Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2008, Article ID 132729, 15 pages
doi:10.1155/2008/132729

Research Article

Adaptive Predistortions Based on Neural Networks Associated with Levenberg-Marquardt Algorithm for Satellite Down Links

Rafik Zayani,[1] Ridha Bouallegue,[1] and Daniel Roviras[2]

[1] *Unité de Recherche 6'Tél/Sup'Com, route de Raoued Km 3.5, 2083 Ariana, Tunisia*
[2] *ENSEEIHT, Laboratoire IRIT 2, rue C. Camichel, BP 7122, 31071 Toulouse Cedex 7, France*

Correspondence should be addressed to Rafik Zayani, rafik.zayani@gmail.com

Received 1 March 2008; Accepted 28 June 2008

Recommended by George Tombras

This paper presents adaptive predistortion techniques based on a feed-forward neural network (NN) to linearize power amplifiers such as those used in satellite communications. Indeed, it presents the suitable NN structures which give the best performances for three satellite down links. The first link is a stationary memoryless travelling wave tube amplifier (TWTA), the second one is a nonstationary memoryless TWT amplifier while the third is an amplifier with memory modeled by a memoryless amplifier followed by a linear filter. Equally important, it puts forward the studies concerning the application of different NN training algorithms in order to determine the most prefermant for adaptive predistortions. This comparison examined through computer simulation for 64 carriers and 16-QAM OFDM system, with a Saleh's TWT amplifier, is based on some quality measure (mean square error), the required training time to reach a particular quality level, and computation complexity. The chosen adaptive predistortions (NN structures associated with an adaptive algorithm) have a low complexity, fast convergence, and best performance.

1. INTRODUCTION

Satellite communication systems are developed to provide connectivity between remote terrestrial networks, direct network access, and Internet services using fixed or mobile terminals [1]. Hence, future generations of these systems are required to support higher transmission data rates for providing multimedia services. Among the most important challenges of satellite communications are spectral and power efficiencies [2]. In order to increase bandwidth utilization, several researchers are working to make use of high-level linear modulation schemes such as M-QAM and its multicarrier variant OFDM which can support bit rates of $\log_2(M)$ bits/s in 1 Hz of bandwidth [3] and resists multipath fading and impulse noise [4]. Even if OFDM presents high power to average power ratio (PAPR), its utilization on satellite down links is interesting because of its spectral efficiency. Furthermore, digital video broadcasting (DVB) for terrestrial networks uses OFDM modulation. Using OFDM for video broadcasting on satellite down links has the advantage of having only one single terminal for receiving television from terrestrial or satellite broadcasting.

Reaching high power efficiency in a satellite communication system often imposes driving a high power amplifier (HPA), such as the traveling wave tube amplifier (TWTA), at or near its saturation point, resulting in severe nonlinear distortion of the signal. It is worth noting that OFDM systems are known to be significantly more sensitive to nonlinear distortions since its time domain signal has large PAPR.

To maintain an acceptable level of linearity, it is possible to operate the HPA with a large back off but this would be detrimental for the power efficiency [5]. Therefore, reliable and realizable compensation techniques are very crucial to the success and acceptance of the future satellite communication systems, which will be based on OFDM techniques.

Many approaches have been proposed for solving the problem of correct reception of the transmitted signal in those cases: PAPR reduction, equalization, and power amplifier linearization.

Up today, all PAPR reduction techniques can be thought of as a mapping from one signal representation to another that has a lower PAPR [6]. Then, various kinds of these techniques (see a review in [7]) were proposed such as the tone reservation [8], tone injection [9], active constellation extension (ACE) [10], some forms of partial transmit sequence (PTS) method [11], and selected mapping (SLM) method [12]. For OFDM with high number of subcarriers, the PAPR can be very important. If nothing is done for reducing it, it will result in a very high back off of the power amplifier. Thus, for OFDM modulations on satellite down links, methods for reducing the PAPR of the emitted signal will be necessary. Together with these reduction methods, the linearization of the HPA is also necessary in order to diminish the back off. PAPR reduction methods are out of the scope of our paper, and we will focus in the following in linearization techniques.

As we know, all equalization techniques proposed such as in [2, 13, 14] refer to processing the signal at the receiver side in order to recover the transmitted data. Nonlinear equalization is interesting if the transmission channel is nonlinear and frequency selective. In most satellite down link, the channel can be considered as a flat fading channel (Rician channel with very strong Line-of- sight (LOS) signal). The equalizer is then used only for compensating the nonlinear distortion introduced by HPA. For an economic point of view, it seems more interesting to compensate the nonlinear distortion on board the satellite instead of implementing an equalizer in every terminal.

Among all linearization techniques illustrated in literature, digital predistortion is one of the most cost effective and its principle is to distort the HPA input signal by an additional device called a predistorter which characteristics are the inverse of those of the amplifier [15].

In this paper, we will deal only with the last approach where its advantage lies in the fact that only a single system is needed for canceling the HPA nonlinearities at the satellite. Recently, some predistortion methods for HPAs have been introduced such as: inverse Volterra series [4], the rational function [16], Wiener-Hammerstein systems [3, 17], and memory polynomials [18], while, various adaptive algorithms including Volterra LMS and Volterra RLS have been released [19]. Accordingly, large-scale matrix computation is indispensable for these adaptive algorithms, yet still makes them impractical when a real-time system is necessary.

Other predistortion structures apply a look-up table (LUT) [5, 19], which is updated using the least-mean-square algorithm and the one proposed by Cavers [5]. This table is used to multiply the signal before feeding it to the HPA by the coefficient depending on the current signal amplitude and phase [20]. The LUT is the less complex of all techniques but it is a memoryless system and it cannot correct the memory effects in the power amplifier. Moreover, in order to implement this approach within a real OFDM systems, a large amount of RAM is required, which contents are updated with a low convergence speed.

It has been discovered that neural networks (NNs), which are nonlinear in their nature in addition to their very developed aspect, could be a good tool to compensate for nonlinearity [15]. More recently, the NN was proposed as an adaptive predistortion technique for communications systems [20–22]. These predistortions are realized with a multilayer perceptron (MLP) neural networks, to linearize stationary HPA, associated with various adaptive algorithms including the standard backpropagation (BP), conjugate gradient (CG), and natural gradient (NG) [23] which show a lower convergence speed and more complexity than the one used in this survey, if very accurate quality level is required.

Novelty of our paper consists in the use of three HPAs (memoryless and stationary, memoryless and nonstationary, and with memory) and new neural network predistorter (PD) structures associated with an adaptive algorithm which has been shown to exhibit a very good performance at a low computation complexity, a low amount of required RAM and faster convergence than other algorithms used in literature [2, 20, 21].

In this paper, we first describe (Section 2) the proposed satellite system considered with the applied neural network architecture. Then, we show (Section 3) how to determine the suitable NN predistortion structure aiming to give the best performance for the compensation of nonlinear distortions (AM/AM and AM/PM) caused by stationary memoryless TWTA.

In reality, the power amplifier characteristics may change over time because of temperature drift, component aging, and so forth. Therefore, we propose (Section 4) a novel adaptive predistortion structure based on NN able to adapt to these changes. The crucial point then is to find a suitable training algorithm able to cope with the described structure and the training data set. Accordingly, we report and compare (Section 5) five neural network training methods in terms of performance and complexity.

In Section 6, we propose two memory predistortion structures to mitigate the signal deterioration caused by nonlinear amplifier with memory effects, determine an optimal set of parameters for these PD, and compare its performances and complexity. We complete this study by conclusions (Section 7).

2. SATELLITE SYSTEM DESCRIPTION

Traditional satellites are nonregenerative transponders, which simply retransmit the received signal back to the ground [1].

New satellite generations have regenerative payloads [21], and the baseband transmitted signals are available on-board. Hence, uplink and downlink can be treated separately. Predistortion will be applied on the baseband transmitted signal (I/Q channels). Nevertheless, if the satellite is a nonregenerative transponder, it is always possible to apply the predistortion to the complex envelope obtained by down conversion.

Figure 1: The discrete communication system model for the downlink.

Figure 1 shows the baseband discrete equivalent communication system model for downlink with predistortion, where c_k is the kth transmitted symbol, which is mapped in 16-QAM, x_n is the nth transmitted OFDM sample, y_n is the same sample at the output of the predistorter, and z_n denotes the amplified one.

2.1. OFDM signal

Typically, an OFDM signal can be represented as

$$x(t) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} C_i e^{j2\pi f_i t}, \tag{1}$$

where C_i denotes the quadrature amplitude modulation (QAM) symbol, N is the number of subcarriers, and f_i is the ith subcarrier frequency which can be represented as follow:

$$f_i = i \cdot \frac{1}{NT_0}, \tag{2}$$

where T_0 is the sampling period of $x(t)$.

By discretizing $x(t)$ at $t = nT_0$, we have

$$x_n = x(nT_0) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} C_i e^{j2\pi in/N}. \tag{3}$$

2.2. HPA models

There are two technologies for the high power amplifiers (HPAs): traveling-wave tube amplifiers (TWTAs) and solid state power amplifiers (SSPAs).

Three different models of the HPA amplifiers are considered.

2.2.1. Stationary memoryless TWTA

If the HPA frequency response exhibits flat characteristics over its entire working frequency range, the nonlinearity is said to be frequency-independent or memoryless [3]. Then, in this situation the TWTA can be characterized by an AM/AM conversion and an AM/PM conversion.

