Pre-Distortion for the compensation of HPA nonlinearity with neural networks: Application to satellite communications

Rafik Zayani† and Ridha Bouallegue††

† National Engineering School of Tunis, Tunis ElManar University, Tunisia
†† National Engineering School of Sousse, Sousse University, Tunisia
6'Tel Laboratory SUP'COM

Summary
Neural networks (NNs) are able to give solutions to complex problems in digital communications due to their nonlinear processing, parallel distributed architecture, self-organization, capacity of learning and generalization, and efficient hardware implementation.

The pre-distortion being at the center of interest of this paper is one of the possible methods to compensate for HPA nonlinearities. The principle of pre-distortion is to distort the HPA input signal by an additional device called a pre-distorter whose characteristics are the inverse of those of the amplifier.

In this paper, we propose a pre-distortion scheme based on a feed-forward neural network. Efficient High Power Amplifiers (HPA) present non-linearities generating amplitude and phase distortions on the HPA output signal; the proposed pre-distortion technique will reduce theses distortions.

The performance of the proposed scheme is examined through computer simulations for 16-QAM OFDM signals. It is confirmed that the proposed pre-distorter with neural network consisting with one hidden layer and nine neurons gives a good performance improvement of quality of the transmission. Specifically, improvements in the reduction of the bit error rate (BER) are demonstrated for the travelling wave tube (TWT) HPA model.

Key words
OFDM, Pre-distorter, HPA, TWTA, SSPA, Neural Networks, Mobile Satellite Communications

1. Introduction

The satellite communications field is getting an enormous attention in the wake of third generation (3-G) and future fourth generation (4-G) mobile communication systems challenges [1, 2]. Currently, when the telecommunications industries are planning to deploy the 3-G system worldwide and researchers are coming up with tons of new ideas for the next-generation wireless systems, a load of challenges are yet to be fulfilled. These include high data rate transmissions, multimedia communications, seamless global roaming, quality of service (QoS) management, high user capacity, integration and compatibility between 4-G components, and so forth. To meet these challenges, presently researchers are focusing their attention in the satellite domain by considering it an integrated part of the so-called information superhighway [2, 3, 4, 5]. As a result, a new generation of satellite communication systems is being developed to support multimedia and Internet-based applications.

These satellite systems are developed to provide connectivity between remote terrestrial networks, direct network access, Internet services using fixed or mobile terminals, and high data rate transmissions [1, 6]. In all these research and development scenarios, non-geostationary satellite networks are considered to provide satellite-based mobile multimedia services for their low propagation delay and low path loss [1,2, 5, 7, 8].

Among the most important challenges of satellite mobile communications are spectral and power efficiencies. Several researchers are working to make use of spectrally efficient modulation schemes, such as M-QAM modulations, for satellite transmissions. Power efficiency represents the ability of a system to reliably transmit information at a lowest practical power level. To reach high power efficiency, satellite communication systems are equipped with high power amplifiers (HPAs), which, unfortunately, cause nonlinear distortions to the transmitted signal. The distortions are particularly significant when

Manuscript received March 5, 2007
Manuscript revised March 25, 2007

multilevel modulation schemes are employed, such as M-QAM (M > 4) modulations [4, 6, 10]. Because of this nonlinear problem, early satellite systems have been restricted to simple (and, therefore, spectrally inefficient) modulation schemes, such as binary phase shift keying (BPSK) modulation, which are less sensitive to the nonlinear problem than spectrally efficient modulation schemes [6].

Given the above facts, this paper proposes a non-linear distortion compensation technique for 16-QAM OFDM signals. The Pre-distorter is based on a feed-forward neural network (FNN); due to the universal approximation property of neural networks. Recently, the FNN has been applied to modeling non-linear memory less channels such as traveling wave tube amplifier (TWTA) and also allow efficient approximation of the inverse TWT transfer function.

