

Fast Complex Adaptive Spline Neural Networks for Digital Signal Processing

Aurelio Uncini, Fulvio Capparelli and Francesco Piazza

Dipartimento di Elettronica e Automatica - Università di Ancona Italy

Via Breccie Bianche, 60131 Ancona-Italy

Fax: +39 (71) 2204464 - email: aurel@eealab.unian.it - Internet: http://nnsp.eealab.unian.it/

Abstract

In this paper, we study the complex-domain artificial neural networks called adaptive spline neural networks (ASNN), defined in the complex domain, which are able to adapt their activation functions by varying the control points of a Catmull-Rom cubic spline. This kind of neural network can be implemented as a very simple structure able to improve the generalization capabilities using few training epochs. Due to its low architectural complexity this network can be used to cope with several nonlinear DSP problems at high sampling rate. In particular, we investigate the application of this new neural network model to the adaptive channel equalization problem. The goal is to design a receiver which compensates the HPA nonlinearities in digital radio links and performs the symbols extraction from the received data (demodulation process), when a 16-QAM is used.

1. Introduction

RECENTLY¹, a new interest in adaptive activation functions has arisen in the neural network community. In fact, such a strategy seems to provide better fitting properties with respect to classical architectures with sigmoidal neurons.

The simplest solution we can imagine consists in involving gain a and slope b of the sigmoid $a(1 - e^{-bx})/(1 + e^{-bx})$ in the learning process [1]. A different approach is based on the use of polynomial functions [2], which allows to reduce the size of the network and, in particular, the connection

complexity; in fact, the digital implementation of the activation function through a LUT (look-up-table) keeps the overall complexity under control. Drawbacks with this solution arise with the non-boundedness of the function (non-squashing) and with the adaptation of the coefficients in the learning phase. In [3] the direct adaptation of the LUT coefficients is proposed. This approach is characterized by a difficult learning process due to the large number of free parameters and the lack of smoothness of the neuron's output. These are also the main reasons for the introduction of Hermite polynomials as substitutes for the so called supersmoother in the Projection Pursuit Learning approach of Hwang *et al.* [4].

Recently a new kind of artificial neural networks called adaptive spline neural networks (ASNN) has been introduced [5-7]. The adaptive activation function makes use of Catmull-Rom spline curve whose shape can be modified through some control points [8].

In particular in [7] this new architecture is presented in the context of the regularization theory: the authors demonstrate that ASNN embodies some regularity characteristics in its own activation function much better than the classical sigmoid can do.

The main advantages of this innovative structure, very useful for nonlinear adaptive signal processing, are: 1) the training sequence may be shorter than that required by the classical MLP; 2) the architecture is general and, unlike others approaches, it does not require a specific design; 3) the low hardware complexity (low overhead with respect to a simple adaptive linear combiner) makes it suitable for high speed data transmissions using a DSP device.

¹ This work is supported by the «Ministero dell'Università e della Ricerca Scientifica e Tecnologica» of Italy

In particular, in this paper, we investigate the complex domain extension of the ASSN (CASSN) and the application of this new neural network model to the adaptive nonlinear channel equalization problem.

2. Complex Domain Spline-based Activation Functions (CASSN)

2.1. Catmull-Rom implementation of the cubic activation function

Our idea consists in realizing a neuron with a more complex activation function than the sigmoid, able to reproduce the shape of a whole cubic spline along the directions specified by weights \mathbf{w}_j , $j=1, \dots, n$.

$$\varphi_j(\mathbf{w}_j \mathbf{x}) = \sum_{i=0}^N c_i |\mathbf{w}_j \mathbf{x} - \alpha_{ij}|^3 \quad j=1, \dots, n. \quad (1)$$

Referring to the papers [5-7] for further details, here we give some notes on such realization of an adaptive activation function. The goal is to give a global approximation using a structure as tractable as possible. In equation (1) we find a spline with $N+1$ tracts: each of them is described by a different combination of the coefficients c_i , because of the change in the sign of the kernels at α_{ij} . We have chosen to represent the activation functions through the concatenation of even more local spline basis functions, controlled by only four coefficients. As we want to keep the cubic characteristic, we have used a Catmull-Rom cubic spline. Using this type of spline we could exactly reproduce function (1), but, of course, this is not the cheapest solution, so we'll take a different approach.