According to Saleh's model [24] which has the advantage of exhibiting greater simplicity and accuracy than other models, AM/AM and AM/PM conversions of the TWTA can be represented as follow:

$$A(r) = \frac{\alpha_a r}{1 + \beta_a r^2}, \qquad P(r) = \frac{\alpha_p r^2}{1 + \beta_p r^2}, \tag{4}$$

where r is the input modulus of the TWTA, α_a and β_a are the parameters to decide the nonlinear level, and α_p and β_p are phase displacements. The values for these parameters are assumed to be $\alpha_a = 2$, $\beta_a = 1$, $\alpha_p = 4$ and $\beta_p = 9$ [15] which can approximate typical TWTA employed in satellites.

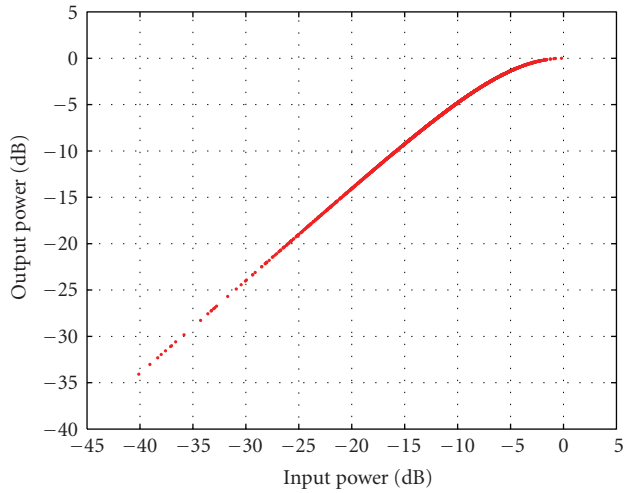

Figure 2: AM/AM characteristic.

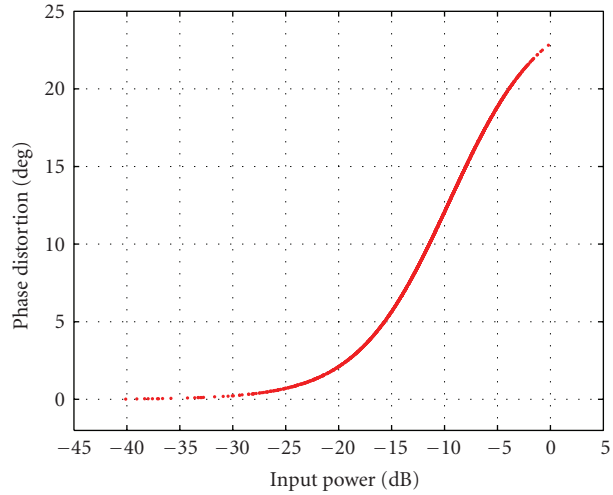

Figure 3: AM/PM characteristic.

The output of the TWTA can be represented as

$$z(t) = A(r) \exp\left(j \cdot \left(\omega_c t + \varphi(t) + P(r)\right)\right), \tag{5}$$

where $\varphi(t)$ is the phase of the input signal.

Figures 2 and 3 show the input and output relationships of AM/AM and AM/PM conversion characteristics for TWTA when the parameters are given by the above values.

2.2.2. Nonstationary memoryless TWTA

HPAs can no longer be considered as stationary in a real satellite system. In fact, power amplifiers operating under more stringent conditions may undergo slow but significant changes in their AM/AM and AM/PM characteristics basically due to factors like temperature, age of components, power level, biasing variations, frequency changes, and so on.

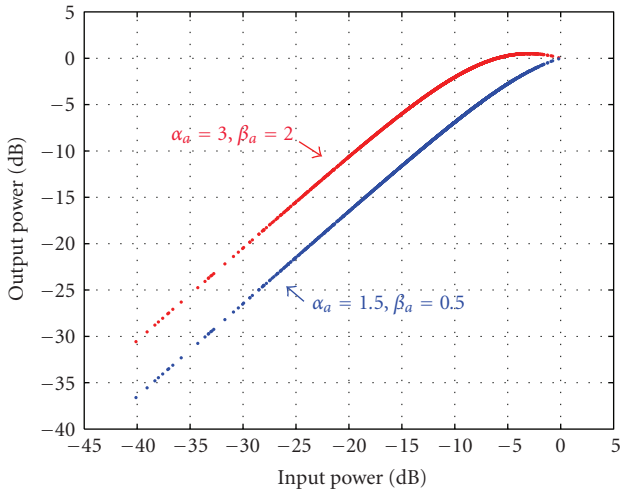

Figure 4: AM/AM characteristic variation.

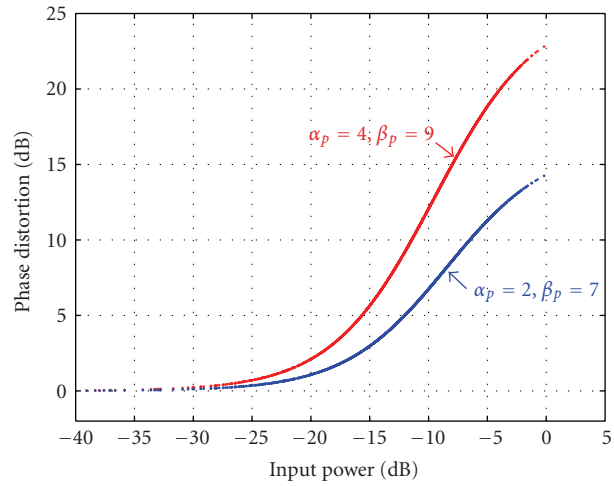

Figure 5: AM/PM characteristic variation.

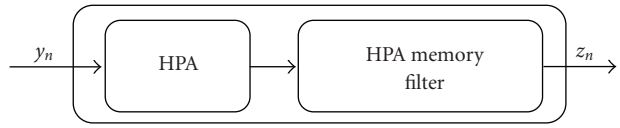

Figure 6: Model of the HPA with memory.

As a nonstationary (time-varying) model, we consider the memoryless model where the four parameters α_a, β_a, α_p, β_p are changing with time.

In this investigation, we assume that these parameters change linearly with time according to the following conditions.

(1) Four parameters change in the following ranges:

$$\begin{aligned} 1.5 \le \alpha_a \le 3, \\ 0.5 \le \beta_a \le 2, \\ 2 \le \alpha_p \le 4, \\ 7 \le \beta_p \le 9. \end{aligned} \tag{6}$$

(2) Input and output normalization conditions, $\beta_a = \alpha_a - 1$.

(3) Saturation condition, signal power is clipped above 0 dB.

The reason why we have chosen these conditions on the amplitude and phase is to maintain normalization constraints in both input and output and the saturation condition in the above range ($r > A_0$), even if the amplitude is changed. Where A_0 is the maximum output amplitude.

Figures 4 and 5 represent, respectively, the variation of AM/AM and AM/PM in order to show the extent of the HPA variations used in this work.

2.2.3. HPA with memory

With increasing bandwidth and average power of signals, HPA memory effects cannot be ignored. These memory effects may be explained by frequency dependence of components or by thermal phenomena [5]. As a model of the HPA with memory, it is considered to be a Hammerstein system which can be represented by a memoryless HPA followed by a linear filter (see Figure 6).

Figures 7 and 8 represent the AM/AM and AM/PM characteristics of the amplifier with memory, where we use a memoryless nonlinear HPA modeled by (4) where $\alpha_a = 2$, $\beta_a = 1$, $\alpha_p = 4$ and $\beta_p = 9$. The linear subsystem in the amplifier that captures the memory effects is modelled by a low-pass filter (see the frequency response in Figure 27) with 3 poles (0.7692, 0.1538, 0.0769) [3, 4]. For Figures 7 and 8, the average input power was set to 8 dB below the HPA saturation power. The memory effects are responsible for the thickening of curves.

2.3. Architecture of the applied neural network

Neural networks (NNs) are very attractive for adaptive predistortions due to their properties which are parallel distributed architecture, adaptive processing, nonlinear approximation, easy integration in large information processing chains, and efficient hardware implementation [25].

In this subsection, we present the most popular neural network architecture used in digital communications, as the multilayer perceptron (MLP).

A multilayer neural network (see Figure 9) is composed of neurons connected to each other. Connections are allowed from the input layer to the hidden layers and from the hidden layers to output layer.

It is well known that each neuron in the network is composed of a linear combiner and an activation function which gives the neuron output:

$$y_{lj} = f\left(\sum_{i=0}^{N_{l-1}} w_{l,j,i} x_{l-1,i} + b_{lj}\right), \tag{7}$$

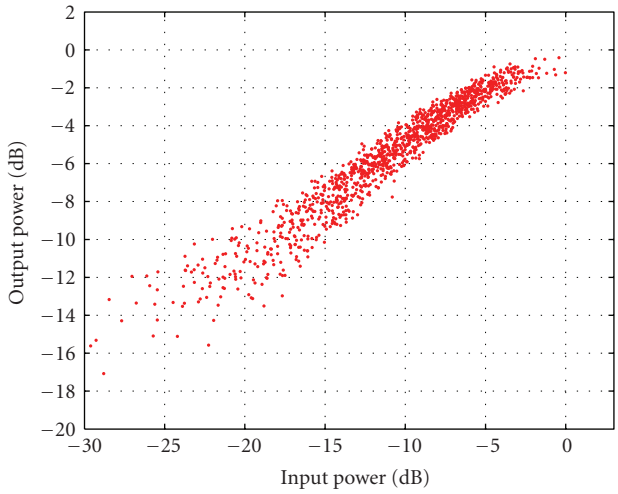

Figure 7: AM/AM characteristic.