The remainder of this paper is organized as follows. In section 2, we present a description of the proposed system. In section 3 and 4 we describe the non-linear model for high power amplifier and the pre-distortion scheme respectively. Section 6 presents and discusses the simulation results. The last section shows the conclusions.

2. System description

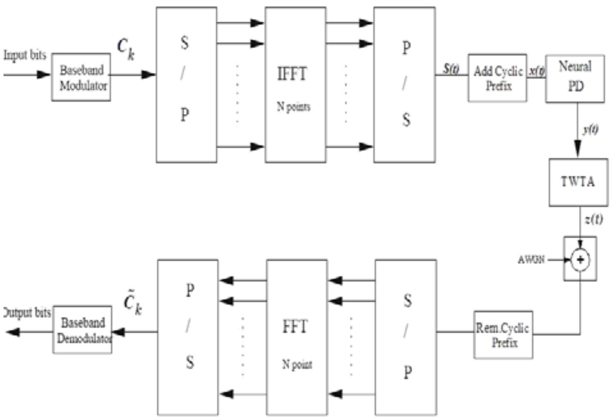

Fig. 1 Block diagram of the transmission system

The serial input bit stream consists of binary data $u_k \in \{0,1\}$ that are mapped to the symbol constellation, resulting in a complex stream c_k. This complex is applied in serial to parallel converted to produce the sequence applied to the inverse fast Fourier transform (IFFT) process, which is a fast implementation of an inverse discrete Fourier transform (IDFT). This signal is extended by a guard interval called cyclic prefix (CP). The nth transmit OFDM block is given by:

$$s_n(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k \phi_n(t - nT) \tag{1}$$

$$\phi_k(t) = \begin{cases} \exp(j \cdot 2\pi \cdot f_k \cdot t), & \forall t \in [-T_g, T] \\ 0 & \text{othersiwe} \end{cases} \tag{2}$$

Where $f_k = f_0 + \frac{k}{T}$.

The choice of the frequency spacing $\Delta f = \frac{1}{T}$ guarantees that the sub-carriers are orthogonal over the elementary symbol time T and the total OFDM block duration is $T_s = T + T_g$. Finally, samples of IFFT output, s_n, are Digital to analog converted and transmitted.

The modulated signal $x(t)$ is first pre-distorted and nonlinearly amplified and then propagated over an AWGN channel corrupted by additive white Gaussian noise.

3. Non linearity models

Power amplifiers are typically the most power-hungry components of RF transceivers. The design of PAs, especially for linear, low-voltage operations, remains a difficult problem defying an elegant solution. Two type's amplifiers are mostly used in communication: TWT and SSPA. TWT is mostly used for high power satellite transmitters while SSPA is used in many other applications including mobile transmitters because of its small size. Several previous papers used saleh's model to analyze the HPA [6,7].

The complex output of RF with non-linear distortion can be expressed as:

$$z(t) = f\left[u_y(t)\right] e^{j\left\{\alpha_y(t) + \phi\left(u_y(t)\right)\right\}} \tag{3}$$

where $u_y(t)$ and $\alpha_y(t)$ are the modulus and phase of the input signal. The measured AM/AM and AM/PM for TWT is well presented by saleh's model [3] as:

$$\begin{cases} f(u_y) = \alpha_A u_y / (1 + \beta_A |u_y|^2) \\ \phi(u_y) = \alpha_\phi |u_y|^2 / (1 + \beta_\phi |u_y|^2) \end{cases} \tag{4}$$

While SSPA's AM/AM and AM/PM can be captured by[8,9]:

$$\begin{cases} f\left[u_y\right] = u_y / [1 + (u_y / A_{\max})^{2p}]^{1/2p} \\ \phi\left[u_y\right] = 0 \end{cases} \qquad (5)$$

Here $A_{\max}$ is the maximum output amplitude and the parameter p controls the smoothness of the transition from the linear region to the limiting region [9].
Among the models available, here we will concentrate in describing the well-known Saleh Model for TWT memory-less HPAs, which is in fact the most commonly used in the literature and thence will be applied later in our Pre-distorter model.