The activation function will be composed of simpler cubic structures of local nature. The i -th tract is expressed by equation (2) where $u \in [0, 1]$ and $\mathbf{Q} = (q_x, q_y)$ (q_x and q_y are defined as control points). Such a spline interpolates the points \mathbf{Q}_{i+1} ($u=0$) and \mathbf{Q}_{i+2} ($u=1$) and has a continuous first derivative; the second derivative is not continuous only at the knots. Equation (2) represents a function if x -coordinates are ordered according to the rule $q_{x,i} < q_{x,i+1} < q_{x,i+2} < q_{x,i+3}$. The third degree equation $F_{x,i}(u) = x_0$, where x_0 is the activation of the neuron, gives the value of the local coordinate u : if the control points are uniformly

spaced along the x -axis (by a fixed step Δx), $F_{x,i}(u)$ reduces to the first degree polynomial (3).

$$F_i(u) = \frac{1}{2} \begin{bmatrix} u^3 & u^2 & u & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{Q}_i \\ \mathbf{Q}_{i+1} \\ \mathbf{Q}_{i+2} \\ \mathbf{Q}_{i+3} \end{bmatrix} \quad (2)$$

$$F_{x,i}(u) = u\Delta x + q_{x,i+1} \quad (3)$$

It is worth-noting that fixing the values of q_x reduces the total number of free parameters and the risk of overfitting.

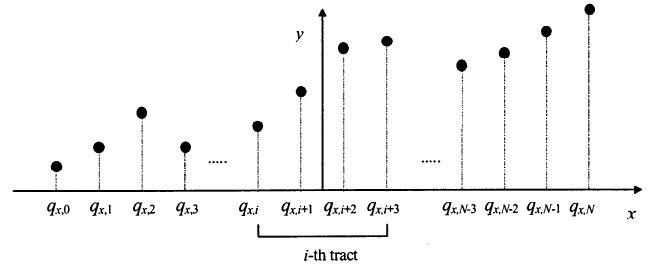


Figure 1. Control points of the Catmull-Rom spline based activation function: along the x -axis there is a fixed step Δx . The i -th tract starts with $q_{x,i}$ and ends with $q_{x,i+3}$, but the controlled interval is only between $q_{x,i+1}$ and $q_{x,i+2}$.

The q_y coordinates are initialized by sampling a function like the sigmoid, which supplies the neural network with universal approximation capability. Along the x -axis, $N+1$ points are taken; outside the sampling interval the neuron's output will be held constant at the values $q_{y,1}$, for the negative x -coordinate, and $q_{y,N-1}$ for the positive x .

In equation (3) the fixed step Δx plays a rule of regularizing parameter (see [7] for more details).

2.2 The Complex Valued Spline Activation Function

The advantage of using complex-valued NNs instead of a real-valued NN counterpart fed with a pair of real values is well known [9-12]. In complex-valued neural networks one of the main problems to deal with, is the complex domain activation function, whose most suitable features have been suggested in [11]. Let $F(S)$ be the complex activation function with $S \in \mathbb{C}$ defined as the complex linear combiner output;

the main constraints that $F(S)$ should satisfy are: 1) $F(S)$ should be non linear and bounded; 2) in order to derive the backpropagation algorithm the partial derivatives of $F(S)$ should exist and be bounded; 3) because of the Liouville's theorem $F(S)$ should not be an analytic function.

According to the previous properties, one possible choice, proposed in [12], [13], consists on the superposition of real and imaginary activation functions $F(S)=f_{Re}(Re[S])+jf_{Im}(Im[S])$; where the functions $f_{Re}(\bullet)$ and $f_{Im}(\bullet)$, can be simple real-valued sigmoids or more sophisticated adaptive functions.

Using a formalism similar to the one introduced in Widrow and Lehr in [14], and following a development similar to [9], [12-13], for the synaptic weights, the learning algorithm is now extended to the spline's control points.

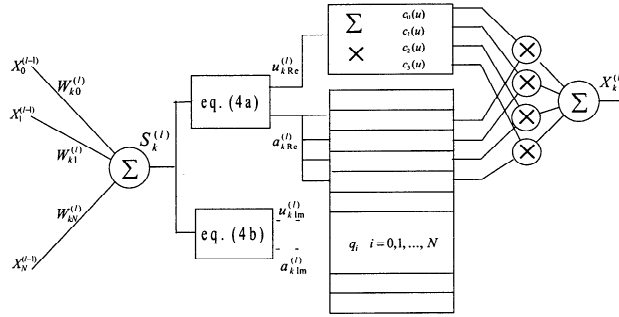


Figure 2. The complex-valued generalized sigmoid (GS) neuron structure (the structure of the imaginary part is omitted because it is identical to that of the real part).