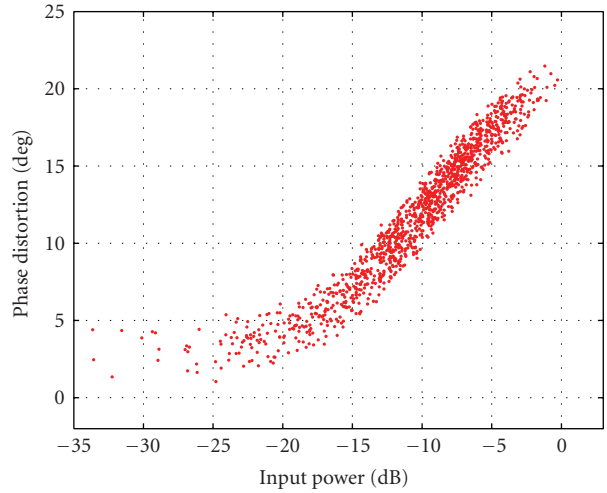

Figure 8: AM/PM characteristic.

where $w_{l,j,i}$ is the weight which connects the ith neuron in layer $l-1$ to the jth neuron in layer l, b_{lj} is the bias term, and $x_{l-1,i}$ denotes the ith component of the input signal to the neuron.

In general, the activation function is a nonlinear function (sigmoid function [20] or hyperbolic tangent [13]). For our NN, activation functions of the hidden layer are nonlinear and given by

$$f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{8}$$

The activation function of the output neurons is linear in our implementation.

In the next sections, we present the neural network predistorter structures with the associated learning algorithms for the three considered HPA models, which have been described previously.

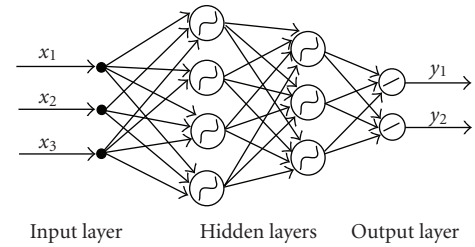

Figure 9: A multilayer neural network: this network has 3 layers, 3 input signals, 4 neurons in the first layer, 3 neurons in the second layer, and 2 neurons in the output layer (2 output signals).

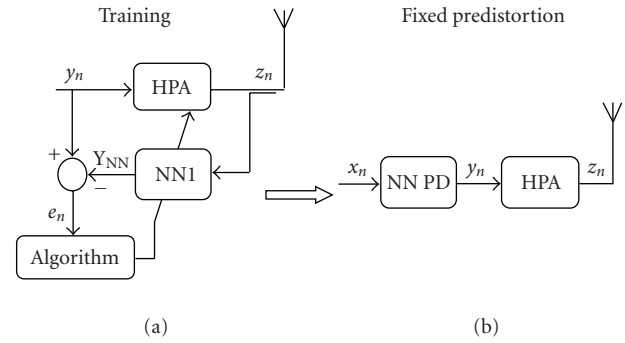

Figure 10: Block diagram for training of the PD with HPA.

3. PREDISTORTION OF STATIONARY MEMORYLESS TWTA

The basic idea proposed is to identify the TWT inverse transfer function with a feed-forward neural network. Therefore, by using this structure (see Figure 10), we aim at obtaining direct estimation of the amplitude and phase nonlinearities.

3.1. Training and generalization

Figure 10 shows the detailed scheme of predistortion system.

Training

We call "training process" the determination of the PD characteristics. Using the structure illustrated in Figure 10(a) we aim to identify the HPA inverse transfer functions, the complex envelope signals are differentiated and the error sent to "learning algorithm" bloc reacts on coefficients of NN1. In this indirect structure, a postdistortion is computed that will be applied as a predistortion system. Also, it has been demonstrated that this indirect approach is much more efficient than a direct one for predistortion system like polynomials or Volterra models [5].

Generalization (see Figure 10(b))

Coefficients of the NN1 are recopied on the NNPD that achieves the predistortion. The training procedure can be done on ground because the HPA is stationary. On

board the satellite, only the system of Figure 10(b) will be implemented.

3.2. Neural network structure

In this section, we present OFDM simulation results with the assumption that parameters α_a, β_a, α_p, β_p are time-invariant in order to determine the suitable neural network predistorter which allows the linearization of the power amplifier presented in Section 2.2.1.

The simulated OFDM system uses 64 carriers, a 16-QAM modulation, and a channel with additive white Gaussian noise (AWGN) was assumed to clearly observe the effect of nonlinearity and performance improvement by the proposed PD.

In the operation of the HPA, a relative level of power back off is needed to reduce distortion. However, this power back off is not so desirable because it reduces the power efficiency.

In our investigation, a compensation solution always exists in the range $r < A_0$, where A_0 is the maximum output amplitude. So, if the input average power is equal to A_0^2, we get maximum power efficiency, but the amplification will be highly nonlinear.

Thus, we need a criterion to show how much power back off is needed for optimum power efficiency. In the simulations, we define the input back off (IBO) as

$$\text{IBO} = 10\log_{10}\left(\frac{A_0^2}{P_{in}}\right), \tag{9}$$

where P_{in} is the average input power.

The neural predistorter considered was a multilayer perceptron (see Figure 11), which has two inputs, namely the I and Q components of the input signal complex envelope. The NN has two outputs that are the predistorted signals I and Q. Different structures have been tested, with first of all one hidden layer with 2 neurons, then while increasing the number of neurons progressively, before testing a network with two hidden layers, and also while increasing the number of neurons progressively on the two layers. As we have said earlier, activation functions of hidden layer are hyperbolic tangent, while the ones of the output layer are linear.

Figure 12 shows the performance of each predistorter on OFDM systems for an IBO of 8 dB.

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

All neural predistorters can reduce the SER compared to the one without any predistorter. The one that gets the best performances is the PD(2,9,2). Figure 13 shows the training curve versus iteration number for 64 and 16-QAM OFDM symbols and a TWT amplifier with an IBO equal to 8 dB. The feed-forward neural predistorter was configured as PD(2,9,2), the training algorithm used is the Levenberg-Marquardt one (see Section 5.1). After 200 training iterations, the MSE was lower than 1.5×10^{-5}, resulting in an accurate estimation of the coefficients for the neural predistorter.

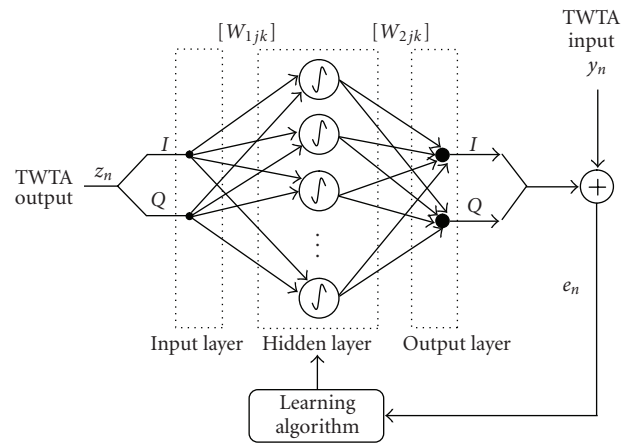

Figure 11: A multilayer neural network. The network has two layers, two input signals, one hidden layer, 2 neurons in the output layer, and 2 output signals. (Indexes I and Q refer to the real and imaginary parts, resp.)

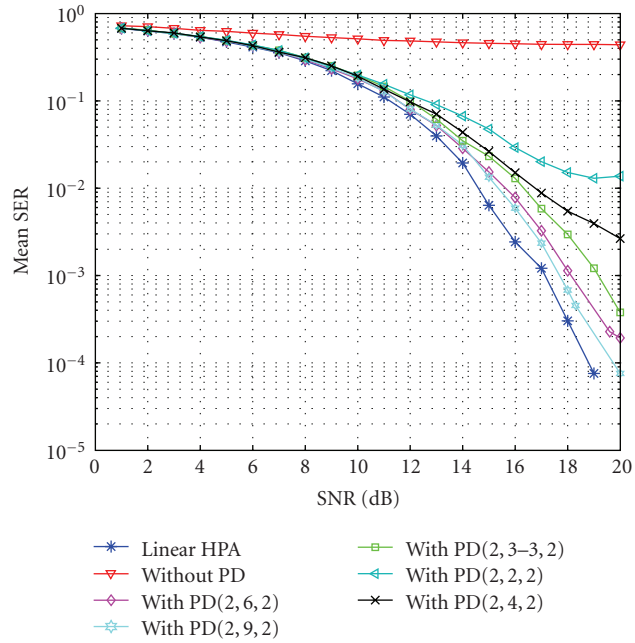

Figure 12: Symbol error rate of the OFDM system with predistorter versus SNR: a 16-QAM modulation is used on 64 carriers and IBO = 8 dB.

The symbol error rate (SER) performance curves, in Figure 14, show that the PD can significantly reduce nonlinear distortion in an OFDM system with low IBOs.

For an IBO of 7 dB, the performance of PD is very close to the performance of an ideal linear amplifier. When IBO is decreasing (5 dB, e.g.), the input signal modulus can be higher than A_0 with a strong probability. This gives an irreducible error at higher SNR.

As we have said in the introduction of the paper, it will be necessary to decrease the PAPR with PAPR reduction methods together with linearization of the HPA.