In the expressions above, we choose to set the signal gain term to $\alpha_A = 2$, $\beta_A = 1$, $\alpha_\phi = 4$, $\beta_\phi = 9$. This represents a typical TWT model used in satellite communications [13]. so that the input saturation voltage $A_s = 1/\sqrt{\beta_A}$ and the maximum output amplitude

$$A_{\max} = \max\left\{ f\left[x(t)\right]\right\} = \alpha_A A_s / 2 \qquad (6)$$

The corresponding AM/AM and AM/PM curves so scaled are depicted in the following figure.

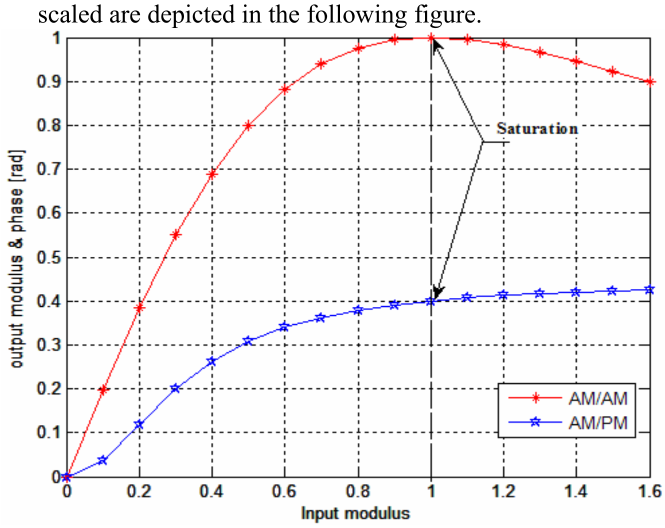

Fig. 2 The AM/AM and AM/PM characteristics of an HPA

The non-linear distortion of a TWT amplifier (TWTA) depends on the back-off. The input back-off (IBO) for the pre-distorted amplifier is defined as

$$IBO = 10\log_{10}\left(\frac{P_{sat,i}}{P_{avg,i}}\right) \qquad (7)$$

where $P_{sat,i}$ is the saturation input power and $P_{avg,i}$ is the average input power of the TWTA.

The pre-distortion technique connects a pre-distorting amplifier in front of the main amplifier. Compared to the compressive main amplifier, this additional amplifier has the opposite output distortion characteristic, i.e. its nonlinearity is expansive, not compressive. These two nonlinear distortions cancel each other when summed, resulting in a linear and distortion-free output from the main RF amplifier.

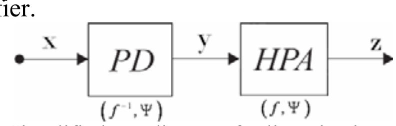

Fig. 3 Simplified pre-distorter for linearization of HPA

As in figure 3, the purpose of pre-distortion is to find another function (f^{-1},ψ), so that the overall effect of signal output will be linear as,

$$z(t) \approx Cx(t) \qquad (8)$$

The theoretical formulation of the pre-distortion is obtained by replacing the saturation input amplitude $A_s = 1/\sqrt{\beta_A}$ in the expression (4). This gives

$$f(u_y) = \frac{A_s^2 \alpha_A u_y}{A_s^2 + u_y^2} \qquad (9)$$

Whence we can find an AM/AM inverse transfer function $f^{-1}[.]$ by solving (9) for $u = f^{-1}(f(u))$. Some straightforward algebraic steps lead us to directly obtain:

$$f^{-1}(u) = \frac{\alpha_A A_s^2}{2u}\left[1 - \sqrt{1 - \left(\frac{2u}{\alpha_A A_s}\right)^2}\right] \qquad (10)$$

Where it is important to note that this inversion will be valid only within the interval $\{0 \le u \le \alpha_A A_s / 2\}$. This defines a restriction for the input range of the theoretical AM/AM PD. However the invertibility of the complex HPA function will not be necessarily restricted to the same limits since they account only for the AM/AM invertiblity.