Considering M total layers and indicating each of them with the index l , $l=1,...,M$, we can find the span a_k and the local variable u_k by

$$\begin{aligned} z_{k\text{Re}}^{(l)} &= \frac{\text{Re}[S_k^{(l)}]}{\Delta x} + \frac{N-2}{2} \\ a_{k\text{Re}}^{(l)} &= \lfloor z_{k\text{Re}}^{(l)} \rfloor \\ u_{k\text{Re}}^{(l)} &= z_{k\text{Re}}^{(l)} - a_{k\text{Re}}^{(l)} \end{aligned} \quad (4a)$$

$$\begin{aligned} z_{k\text{Im}}^{(l)} &= \frac{\text{Im}[S_k^{(l)}]}{\Delta x} + \frac{N-2}{2} \\ a_{k\text{Im}}^{(l)} &= \lfloor z_{k\text{Im}}^{(l)} \rfloor \\ u_{k\text{Im}}^{(l)} &= z_{k\text{Im}}^{(l)} - a_{k\text{Im}}^{(l)} \end{aligned} \quad (4b)$$

where the symbol $\lfloor \cdot \rfloor$ is the floor operator. We find for each neuron j the local tract a_j which s_j belongs to, and the local coordinate u_j . These expressions lead to neuron structure reported in Figure 2.

2.3 The CASNN Learning Algorithm

Following a development similar to [12], [13], for the synaptic weights, the learning algorithm is now extended to the spline's control points.

$$\begin{aligned} E_k^{(l)} &= \begin{cases} D_k - X_k^{(l)} & l = M \\ \sum_{m=1}^{N_l} \Lambda_m^{(l+1)} W_{mk}^{(l+1)} & l = M-1, \dots, 1 \end{cases} \\ \Delta_k^{(l)} &= \text{Re}[E_k^{(l)}] \left(\frac{d \text{Re}[F_{k,a_k^{(l)}}(u)]}{du} \right)_{u=u_{k\text{Re}}^{(l)}} \frac{1}{\Delta x} + j \text{Im}[E_k^{(l)}] \left(\frac{d \text{Im}[F_{k,a_k^{(l)}}(u)]}{du} \right)_{u=u_{k\text{Im}}^{(l)}} \frac{1}{\Delta x} \\ W_{kj}^{(l)}[p+1] &= W_{kj}^{(l)}[p] - 2\mu_w \Delta_k^{(l)} X_j^{(l-1)} \end{aligned} \quad (5)$$

with $0 \leq k \leq N_l$ and $0 \leq j \leq N_{l-1}$. The adaptation of the control points is ruled by

$$\begin{aligned} q_{k,(a_k^{(l)}+m)\text{Re}}^{(l)}[p+1] &= q_{k,(a_k^{(l)}+m)\text{Re}}^{(l)}[p] - 2\mu_q \text{Re}[E_k^{(l)}] \left(\frac{\partial \text{Re}[F_{k,a_k^{(l)}}(u)]}{\partial q_{k,(a_k^{(l)}+m)\text{Re}}^{(l)}} \right) = \\ &= q_{k,(a_k^{(l)}+m)\text{Re}}^{(l)}[p] - 2\mu_q e_k^{(l)} c_{k,m\text{Re}}^{(l)}(u_k^{(l)}); \\ q_{k,(a_k^{(l)}+m)\text{Im}}^{(l)}[p+1] &= q_{k,(a_k^{(l)}+m)\text{Im}}^{(l)}[p] - 2\mu_q \text{Im}[E_k^{(l)}] \left(\frac{\partial \text{Im}[F_{k,a_k^{(l)}}(u)]}{\partial q_{k,(a_k^{(l)}+m)\text{Im}}^{(l)}} \right) = \\ &= q_{k,(a_k^{(l)}+m)\text{Im}}^{(l)}[p] - 2\mu_q e_k^{(l)} c_{k,m\text{Im}}^{(l)}(u_k^{(l)}); \end{aligned} \quad (6)$$

where the patch index $m=0, \dots, 3$. The adaptation rates are μ_w for the connection weights and biases and μ_q for the control points. The control point with index 0, 1, $(N-1)$ and N are fixed.