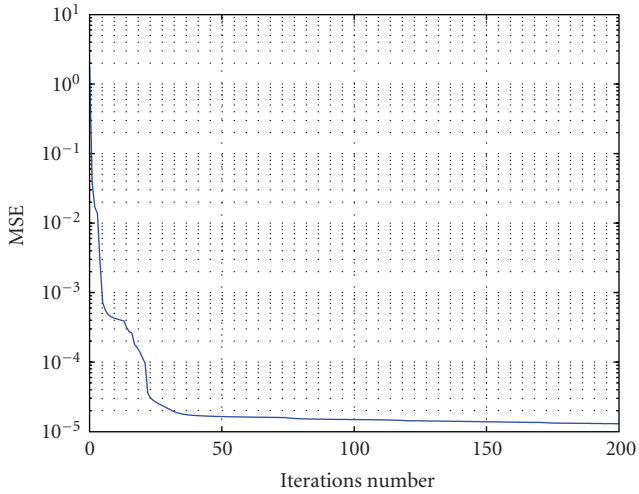

Figure 13: MSE versus iterations number for 64 carriers and 16-QAM OFDM, PD(2,9,2).

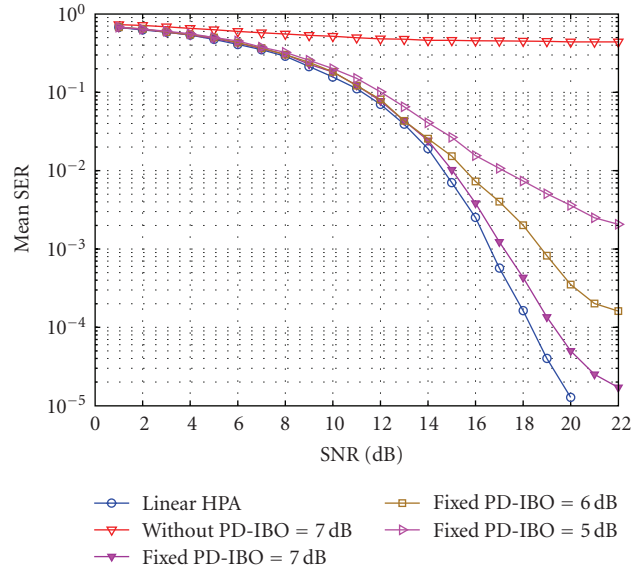

Figure 14: SER performance of PD in OFDM system, with time-invariant TWTA.

4. PREDISTORTION OF NONSTATIONARY MEMORYLESS TWTA

As we mentioned previously, the HPA can be a time-varying system. In this subsection, we assume that the four parameters α_a, β_a, α_p, β_p are time-varying; thus, the PD must track variations of α_a, β_a, α_p, and β_p.

Previously, we took into account the convenience of performing the estimation of the inverse HPA characteristics in a postdistortion stage rather than in a simple predistortion one. According to this, the predistortion architectures presented here are basically derived from a postdistortion adaptive structure which may employ two general alternatives for its operation. These alternatives are the following.

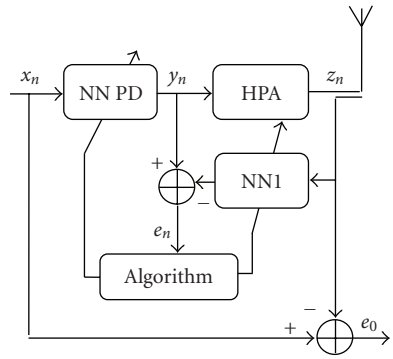

Figure 15: Simultaneous PD updating.

1st Alt: loading the predistorter with completely trained coefficients after a complete learning stage (PD is here stationary (see Figure 10)).

2nd Alt: simultaneous updating of the predistorter during the adaptation at the postdistortion loop (PD is here adaptive (see Figure 15)).

Figure 15 shows the detailed scheme of an adaptive predistortion system based on feed-forward neural network. x_n denotes the input signal to the predistorter, y_n denotes the output signal from the predistorter and sent as input to the TWTA, and z_n denotes the TWTA output signal.

The weights of the neural network predistorter (NN PD) are determined by copying the weights of NN1. These weights are adjusted using an adaptive algorithm based on Levenberg -Marquardt method.

As we mentioned earlier that HPA parameters change linearly with time, we use then the following function to define the temporal variation for each parameter while respecting the conditions shown in Section 2.2.2:

$$f(t) = A * t + C, \tag{10}$$

where A is the constant that defines the speed of the temporal variation and C is the constant that defines the initial value.

The SER performance of the adaptive predistortion structure in OFDM system is compared to the fixed predistortion which has determined the inverse transfer function of the initial HPA with ($\alpha_a = 2$, $\beta_a = 1$, $\alpha_p = 4$, $\beta_p = 9$). The neural network structure used here is the one determined in Section 3.2.

The following figure shows the SER performance of the proposed adaptive predistortion compared to the one of the fixed predistortion with a time-varying HPA where ($\alpha_a = 1.5$, $\beta_a = 0.5$, $\alpha_p = 2.5$, $\beta_p = 7.5$).

We deduce from Figure 16 that if the variation of the HPA is not tracked, the performance degrades strongly at higher SNR. The simulation results thus show that this ability to use an adaptive predistortion adds value to the system performance.

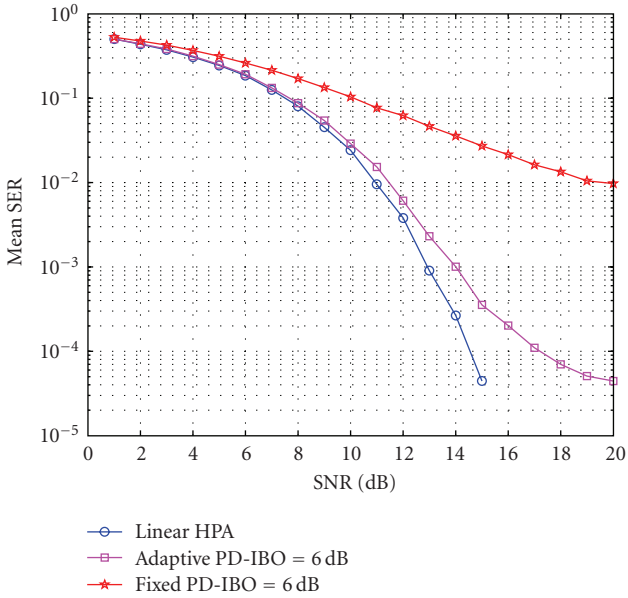

Figure 16: SER performance of PD in OFDM, with time-varying TWTA, IBO = 6 dB.

5. COMPARISON OF NN ALGORITHMS FOR ADAPTIVE PREDISTORTION

Previously, we have determined the suitable structure of NN that forms an adaptive nonlinear device which can approximate the inverse transfer functions of HPA nonlinearities (AM/AM and AM/PM). It consists of two inputs and two outputs (I and Q), one hidden layer of nine neurons whose activation functions on the hidden layer are hyperbolic tangent, while output layers are linear.

After the determination of the global neural network architecture and its suitable structure, the crucial point next is to find a suitable training algorithm able to cope with the described network and the training data set. In short, this section compares the performance of five neural network training methods to enhance the compensation for nonlinear distortions (AM/AM and AM/PM) caused by HPAs used in satellites.

5.1. Training algorithms

In this subsection, we review the different algorithms used in this investigation: gradient descent backpropagation (GD), gradient descent backpropagation with the momentum (GDm), conjugate gradient BP (CGF), quasi-Newton (BFG), and Levenberg-Marquardt (LM).

(i) Gradient descent BP (GD)

The gradient-based methods are the most straightforward training algorithms for feed-forward multilayer perceptron networks [26], and there are two different methods in which this gradient descent algorithm can be implemented: incremental mode and batch mode. The simplest implementation of back-propagation learning updates the network weights and biases in the direction in which the performance function decreases more rapidly. The new weight vector x_{k+1} can be adjusted as

$$x_{k+1} = x_k - \mu g_k, \tag{11}$$

where x_k is the vector of current weights and biases, μ is the learning rate, and g_k is the gradient of the error with respect to the weight vector. The computation of g_k is presented in [26]. The negative sign indicates that the new weight vector x_{k+1} is moving in a direction opposite to that of the gradient.

(ii) Gradient descent BP with momentum (GDm)

The convergence of the network by backpropagation is a crucial problem because it requires many iterations. To mitigate this problem, a parameter, called "Momentum," can be added to BP learning method by making weight changes equal to the sum of fraction of the last weight change and the new change suggested by the gradient descent BP rule (see equation (6)) [26]. The momentum is an effective means not only to accelerate the training but also to allow the network to respond to the (local) gradient.

Then, the new weight vector is adjusted as [26]

$$x_{k+1} = x_k - \mu g_k + \alpha(x_k - x_{k-1}), \tag{12}$$

where the parameter α is the momentum constant, which can be any number between 0 and 1.

(iii) Conjugate gradient BP (CGF)

The standard back-propagation algorithm adjusts the weights in the steepest descent direction, which does not necessarily produce the fastest convergence [27]. And it is also very sensitive to the chosen learning rate, which may cause an unstable result or a long-time convergence [15]. As a matter of fact, several conjugate gradient algorithms have recently been introduced as learning algorithms in neural networks [26]. They use, at each iteration of the algorithm, different search directions in a way which produce generally faster convergence than steepest descent directions [28]. In the conjugate gradient algorithms, the step size is adjusted at each iteration. The conjugate gradient used here was proposed by Fletcher and Reeves [26, 27].