The ideal AM/PM PD characteristic reated to the AM/AM theoretical inverse given in (10) is much simpler to obtain but not as trivial as taking (4) and inverting its sign.

Thus, letting $\psi[.]$ denote the AM/PM characteristic of the PD block, we have:

$$y = f^{-1}(u_x)e^{j\left(\alpha_x + \psi\left(u_x\right)\right)} \tag{11}$$

$$z = f\left(f^{-1}\left(u_x\right)\right)e^{j\left(\alpha_x + \psi(u_x) + \phi\left(f^{-1}\left(u_x\right)\right)\right)} \tag{12}$$

Wherein the AM/PM correction requires that:

$$\psi(u_x) = -\phi\left[f^{-1}\left[u_x\right]\right] \tag{13}$$

Is a general expression for the AM/PM pre-distortion function $\psi[.]$ that compensates for the amplitude to phase distortion $\phi[.]$ introduced by an HPA whose AM/AM nonlinear characteristic $f[.]$ has an exact inverse counterpart $f^{-1}[.]$. Thence, with these conditions fulfilled for the example of figure 3, it is true that:

$$z = f\left(f^{-1}\left(u_x\right)\right)e^{j\left(\alpha_x + \psi(u_x) + \phi\left(f^{-1}\left(u_x\right)\right)\right)} \approx x \tag{14}$$

The corresponding AM/AM and AM/PM transfer characteristics of pre-distorter, valid for the normalized Saleh's HPA model in the interval $\{0,1\}$, is shown in the following figure.

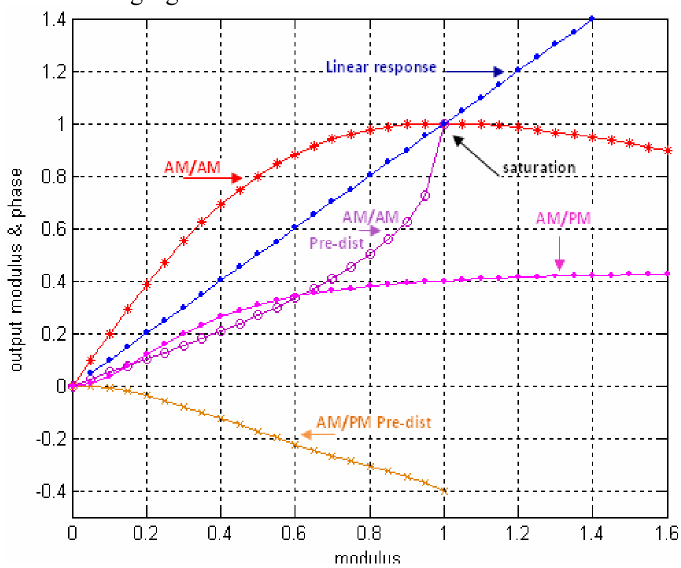

Fig. 4 AM/AM et AM/PM theoretical pre-distortion characteristics for the Saleh model

4. Neural Network pre-distortion scheme

The basic idea proposed is to identify the TWT inverse transfer function with a feed-forward neural network. Therefore, by using this structure, we aim at obtaining direct estimation of the amplitude and phase nonlinearities.

The following figure shows the detailed scheme of pre-distortion system.

5.1. *Training and generalization*

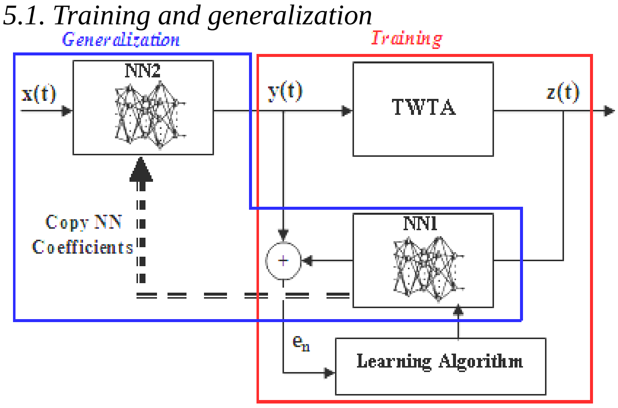

Fig. 5 Block diagram for training of the PD with TWTA

Training: where NN1 aims to identify the TWTA inverse transfer function, the error sent to "learning algorithm" bloc that reacts on coefficients of NN1.