The spline curve control points are initialized by sampling a bipolar sigmoid (or other well-known activation functions); outside the sampling interval, the asymptotic shape of the sigmoid is maintained.

3. Digital Radio Links Nonlinear Equalization

In microwave digital radio links, since on-board satellite power is a precious resource, to have an high efficiency, the transmitter high-power amplifier (HPA) operates near the saturation point:

nonlinearities are introduced that can cause serious performance degradation of the received signal [15].

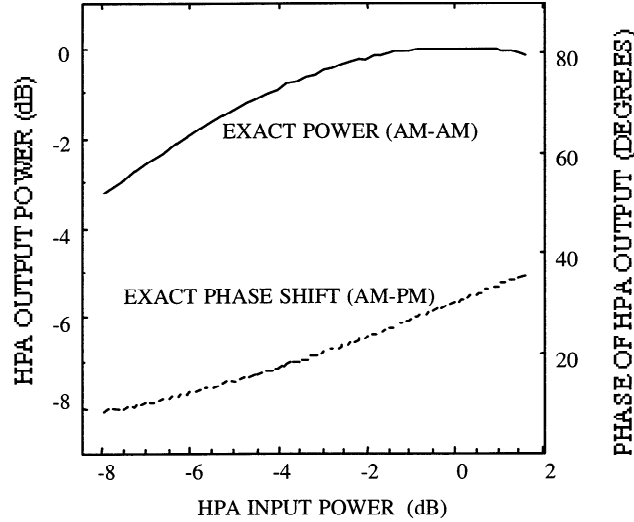


Figure 3. Typical traveling-wave tube (TWT) HPA memoryless input-output characteristic.

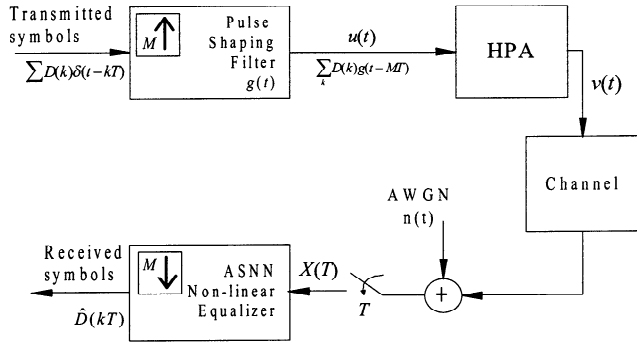


Figure 4. Block diagram of the digital radio link using a complex ASNN as nonlinear equalizer.

The nonlinearity of a typical HPA, a traveling-wave tube (TWT) or a GaAs FET amplifier, affects both amplitude (AM/AM conversion) and phase (AM/PM conversion) of the amplified signal, and can be considered as memoryless, i.e. the HPA is a nonlinear system without memory under a wide range of operational conditions [16]. In Figure 3 the AM-AM and AM-PM characteristics of a typical HPA is reported. However, in practice, the transmitter contains a pulse shaping circuit (modulator) at the baseband or at the intermediate frequency (IF) stage virtually in all digital radio systems (see Figure 4). Therefore, the overall

baseband-equivalent system (the cascade of transmitter, HPA nonlinearity and receiver) is a nonlinear system with memory.

It is well known that the 16 and 64 Quadrature Amplitude Modulation (QAM) links are very sensitive to nonlinear distortion, although they can exhibit better bit-error rate (BER) performance than equivalent phase-shift keying (PSK) systems on additive white noise Gaussian channels. The effects of the nonlinear channel with memory on the QAM signals are manifold, but three of them have particular importance:

Spectral spreading. The spectrum of the amplified signal after the HPA is much wider than the one of the signal before the HPA ;

Intersymbol interference (ISI). Since the overall system has memory, each symbol of the QAM alphabet, usually referred to as a constellation point, becomes at the receiver a cluster of points due to the interference among symbols at the sampling instants.

Constellation warping. The respective centers of gravity of the clusters caused by the ISI are no longer on a rectangular grid as in the original constellation.

The block diagram of the radio link used in our experiments is depicted in Figure 3. The complex input data sequence $D(k)$, represents the points of the QAM constellation.