All conjugate gradient algorithms start out by searching in the steepest descent direction on the first iteration:

$$p_0 = -g_0. \tag{13}$$

The search direction at each iteration is determined by updating the weight vector as

$$x_{k+1} = x_k - \mu_k p_k, \tag{14}$$

where

$$p_k = -g_k + \beta_k p_{k-1}. \tag{15}$$

For the Fletcher-Reeves update, the constant β_k is computed by

$$\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}. \tag{16}$$

This is the ratio of the norm squared of the current gradient to the norm squared of the previous gradient.

(iv) BFGS quasi-Newton (BFG)

In Newton methods, the update step is adjusted as

$$x_{k+1} = x_k - H_k^{-1} g_k, \tag{17}$$

where H_k is the Hessian matrix (second derivatives) of the performance index at current values of the weights and biases.

Newton's methods often converge faster than conjugate gradient methods. Unfortunately, they are computationally very expensive, due to the extensive computation of the Hessian matrix H coming along with the second-order derivatives [27]. The quasi-Newton method that has been the most successful in published studies is the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) update [26].

(v) Levenberg Marquardt (LM)

Similarly to quasi-Newton methods, the Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. Under the assumption that the error function is some kind of squared sum, then the Hessian matrix can be approximated as

$$H = J^T J, \tag{18}$$

and the gradient can be computed as

$$g = J^T e, \tag{19}$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, the Jacobian matrix determination is computationally less expensive than the Hessian matrix, e is a vector of network errors.

Then, the update can be adjusted as

$$x_{k+1} = x_k - \left[J^T J + \mu I\right]^{-1} J^T e. \tag{20}$$

The parameter μ is a scalar controlling the behavior of the algorithm. For $\mu = 0$, the algorithm follows Newton's method, using the approximate Hessian matrix. When μ is high, this becomes gradient descent with a small step size.

5.2. Results of performance comparison

It is very difficult to know which training algorithm will be the fastest and the most adequate for a given problem. It will depend on several factors including the complexity and the type of the problem, the data set of the training base, the number of weights and biases in the network, and the required training time, hardware resources, and mean squared error between the actual and desired network responses.

In this subsection, we carry out a certain number of comparisons of the various training algorithms. The neural network consists in a predistortion to compensate for nonlinearities caused by the power amplifier in OFDM system with 64 carriers and 16-QAM.

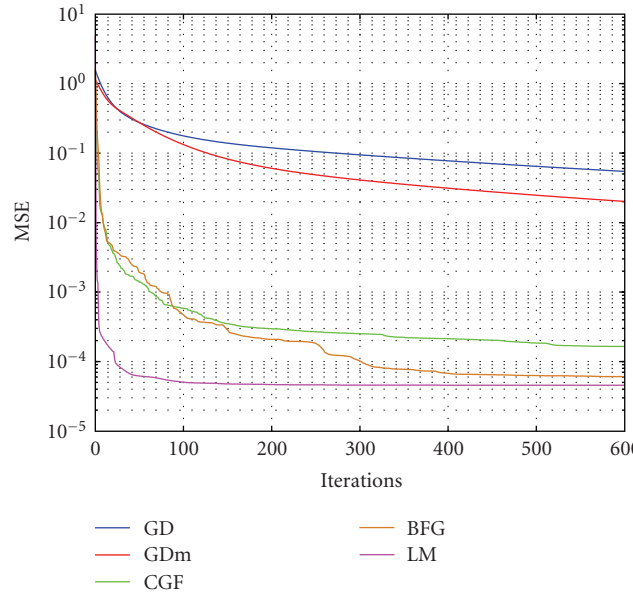

Figure 17: Mean square error versus iteration for different algorithms.

In fact, the goal of the network in this case is to estimate inverse transfer functions (AM/AM and AM/PM) of the amplifier in question.

According to [5], the most suitable neural network to this problem is of feed-forward type with two inputs, two outputs, and a hidden layer of nine neurons (2-9-2). The activation function is sigmoid for hidden layer neurons and linear for the output ones. Also, 312 16-QAM OFDM samples were employed for the learning process.

In this investigation, (all experiments were carried out Matlab running on HP pavilion ze5500 with a Mobile Intel Pentium IV 2.66 GHz processor and 512 Mo RAM) the neural network is employed to approximate inverse transfer functions of the amplifier used in satellite transmitter. Accordingly, the accuracy expected from the approximation can affect the performance of the various algorithms. Figure 17 plots for each method, the mean square error versus iteration number averaged over 30 simulations. We can see that the error in the LM algorithm decreases much more rapidly than the other algorithms.

At this point, we can say that the LM algorithm gives more accurate results in terms of convergence speed. Moreover, it is important to consider the algorithmic complexity. The following table summarizes the results of the comparative study of the five-mentioned algorithms in terms of complexity. The variable number of floating operations (*Nflops*) is the number of computation that each method required to run per epoch while *Ntflops* is the number of computations that each method required to reach the minimum MSE. In each case, the network is trained until the squared error is less than 10^{-3}.

Table 1: Computation comparison for different algorithms.

Algorithm	*Nflops*	*Ntflops*
LM	5973400	1.5651e + 007
BFG	402710	2.225e + 007
CGF	285300	2.472e + 007
GDm	296574	*
GD	197663	*

(*required training goal was not reached with 2×10^5 Epochs).

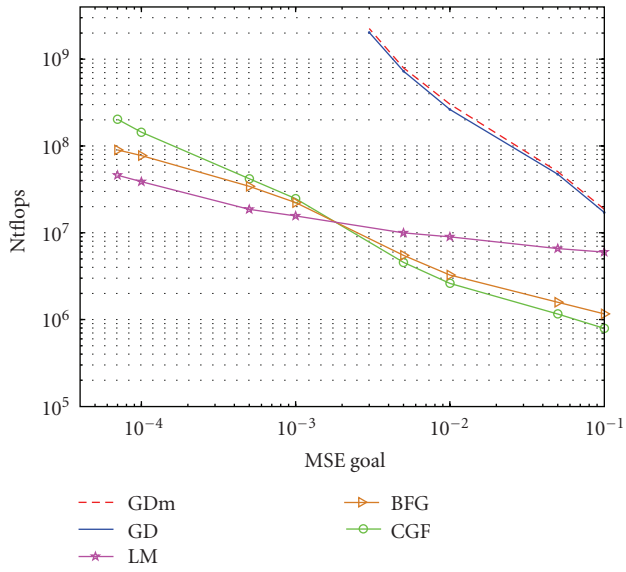

Figure 18: Computation number required versus mean square error.

For the calculation of the number of floating operations, additions and subtractions are one flop if real and two if complex. Multiplications and divisions count one flop each if the result is real and six flops if it is not.

As can be seen in Table 1, the Levenberg-Marquardt algorithm is obviously quite well suited for the used neural network training. Although it requires the most significant number of computation per epoch (because of the Hessian computation), it requires the least amount of total computation flop (*Ntflops*) for the mean square error convergence goal.

Figure 18 indicates the number of computation (*Ntflops*) required to be converged versus the mean square error convergence goal. Again, we observe, as the error goal is reduced, the improvement provided by the LM algorithm becomes more pronounced. LM algorithm performs better than other algorithms as the MSE goal is reduced.

As we mentioned before, the HPA can be a time-varying system. In this subsection, we assume that the four parameters α_a, β_a, α_p, β_p are time varying as presented in Section 2.2.2. Thus, we study the performances of these algorithms for an adaptive predistortion (see Figure 15), in the case of a nonstationary amplifier.

In a first phase, we train the neural network during a "T_0" time in order to fix the "NN PD" (see Figure 10). Figure 19

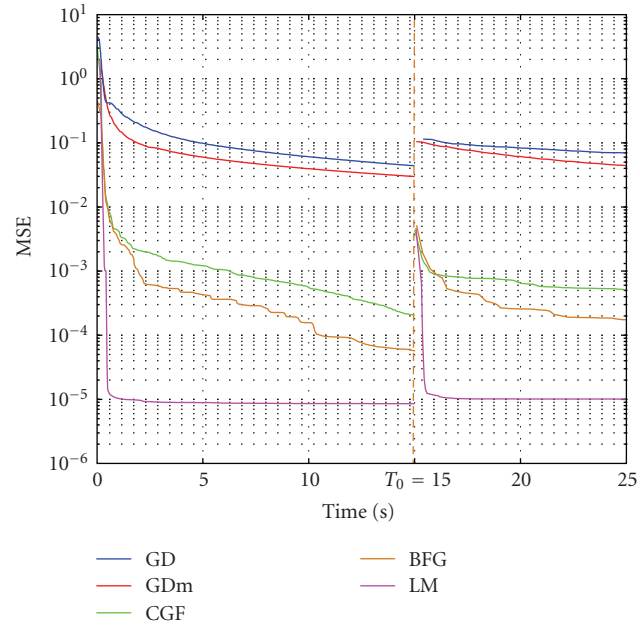

Figure 19: Mean square error versus time.

represents the learning curves of the neural network with various algorithms according to time "t" such as $(0 < t < T_0)$.

The amplifier used is a Saleh's TWTA model with a given set of parameters ($\alpha_a = 2$, $\beta_a = 1$, $\alpha_p = 4$, $\beta_p = 9$). We note that all algorithms converge towards different MSE and that of LM is the smallest.

At time "T_0," we abruptly change the parameters of the amplifier such as ($\alpha_a = 3$, $\beta_a = 2$, $\alpha_p = 2$, $\beta_p = 7$) and we compare the convergence of the various algorithms. According to Figure 19, we clearly notice that the Levenberg-Marquardt algorithm is the fastest and ensures the best convergence towards a minimum of error.