Generalization: coefficients of the NN1 are recopied on the NN2 that achieves the pre-distortion.

5.2. *Neural networks structure*

The multi-layer [7] feed forward neural network (MLNN), called also multi-layer perceptron (MLP), is one of the most popular neural network architectures used in digital communications. Its basic unit, the neuron (Fig. 6), is composed of a linear combiner and an activation function. The neuron receives inputs from other processors. The linear combiner output is the weighted sum of the inputs plus a bias term. The activation function gives then the neuron output:

$$z = g(d)$$

where

$$d = \sum_{j=1}^{N} w_j z_j + b,$$

where z_j is the j^{th} input value of the neuron, w_j the corresponding synaptic weight, and b the bias term. $\{w_j\}$ and $\{b\}$ form the free parameters of the neuron.

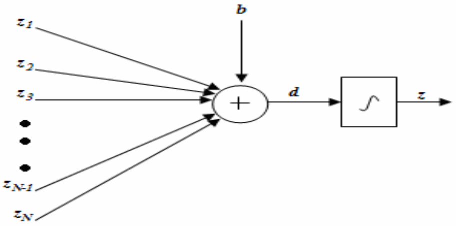

Fig. 6. The neuron

A multi-layer neural net (see Fig. 7) is composed of neurons connected to each other.

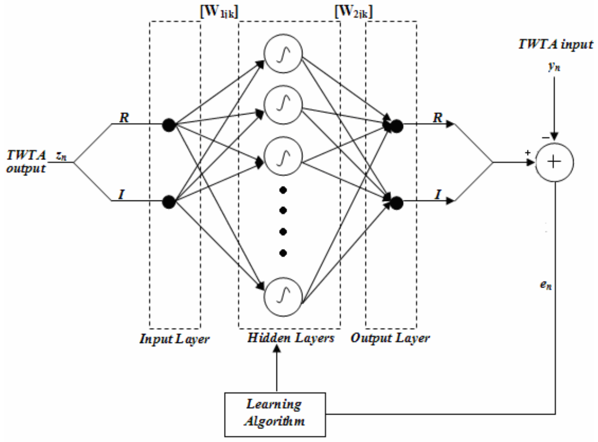

Fig. 7. A multi-layer neural network: The network has two layers, two input signals, one hidden Layers, 2 neurons in the output layer, and 2 output signals. (Indexes *R* and *I* refer to the real and imaginary parts, respectively)

The input information is processed from the input layer to the output layer. The network inputs are the inputs of the first layer. The outputs of the neurons in one layer form the inputs to the next layer. The network outputs are the outputs of the output layer. The layer index is denoted by i. z_{li} is the output of neuron i of layer l. w_{lji} is the weight that links the output z_{i-lj} to neuron i of layer l. *N(l)* is the number of neurons in layer l. With these notations, the output z_{li} of neuron *(l, i)* is given by:

$$z_{li} = g(d_{li}) \tag{15}$$

where

$$d_{li} = \sum_{j=1}^{N(l-1)} w_{lji} z_{l-1j} + b_{li} \tag{16}$$

5.3. Learning algorithm

The neural network is used to identify the TWTA inverse transfer function using supervised learning. At each iteration, a pair of TWTA input - TWTA output signals is presented to the neural network. We use a Levenberg Marquardt (LM) algorithm [10] to train the NN.

In the LM method, the change (Δ) in the weights (w) is obtained by solving

$$\alpha\Delta = -\frac{1}{2}\nabla E \tag{17}$$

where E is the mean-squared network error

$$E = \frac{1}{Na}\sum_{k=1}^{Na}\left[y(z_k) - y_k\right]^2 \tag{18}$$

N_a is the number of examples, $y(z_k)$ is the network output corresponding to the example z_k, and y_k is the desired output for that example.