The function $g(t)$ represents the modulator filter impulse response: in the simulation we use a square-root of a raised-cosine having a roll-off factor α equal to 0.5, and the over-sampling factor M is chosen equal to 3. The $g(t)$ filter length is equal to $5M$ taps. The model for the HPA is described in [16], the input-output HPA (memoryless) response is described by the AM-AM response, represented by the module, and the AM-PM response represented by the phase. This description assumes, for convenience, that the maximum possible HPA input power $W_{in} = |u(t)|^2$ is equal to 1W, and the maximum shift is $\Phi_0 = \frac{\pi}{6}$, which are typical values [15].

In the simulations the transmission of QAM signals are considered and, for the sake of brevity, only the case of 16 QAM has been reported. The maximum input power to the HPA W_{inMAX} is very close to the saturation point, and fixed equal to -1 dB (see Figure 3).

Three complex-valued equalizers have been tested:

1. complex-valued linear combiner (A15_1);
2. complex-value standard multilayer neural network with one hidden layer composed by 10 sigmoid neurons and linear output (N15_10_1);
3. complex-valued ASNN composed by only one complex GS neuron (S15_1).

The training set consists of 6144 ($2048 \times M$) input samples, $v(t) + n(t)$ in the scheme of Figure 3, corresponding to 2048 target QAM symbols. Since the neural network output are the QAM complex constellation points, the network performs also the down-sampling conversion. For each epoch, a different realization of white zero-mean Gaussian noise $n(t)$ is added, to obtain a S/N equal to 20 dB.

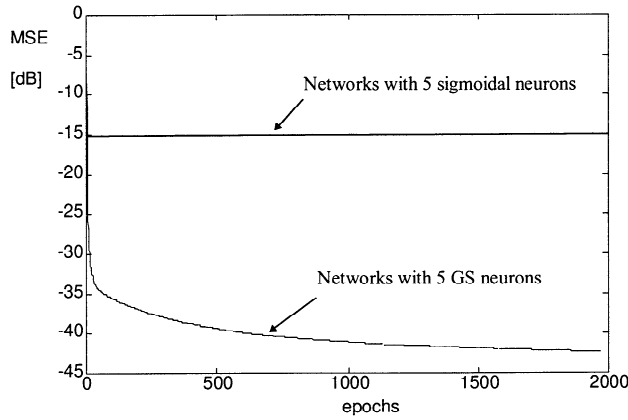


Figure 5. Training error for the Back & Tsoi system: comparison of architectures with the same number of hidden neurons.

During the learning phase, the S15_1 and A15_1 networks are trained for 100 epochs, while, in order to reach a suitable convergence, the standard MLP N15_10_1 is trained for 10^4 epochs. Both adaptation rates μ_q and μ_w for the S15_1 are chosen to be equal to 0.001; the same value is used for the MLP N15_10_1, while for the adaptive linear combiner A15_1 μ_w is equal to 0.0001.

Several tests using different networks initialization weights have been carried out. Figure 5 shows the plot of the MSE expressed in dB vs. the training epochs for the three architectures. The N15_10_1 x-axis scale must be multiplied by a factor of 100 for a total of 10^4 epochs.

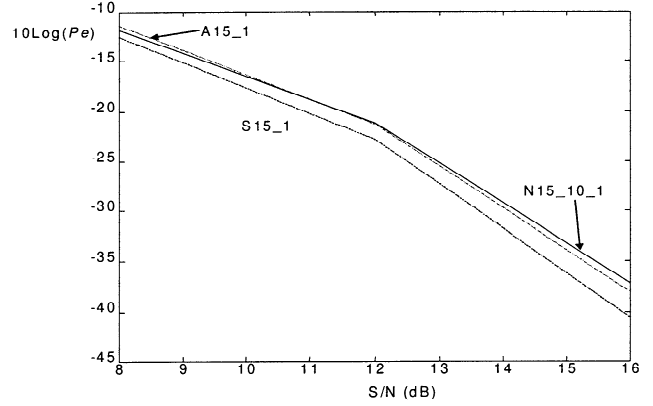


Figure 6. The symbols error probability (P_e) plotted vs. signal to noise ratio (S/N) at the input of the receiver using various equalization schemes.

In order to evaluate in a more realistic way the radio link performances, the symbol error probability (P_e) is computed. Figure 6 reports the P_e values vs. the S/N, both expressed in dB. From this figure we can observe that the proposed approach leads to significative improvements not only with respect to the classical linear adaptive filter, but also with respect to the already known sigmoid MLP based equalization technique.