6. HPA WITH MEMORY PREDISTORTION

Earlier, we have presented the suitable neural network structures and the associated algorithm to enhance the predistortion of AM/AM and AM/PM nonlinearities due to stationary and nonstationary HPAs used in satellites.

In future satellite generations which use OFDM as modulation scheme, memory effects of high power amplifiers can no longer be ignored due to the broadband input signal. So, in this section we analyzed neural network structures with tap delay (with memory) used as predistorter.

The general feedback loop structure presented in Figure 15 is applied here to identify the predistorter with memory where NNPD and NN1 have the same structure. Thus, the input and output of the HPA (y_n and z_n) are accessed with the aim of finding NNPD by using NN1. The objective is to obtain a linear behavior of the cascaded NNPD and HPA by matching NN1 to the inverse nonlinearities with memory in the HPA.

For the adaptive training algorithm, we have used the Levenberg-Marquardt which exhibits the best performance

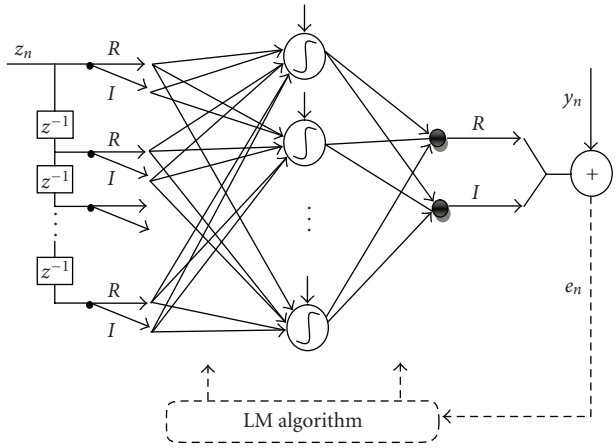

Figure 20: Fully connected NN predistorter structure.

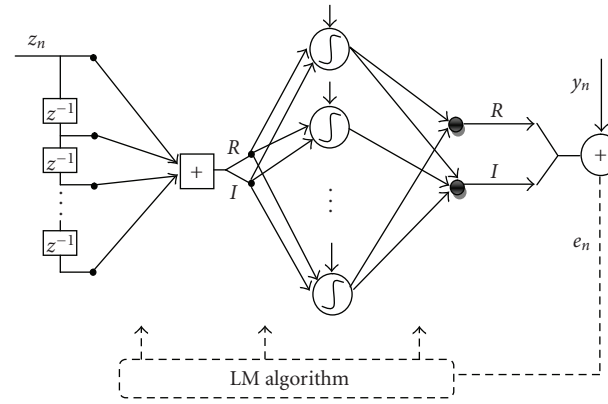

Figure 21: Linear network LN + nonlinear network NLN predistorter structure.

compared to other training algorithms as we demonstrated in the last section.

In this investigation, the Hammerstein model is used to represent the nonlinear HPA with memory effects (see Figure 6). A 16-QAM modulation scheme is selected in OFDM with a total of 64 subcarriers. The memoryless nonlinear model for HPA selected is the Saleh's TWTA modeled by (4), where $\alpha_a = 2$, $\beta_a = 1$, $\alpha_p = 4$, and $\beta_p = 9$. The linear subsystem in the HPA that captures the memory effects is implemented using a low-pass filter with 3 poles (0.7692, 0.1538, 0.0769) [3, 4].

6.1. Applied neural network structures

Two neural network predistortion structures have been proposed. They are the following.

(1) A fully connected multilayer NN (FCNN) predistorter with memory (see Figure 20). The input z_n is connected to nine neurons in the hidden layer. The output neurons are real and imaginary parts. The fully connected NN aims at simultaneously mitigating memory and HPA nonlinear effects.

(2) A neural network mimetic structure (see Figure 21), which combines a linear network (LN) and a memoryless nonlinear neural network (NLN). The LN-NLN predistorter is composed of a linear filter H followed by a memoryless neural network, with one hidden layer, with nine neurons (with sigmoid activation function), and with two linear neurons in the output layer. Using this mimetic scheme (LN-NLN), we realize separately the memory predistortion with the linear network and the predistortion of the memoryless HPA nonlinearities with the nonlinear neural network. A comparative study of these two structures for predistortion has been made in terms of performance and complexity. To ensure a good comparison between the different structures, the same length of the tap delay line has been chosen (4 memory cells). These memory predistortion structures have been trained with LM algorithm.

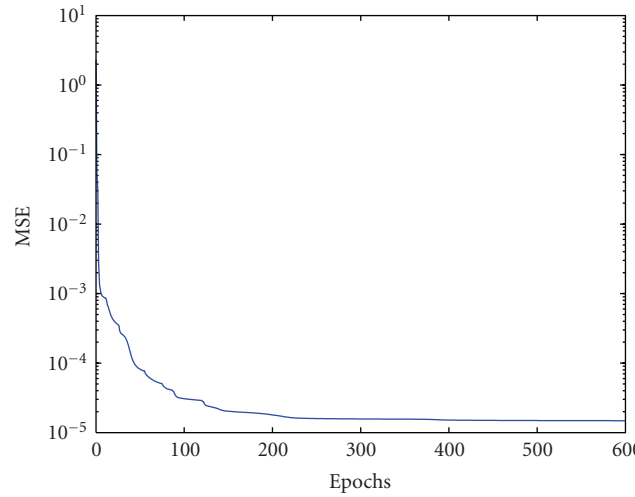

Figure 22: FCNN structure learning sequence.

6.2. Performance comparison

Figure 22 shows the fully connected NN (FCNN) structure MSE evolution versus iteration number for the proposed systems. The training algorithm was the Levenberg-Marquardt one, only 300 training iterations, and the MSE was 1.32×10^{-5}, resulting an accurate estimation of the coefficients for the memory neural predistorter.

For the NN mimetic structure (LN-NLN), we present on Figure 23(a) the linear NN (LN) MSE evolution. After one iteration, LN has converged this well, demonstrating the need to use Levenberg-Marquardt as training algorithm than the other ones. Figure 23(b) shows the memoryless NN (NLN) MSE evolution curve in function of the iteration number. The NLN has converged after 300 iterations, and the MSE was 8×10^{-6}.

Symbol error rate (SER) diagrams are a typical performance measure for qualifying the compensating ability of proposed predistortion structures to reduce HPA distortions. Then, Figure 24 shows the SER performance versus signal to noise ratio (SNR) in systems with a linear HPA along with nonlinear memory HPA without predistortion

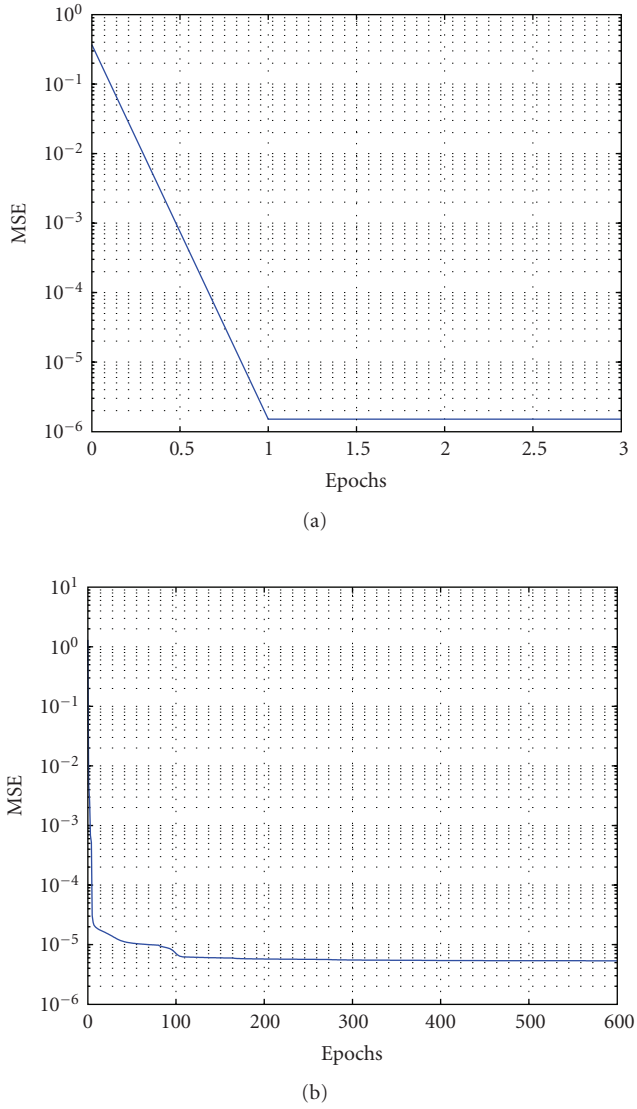

Figure 23: (a) LN learning sequence and (b) NLN learning sequence.