The elements of the α matrix are given by

$$\alpha_{ij} = \left(1 + \lambda\delta_{ij}\right)\sum_{r=1}^{Ns}\sum_{k=1}^{N_a}\left[\frac{\partial y_r(z_k)}{\partial w_i}\frac{\partial y_r(z_k)}{\partial w_j}\right] \tag{19}$$

where Ns is the number of outputs of the network and δ_{ij} is the learning rate.

Starting from initial random weights, both α and ∇E are evaluated, and solving (17), a correction for the values of the weights is obtained $\left(w(n+1) = w(n) + \Delta\right)$. This is known as an LM learning cycle. Each iteration of this cycle reduces the error until the desired goal is achieved or a minimum is found. The λ variable in (19) is a parameter that is adjusted at each cycle, according to the error evolution. If it is very small the α matrix becomes an approximation to the Hessian, and the method is the inverse-Hessian method. If $\lambda \gg 1$, the method becomes analogous to steepest descent.

It can be easily seen from (17) that, if m is the number of weight, we have to calculate and store the m^2 elements of the α matrix at each iteration and find its inverse, which needs about m^3 operations to be performed. This fact makes the LM method very expensive both in memory and number of operations required when the network to be trained has a significant number of adaptive weights.

5.4. Constitution of learning base

To constitute a learning base to Neural Network, a simulation of an OFDM system with a TWT amplifier is required. For every OFDM symbol simulated, the pair of TWTA input – TWTA output symbols is stored in the base. TWTA output will be provided as an input of the neural network in training phase, TWTA input will be given like goal of NN. So the network will learn the inverse function of non-linearities.

Several parameters are important for the generation of the training basis. At first, features of the OFDM modulation: indeed parameters as the number of carriers, the binary coding to symbol and the model of the amplifier are able all to modify behavior of the no-linearity, and therefore the function that must achieve the network of neurons. In presentations of results, these parameters will be indicated systematically.

5. Discussion of Results

This architecture has been tested successfully on a 64-carrier OFDM system. To train the network, a learning base has been created using MATLAB. The OFDM system simulated uses 64 carriers, a 16-QAM modulation, and a channel with Additive White Gaussian Noise (AWGN). The Signal to Noise Ratio (SNR) used is Eb/N0=13dB. 20000 OFDM symbols are used as learning base, and 20000 others are used as validation set. Then the trained neural network is simulated in a complete OFDM system, using MATLAB. The Bit Error Rate (BER) is used to measure the system performance.

The OFDM followed by a pre-distorter was first trained the parameters of the pre-distorter were estimated using $N_a = 200$ samples of amplitude $A(.)$ 16QAM modulation.

The neural pre-distorter consists of two input and two output (R and I). Different architectures have been tested, with first of all a hidden layer of 2 neurons, then while increasing the number of neurons progressively, before testing a network with two hidden layers, also while increasing the number of neurons progressively on the two layers. Functions of hidden layer activation are tangents hyperbolic, and those of output layers are linear.

The Fig. 8 shows the performance of each pre-distorter on OFDM systems at OBO = 5.31dB. The OBO can be defined as

$$OBO[dB] = 10\log_{10}\frac{P_{\max}}{P_{avg}} \quad (17)$$

where $P_{\max}$ and P_{avg} represent the maximum output power and average signal power at the HPA output.

PD(2,x,2) represents a neural network with a hidden layer of x neurons, PD(2,x-y,2) neural network with two hidden layers of x and y neurons.