4. Conclusions

In this paper, a new complex domain architecture is proposed, based on an adaptive spline activation function approach, characterized by a limited complexity. In fact, the main problem using standard neural networks as nonlinear adaptive filters is the high degree of overall complexity, which accounts for a very long adaptation time and a high number of interconnections, that can hinder any practical application. This fact allows to overcome the previous drawbacks making possible to use effectively neural structures in real world problems.

For what concerns the QAM equalization problem, this advantage is even more evident, since the network reduces to a single complex neuron. Moreover, the reduced complexity is responsible for the shorter adaptation phase in terms of training epochs, as experimentally observed. Comparing our technique with classical linear approaches, we can notice that there is a low implementation overhead with respect to the adaptive linear filter, but with a significant improvement in the performance, both in terms of MSE (not reported in this paper) and symbol

error probability. As far as nonlinear methods are concerned, the proposed architecture exhibits performance comparable to finite-order inverse Volterra filters and to global compensation but with a much lower complexity that makes it suitable for an effective implementation.

References

- [1] C. T. Chen, W. D. Chang, «A Feedforward Neural Network with Function Shape Autotuning», *Neural Networks*, Vol.9, No 4, pp. 627-641, June 1996.
- [2] F. Piazza, A. Uncini, M. Zenobi, «Artificial Neural Networks with Adaptive Polynomial Activation Function», in *Proceedings of IJCNN*, Beijing, Cina, pp. II-343-349, Nov. 1992.
- [3] F. Piazza, A. Uncini, M. Zenobi, «Neural Networks with Digital LUT Activation Function», in *Proceedings of IJCNN*, Nagoya, Japan, pp. II-1401-1404, 1993.
- [4] J-N Hwang, S-R Lay, M Maechler, R.D. Martin, J. Schimert, «Regression Modeling in Back-Propagation and Projection Pursuit Learning» *IEEE Transactions on Neural Networks*, 5(2), 342-353.
- [5] S. Guarnieri, F. Piazza, A. Uncini, «Multilayer Neural Networks with Adaptive Spline-based Activation Functions» In *Proceedings of the WCNN 95*, Washington D.C., USA, I695-I699, 1995.
- [6] L. Vecchi, P. Campolucci, F. Piazza, A. Uncini, «Approximation Capabilities of Adaptive Spline Neural Networks» In *Proceedings of ICNN'97* Huston TX 96, ,1997.
- [7] L. Vecchi, F. Piazza and A. Uncini, «Learning and Approximation Capabilities of Adaptive Spline Activation Function Neural Networks», accepted for publication in *Neural Networks*.
- [8] E. Catmull, R. Rom, «A Class of Local Interpolating Splines», in R. E. Barnhill, R. F. Riesenfeld (ed.), *Computer Aided Geometric Design*, Academic Press, NewYork, 1974, pp. 317-326.
- [9] N. Benvenuto, M. Marchesi, F. Piazza and A. Uncini «A Comparison between Real and Complex valued Neural Networks in Communication Applications», *Proc. of Intern. Conference on Neural Networks*, ICANN91, Espoo, Finland, June 1991.
- [10] H. Leung, S. Haykin, «The Complex Backpropagation Algorithm», *IEEE Trans Acoust. Speech and Signal Process*, Vol.ASSP-39, pp.2101-2104, Sept. 1991.
- [11] S. Haykin, «Adaptive Filter Theory» Third Edition, Prentice Hall ed., 1996.
- [12] N. Benvenuto, M. Marchesi, F. Piazza, A. Uncini, «Non Linear Satellite Radio Links Equalized using Blind Neural Networks», in *Proceedings of IEEE International Conference on Acoustic Speech and Signal Processing - ICASSP91*, May 14-17, Toronto, Canada, 1991.
- [13] N. Benvenuto and F. Piazza «On the Complex Backpropagation Algorithm», *IEEE Trans Signal Processing*, Vol.40, pp.967-969, Apr. 1992.
- [14] B. Widrow, M. Lehr, «30 Years of Adaptive Neural Networks: Perceptron, Adaline and Backpropagation», in *Proceedings of the IEEE*, Vol.78, No.9, Sept. 1990.
- [15] S. Pupolin, L. J. Greenstein, «Performance Analysis of Digital Radio Links with Nonlinear Transmit Amplifiers», *IEEE Journ. on Selected Areas in Communications*, Vol. SAC-5, No 3, pp 534-456, April 1987.
- [16] A.A.M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers", *IEEE Trans. on Communications.*, vol. COM-29, pp. 1715-1720, Nov. 1981.