitable architecture for very large-scale artificial neural nets to be implemented in VLSI technology. The structure of such neural nets is composed of pairs of lossless neurons. Hence, they have the form of Hamiltonian Systems with weight matrices **W** particularly belonging to a set of Hurwitz-Radon matrices. The unique feature of HNN seems to be the fact that they can exist as algorithms or Hamiltonian physical devices performing the Haar-Walsh analysis in real-time. Parity functions, known from the literature [5,6] can be an alternative tool for Walsh analysis. Finally, it is worth noting that the use of different type neural networks in signal processing is becoming increasingly widespread [7,8].

REFERENCES

[1] G. Cauwenberghs, M. A. Bayoumi M, Editors; *Learning on Silicon*, Kluwer A. P. USA , 1999.

[2] C. Diorio, D. Hsu, M. Figueroa, "Adaptive CMOS: From Biological Inspiration to Systems-on-a-Chip", *Proc. IEEE*, vol. 90, no 3, 2002.

[3] I. Daubechies, Ed., *Different Perspectives on Wavelets*, AMS, vol. 47, 5 (1993).

[4] B. Eckmann,, "Topology, Algebra, Analysis- Relations and Missing Links", *Notices AMS*, 46, 5, 1999.

[5] V. Roychowdhury, Kai-Yeung Siu, A. Orlitsky, "Neural Models and Spectral Methods" in *Theoretical Advances in Neural Computing and Learning*, Kluwer Academic Publishers, Boston, London, Dordrecht , 1994, p. 3.

[6] Y. Mansour, "Learning Boolean Functions via the Fourier Transform" in *Theoretical Advances in Neural Computing and Learning*, Kluwer Academic Publishers, Boston, London, Dordrecht , 1994, p. 391.

[7] Fa-Long Luo, R. Unbehauen, *Applied Neural Networks for Signal Processing*, Cambridge, University Press 1998.

[8] W. Sienko, W. Citko, *On Very Large Scale Hamiltonian Neural Nets*, Proc. of 6-th International Conference on Neural Networks and Soft Computing, Zakopane, Poland, June, 2002.