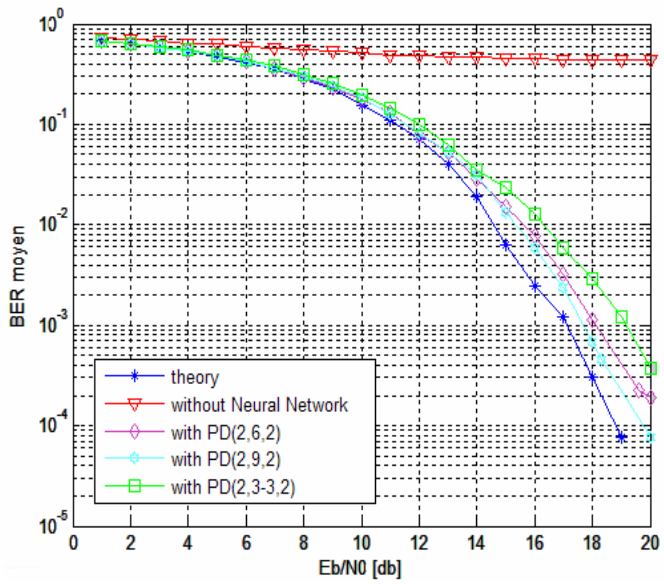

Fig. 8. Bit Error Rate of the OFDM system with pre-distorter vs. SNR: a QAM16 modulation is used on 64 carriers and OBO=5.31dB.

All neural pre-distorter arrive to reduce the BER in relation to the one without any pre-distorter. The one that gets the best performances is the PD(2,9,2). Fig. 9 shows the training curve versus iteration number for 16-QAM OFDM symbols and a TWT amplifier with an OBO=5.31dB. The feed-forward neural pre-distorter was configured as PD(2,9,2), the algorithm of training (here a Levenberg Marquardt), only 200 training iterations and the MSE was 1.32201e-007, resulting an accurate estimation of the coefficients for the neural pre-distorter.

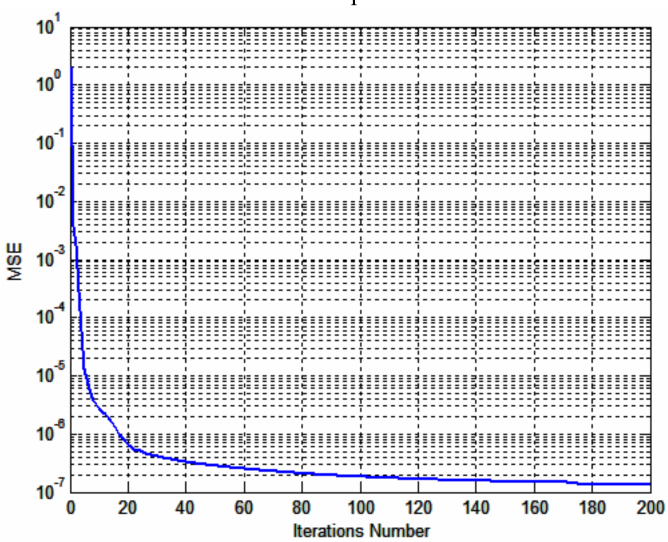

Fig. 9: MSE vs Iterations number for 16QAM OFDM-PD(2,9,2)

Fig. 10 show constellations for signals received either with or without a pre-distorter in the transmitter for an OFDM system on 64-carrires at SNR = 13dB.

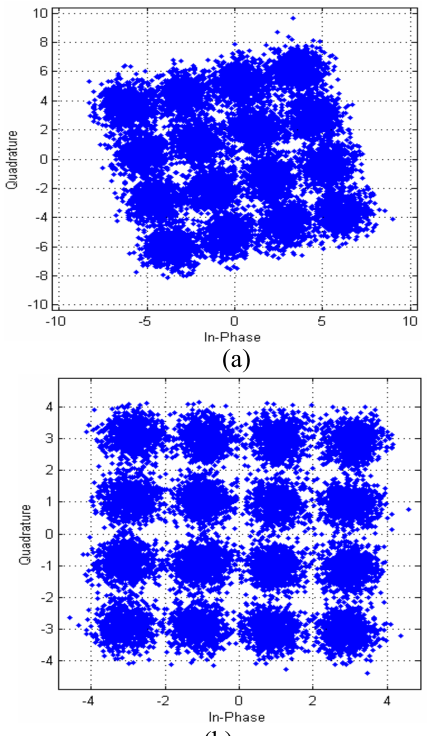

Fig. 10: Constellation of received signals with/without pre-distorter: 16 QAM OFDM, 64 carriers, SNR = 13dB. (a) HPA only, (b) with neural network Pre-distorter.