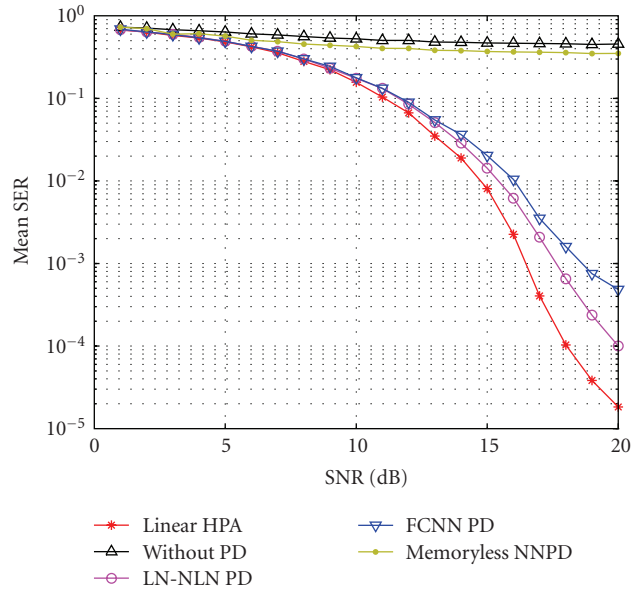

Figure 24: SER performance for 16-QAM OFDM with 64 carriers at IBO = 7 dB.

and a nonlinear memory HPA with NN memoryless predistortion, FCNN predistortion, and LN-NLN predistortion. The realistic level of memoryless nonlinear distortions is considered by working with input back off (IBO) equal to 7 dB.

The 16-QAM constellation of the received signal in system with memory HPA is shown in Figure 25(a), with an apparent wholly phase rotation caused by AM/PM distortion. We see in Figure 25(b) that a memoryless PD has limited effects on the constellation recovery, while Figures 25(c) and 25(d) convince LN-NLN PD and FCNN PD, respectively, achieve superior performance in rectifying the received constellation.

We note from Figures 24 and 25 that the two memory NN structures (FCNN and LN-NLN) succeed in reducing considerably the SER to the one without any predistortion and provide us with a better performance in correcting the received constellation. Nevertheless, the mimetic structure (LN-NLN) performs slightly better than the FCNN structure when they are both trained with LM.

Figure 26 represents AM/AM curves of the amplified signal versus input signal without predistortion and predistorted signal versus input signal for two studied predistortions (memoryless NN PD and LN-NLN).

The FCNN predistortion is not included in this comparison since it has a slightly lower performance than the LN-NLN predistortion.

Memory effects are not taken into account in the memoryless NN PD structure. Thus, we can see on Figure 26(a) that the AM/AM curve of the concatenated system (memoryless NN PD + HPA) is thicker than the resulting AM/AM curve of Figure 26(b) obtained with an LN-NLN PD.

We show on Figure 27 the response of both the HPA memory filter and the converged linear network (LN). We see that the HPA memory has also been successfully identified and compensated by the LN.

6.3. Complexity comparison

Complexity is evaluated by computing the number of floating operations (addition, multiplication, division) required to learn or to run these structures. The training algorithm was the Levenberg-Marquardt one, 312 16-QAM OFDM samples were employed for the learning and generalization phase.

Experiments were carried out by Matlab using the same computer as in Section 5.2.

Table 2 summarizes the results of the comparative study of the two-mentioned structures in terms of complexity. The variable *Nlearnflops* and *Nrunflops* are the number of computation (number of floating operations) that each

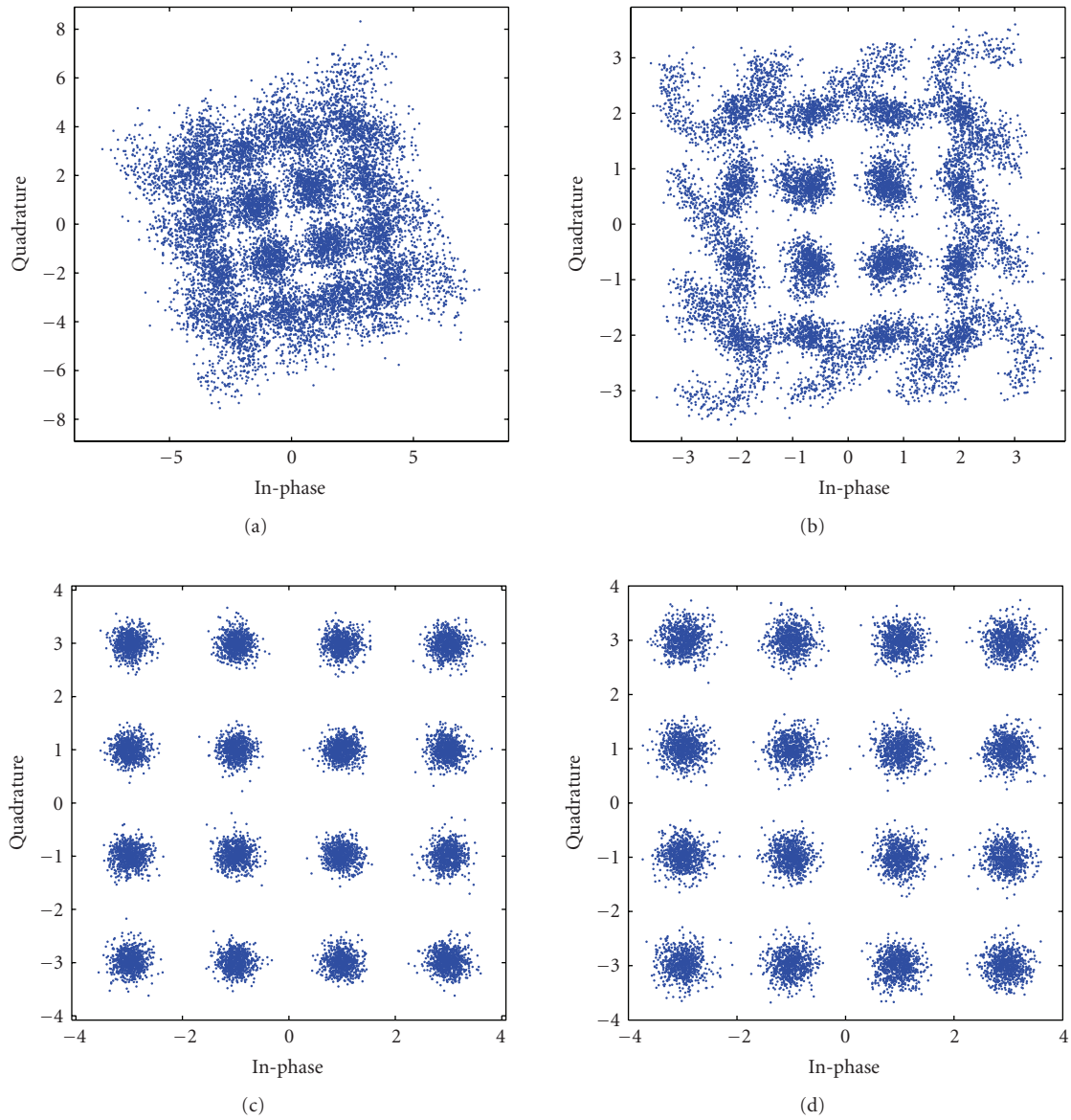

Figure 25: Constellation of received signal at IBO = 7 dB and SNR = 20 dB: (a) polluted by HPA, (b) memoryless NN PD, (c) LN-NLN PD, (d) FCNN PD.

Table 2: Complexity comparison of predistorter structures.

Structure	*Nlearnflops*		LRatio	*Nrunflops*		RRatio
LN-NLN	LN	57	1	LN	25	1
	NLN	5059		NLN	64	
	Total	5116		Total	89	
FullyCNN	32981		6.45	145		1.63

structure required to learn per epoch and to run, respectively, per OFDM sample.

Let us recall for the number of floating operations, additions and subtractions are one flop if real and two if complex. Multiplications and divisions count one flop each if the result is real and six flops if it is not.

We deduce from Table 2 that the computational complexity of the mimetic structure (LN-NLN) is much lower than the fully connected one (FCNN) with an approximate ratio more than 80% in learning phase and nearly 40% in generalization phase.

7. CONCLUSION

In this paper, we have proposed adaptive predistortion structures based on feed-forward neural networks. These structures were applied to 64 carriers and 16-QAM OFDM transmission over three different power amplifiers used in satellite (stationary memoryless TWTA, nonstationary memoryless TWTA, and memory HPA).

For memoryless PD, we have confirmed that the proposed neural network predistorter consisting with nine

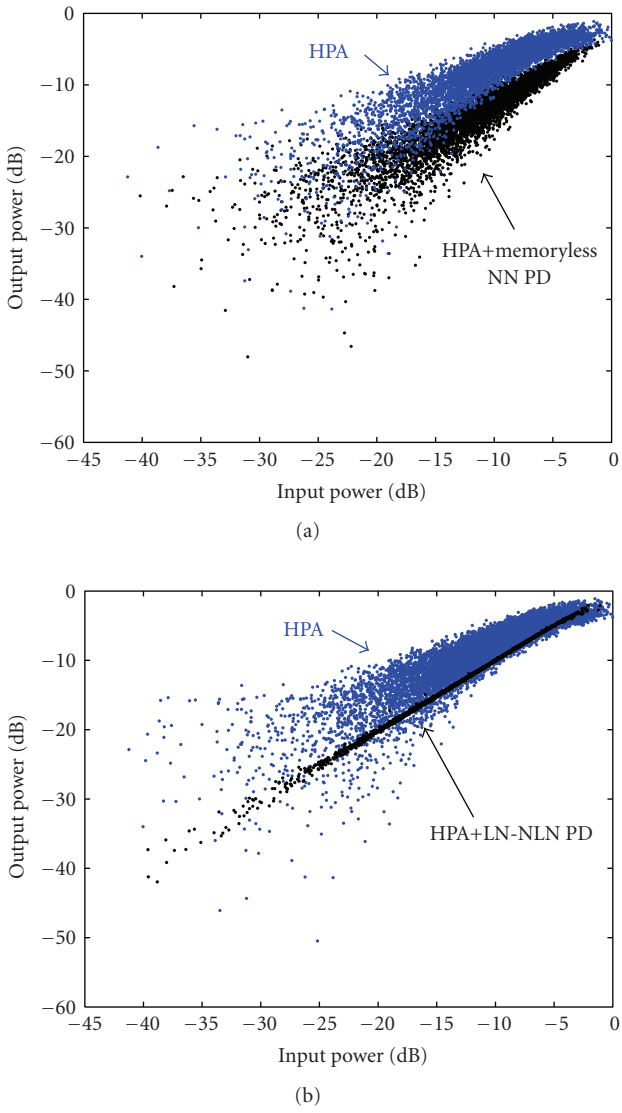

Figure 26: AM/AM curves for: (a) memoryless NN PD, (b) LN-NLN PD.