6 Conclusions

In this paper, we have proposed an adaptive pre-distortion technique based on feed forward neural networks. This structure was applied to 16 QAM OFDM transmission over non-linear TWT amplifier. The Levenberg Marquard algorithm has been used to update the neural network weights.

From simulation results, it is confirmed that the proposed pre-distorter with neural network consisting with one hidden layer and nine neurons gives a good performance improvement of quality of the transmission compared to 16QAM OFDM without a pre-distortion scheme.

Acknowledgment

The authors would like to express their cordial thanks to Pr. Daniel ROVIRAS for his valuable advice.

References

[1] J.A.C. Bingham, Multicarrier modulation for data transmission: an idea whose time has come, IEEE Comm. Mag., Vol. 28, Nº 5, pp. 5-14, May 1990.
[2] ETSI, Radio Broadcasting Systems: Digital Audio Broadcasting (DAB) to mobile, portable and fixed receiver, European Telecommunication Standard ETS 300-401, August 1997.
[3] Saleh M. and Adel A, Frecuency-Independent and Frecuency-Dependent Nolinear Models TWT Amplifiers, IEEE Trans. on Comm.., Vol. COM-29, pp. 1715-1719, Nov. 1981.
[4] M.G. DI Benedetto and P. Mandarini, An application of MMSE predistortion to OFDM systems, IEEE Trans. Comm., vol. 44, Nº 11, pp. 1417-1420, Nov. 1996.
[5] W.G. Jeon K. H. Chang, and Y.S. chao, An adaptive data predistorter for compensation of nonlinear distortion in OFDM systems, IEEE Trans. Comm., Vol. 45, Nº 10, pp. 1167-1171, Oct. 1997.
[6] Lei Ding, «Digital Predistortion of Power Amplifiers for Wireless Applications», a Thesis presented to the Academic Faculty, March 2005.
[7] Mohamed Ibnkahla, " Applications of neural networks to digital communications: a survey", 1997.
[8] A. N. D'Andrea, V. Lottici and R. Reggiannin, «Nonlinear Predistortion of OFDM Signals over Frequency-Selective Fading Channels», IEEE Transactions on Communications. Vol. 49. N° 5. pp. 837-843. 2001.
[9] R. zayani, S. Zid, R. Bouallegue, « Simulateur des non-linéarités HPA sur un système OFDM » OHD Conference, septembre 2005.
[10] G. Lera and M. Pinzolas, « Neighborhood Based Levenberg–Marquardt Algorithm for Neural Network Training», IEEE transaction on Neural Networks vol13 No5, 2002.

Biographical notes

Rafik ZAYANI was born in Gafsa, Tunisia. He received the Engineer degree in Telecommunications in 2003, the M.Sc. degree in Cellular Mobile Network in 2004, all from the National Engineer School of Tunis (ENIT), Tunisia. He is currently a research assistant in ISI, his current research interests include mobile and satellite communications, neural network, OFDM and UWB systems.

Ridha BOUALLEGUE was born in Tunis, Tunisia. He received the M.S degree in Telecommunications in 1990, the Ph.D. degree in telecommunications in 1994, and the Habilitation a Diriger des Recherches (HDR) degree in Telecommunications in 2003, all from the National Engineer School of Tunis (ENIT), Tunisia. He is currently Director of National Engineer school of Sousse and Research Laboratory 6'Tel / Sup'Com. His current research interests include mobile and satellite communications, Access technique, intelligent signal processing, CDMA, MIMO, OFDM and UWB system