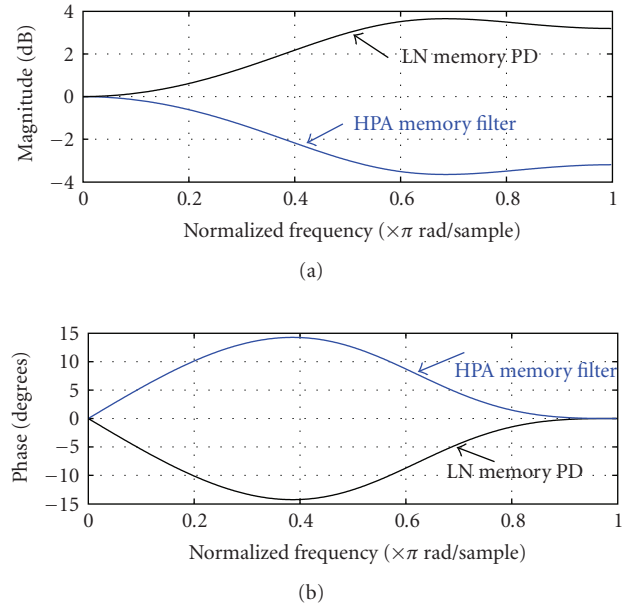

Figure 27: Memory PD identification.

neurons in one hidden layer and two linear neurons in the output layer gives a good performance improvement of the transmission quality.

Performances of the proposed neural network predistorter depend on the BP training algorithm. So this dissertation compares the performance of five neural network training methods in adaptive predistortion.

This comparison examined through computer simulations with a Saleh's TWT amplifier is based on learning speed versus iteration number, floating operation number per epoch, and the total number of computation that each method required to reach a fixed MSE.

We demonstrated by simulations that the Levenberg-Marquardt algorithm has the fastest convergence in terms of iteration number. In many cases, LM is able to obtain lower mean square errors than any of the other algorithms tested. Also LM requires the lower amount of computation for low MSE. This advantage is mainly noticeable if every accurate quality level is required.

In short, LM is the winner of all comparisons with the other algorithms. Well, we demonstrated that the performance of an OFDM system suffering from nonlinear distortions, caused by nonstationary HPA, can be greatly improved by the proposed adaptive neural predistorter using Levenberg-Marquardt algorithm which proved to be efficient.

For HPA with memory, we have proposed two neural network structures (LN-NLN and FCNN). The LN-NLN tries to predistort separately HPA nonlinear distortions and memory by two separate tools (linear network (LN) to mitigate memory memoryless NN (NLN) to invert the nonlinearities). The fully connected NN (FCNN) deals with these two problems as a whole and yields a multidimensional function with memory to predistort both memory and nonlinear distortions.

According to the found results, we come to conclude that the LN-NLN structure performs slightly better than the fully connected NN (when they are both trained with LM), with an important advantage that its computational complexity is much lower than the FCNN.

REFERENCES

[1] E. Kofidis, V. Dalakas, Y. Kopsinis, and S. Theodoridis, "A novel efficient cluster-based MLSE equalizer for satellite communication channels with M-QAM signaling," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 34343, 16 pages, 2006.

[2] M. Ibnkahla and J. Yuan, "A neural network MLSE receiver based on natural gradient descent: application to satellite communications," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 16, pp. 2580–2591, 2004.

[3] T. Wang and J. Ilow, "Compensation of nonlinear distortions with memory effects in OFDM transmitters," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '04)*, vol. 4, pp. 2398–2403, Dallas, Tex, USA, November-December 2004.

[4] J. Li and J. Ilow, "Adaptive Volterra predistorters for compensation of non-linear effects with memory in OFDM transmitters," in *Proceedings of the 4th Annual Communication Networks and Services Research Conference (CNSR '06)*, pp. 100–103, Moncton, Canada, May 2006.

[5] P. Jardin and G. Baudoin, "Filter look up table method for power amplifiers linearization," *IEEE Transaction on Vehicular Technology*, vol. 56, no. 3, pp. 1076–1087, 2007.

[6] R. J. Baxley, *Analyzing selected mapping for peak-to-average power reduction in OFDM*, M.S. thesis, School of Electrical and Computer Engineering, Georgia Institute Technology, Atlanta, Ga, USA, April 2005.

[7] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Communications*, vol. 12, no. 2, pp. 56–65, 2005.

[8] A. Gatherer and M. Polley, "Controlling clipping probability in DMT transmission," in *Proceedings of the 31st Asilomar Conference on Signals, Systems & Computers (ACSSC '97)*, vol. 1, pp. 578–584, Pacific Grove, Calif, USA, November 1997.

[9] J. Tellado, *Multicarrier Modulation with Low PAR: Applications to DSL and Wireless*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[10] B. S. Krongold and D. L. Jones, "PAR reduction in OFDM via active constellation extension," *IEEE Transactions on Broadcasting*, vol. 49, no. 3, pp. 258–268, 2003.

[11] W. S. Ho, A. S. Madhukumar, and F. Chin, "A novel two-layered suboptimal approach to partial transmit sequences," in *Proceedings of the 57th IEEE Semiannual Vehicular Technology Conference (VTC '03)*, vol. 2, pp. 1268–1272, Jeju, Korea, April 2003.

[12] R. W. Bäuml, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electronics Letters*, vol. 32, no. 22, pp. 2056–2057, 1996.

[13] Q. M. Rahman, M. Ibnkahla, and M. Bayoumi, "Neural network based channel estimation and performance evaluation of a time varying multipath satellite channel," in *Proceedings of the 3rd Annual Communication Networks and Services Research Conference (CNSR '05)*, pp. 74–79, Halifa, Canada, May 2005.

[14] S. Bouchired, M. Ibnkahla, D. Roviras, and F. Castanie, "Equalization of satellite UMTS channels using neural network devices," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '99)*, vol. 5, pp. 2563–2566, Phoenix, Ariz, USA, March 1999.

[15] R. Zayani, R. Bouallegue, and D. Roviras, "An adaptive neural network pre-distorter for non stationary HPA in OFDM systems," in *Proceedings of the 15th European Signal Processing Conference (EUSIPCO '07)*, Poznan, Poland, September 2007.

[16] A. Zhu and T. J. Brazil, "An adaptive Volterra predistorter for the linearization of RF high power amplifiers," in *Proceedings of the IEEE MTT-S International Microwave Symposium Digest (MWSYM '02)*, vol. 1, pp. 461–464, Seattle, Wash, USA, June 2002.

[17] L. Ding, R. Raich, and G. T. Zhou, "A Hammerstein predistortion linearization design based on the indirect learning architecture," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 3, pp. 2689–2692, Orlando, Fla, USA, May 2002.

[18] M. A. Nizamuddin, P. J. Balister, W. H. Tranter, and J. H. Reed, "Nonlinear tapped delay line digital predistorter for power amplifiers with memory," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '03)*, vol. 1, pp. 607–611, New Orleans, La, USA, 2003.

[19] T. Hoh, G. Jian-Hua, G. Shu-Jian, and W. Gang, "A nonlinearity predistortion technique for HPA with memory effects in OFDM systems," *Nonlinear Analysis: Real World Applications*, vol. 8, no. 1, pp. 249–256, 2007.

[20] J. Pochmara, "Improving compensation of nonlinear distortions in OFDM system using recurrent neural network with conjugate gradient algorithm," in *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '04)*, vol. 1, pp. 180–185, Barcelona, Spain, September 2004.

[21] F. Langlet, H. Abdulkader, D. Roviras, A. Mallet, and F. Castanié, "Comparison of neural network adaptive predistorsion techniques for satellite down links," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '01)*, vol. 1, pp. 709–714, Washington, DC, USA, July 2001.

[22] M. Ibnkahla, "Adaptive predistortion techniques for satellite channel equalization," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '00)*, Istanbul, Turkey, June 2000.

[23] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.

[24] A. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, 1981.

[25] M. Ibnkahla, "Applications of neural networks to digital communications—a survey," *Signal Processing*, vol. 80, no. 7, pp. 1185–1215, 2000.

[26] Y. Wang, S.-P. Kim, and J. C. Principe, "Comparison of TDNN training algorithms in brain machine interfaces," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '05)*, vol. 4, pp. 2459–2462, Montreal, Canada, July 2005.

[27] U. Seiffert, "Training of large-scale feed-forward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '06)*, pp. 5324–5329, Vancouver, Canada, July 2006.

[28] E. Castillo, B. Guijarro-Berdiñas, O. Fontenla-Romero, and A. Alonso-Betanzos, "A very fast learning method for neural networks based on sensitivity analysis," *Journal of Machine Learning Research*, vol. 7, pp. 1159–1182, 2006.