

A Neural Network MLSE Receiver Based on Natural Gradient Descent: Application to Satellite Communications

Mohamed Ibnkahla

*Electrical and Computer Engineering Department, Queen's University, Kingston, Ontario, Canada K7L 3N6
Email: mohamed.ibnkahla@ece.queensu.ca*

Jun Yuan

*Electrical and Computer Engineering Department, Queen's University Kingston, Ontario, Canada K7L 3N6
Email: steveyuan@comm.utoronto.ca*

Received 30 August 2003; Revised 12 February 2004

The paper proposes a maximum likelihood sequence estimator (MLSE) receiver for satellite communications. The satellite channel model is composed of a nonlinear traveling wave tube (TWT) amplifier followed by a multipath propagation channel. The receiver is composed of a neural network channel estimator (NNCE) and a Viterbi detector. The natural gradient (NG) descent is used for training. Computer simulations show that the performance of our receiver is close to the ideal MLSE receiver in which the channel is perfectly known.

Keywords and phrases: neural networks, satellite communications, high-power amplifiers.

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. Spectral efficiency demonstrates the ability of a system (e.g., modulation scheme) to accommodate data within an allocated bandwidth. 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 multilevel modulation schemes are employed, such as M-QAM ($M > 4$) modulations [6, 9, 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]. Moreover, the propagation channel causes frequency-selective multipath fading which generates intersymbol interferences (ISI). This again limits the transmission rates of existing satellite mobile systems [7, 9].

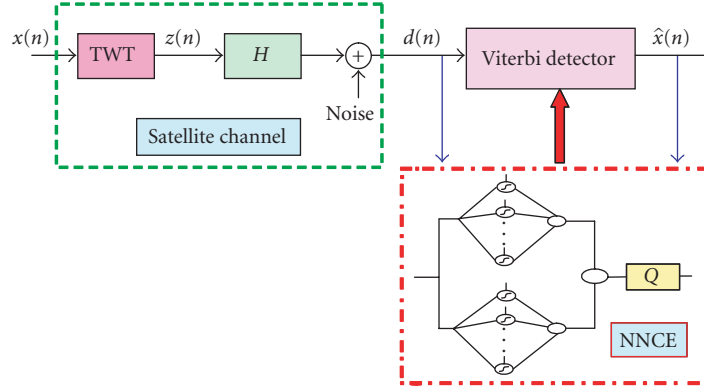


FIGURE 1: Satellite channel and MLSE receiver.

To improve power and spectral efficiencies, researchers have proposed different techniques at both transmitter and receiver sides [1, 3, 4, 9, 10, 11, 12, 13].

This paper proposes an MLSE receiver for M-QAM satellite channels equipped with TWT amplifiers. The receiver is composed of a neural network channel estimator (NNCE) and a Viterbi detector. The NNCE is trained using natural gradient (NG) descent [14, 15].

Our receiver is shown to outperform the fully connected multilayer neural network equalizer, the LMS combined with a memoryless neural network equalizer, and the LMS equalizer. Computer simulations show that it performs close to the ideal MLSE (IMLSE) receiver (which assumes perfect channel knowledge).

In the following section, we describe the system model and derive the learning algorithm. In Section 3, we present simulation results and illustrations.

2. SYSTEM MODEL

2.1. Satellite channel model

The satellite channel model [1, 6, 9] is composed of an on-board traveling wave tube (TWT) amplifier, followed by a propagation channel which is modeled by an FIR filter H (Figure 1). The transmitted signal $x(n) = r(n)e^{j\phi(n)}$ is M-QAM modulated.

The TWT amplifier behaves as a memoryless nonlinearity which affects the input signal amplitude. Its output can then be expressed as

$$z(n) = A(r(n)) \exp j\{P(r(n)) + \phi(n)\}, \quad (1)$$

where $A(\cdot)$ and $P(\cdot)$ are the TWT amplitude conversion (AM/AM) and phase conversion (AM/PM), respectively. These nonlinear conversions, which are assumed to be unknown to the receiver, have been modeled in this paper as

$$\begin{aligned} A(r) &= \frac{\alpha_a r}{1 + \beta_a r^2}, \\ P(r) &= \frac{\alpha_p r^2}{1 + \beta_p r^2}, \end{aligned} \quad (2)$$

where $\alpha_a = 2$, $\beta_a = 1$, $\alpha_p = 4$, $\beta_p = 9$. This represents a typical TWT model used in satellite communications [9].

The TWT amplifier gain is defined as $G(r) = A(r)/r$. The TWT backoff (BO) is defined as the ratio (in dB) between the signal power at the TWT saturation point and the input signal power: $\text{BO} = 10 \log(P_{\text{sat}}/P_{\text{in}})$. The TWT behaves as a hard nonlinearity when the BO is low, and as a soft nonlinearity when the BO is high.

Filter H output is given by $d_0(n) = H^t Z(n)$, where $H = [h_0, h_1, \dots, h_{N_H-1}]^t$, and $Z(n) = [z(n), z(n-1), \dots, z(n-N_H+1)]^t$ (where the superscript “ t ” denotes the transpose).

Finally, the channel output can be written as $d(n) = d_0(n) + n_0(n)$, where $n_0(n)$ is a zero-mean white Gaussian noise.

The MLSE receiver is composed of an NNCE and an MLSE detector. The NNCE performs an on-line estimation of the satellite channel. The estimated channel is provided to the MLSE detector (Figure 1), which gives an estimation of the transmitted symbol using a Viterbi detector [9].

2.2. Neural network channel estimator

The NNCE is composed of a memoryless neural network followed by an adaptive linear filter Q (Figures 1 and 2). The NN aims at identifying the TWT transfer function; while the adaptive filter Q aims at identifying the linear part of the system (i.e., filter H).

The memoryless NN consists of two subnetworks called NNG and NNP (Figure 2), each has M (real-valued) neurons in the first layer and a scalar output. NNG aims at identifying the amplifier gain, while NNP aims at identifying the phase conversion. Therefore, by using this structure, we aim at obtaining direct estimation of the amplitude and phase nonlinearities.

The filter-memoryless neural network structure has been shown to outperform fully connected complex-valued multilayer neural network with memory when applied to satellite channel identification (see, e.g., [12, 16]).

The two subnetworks have the same input which is the amplitude of the transmitted symbol, (i.e., $r(n) = |x(n)|$), in

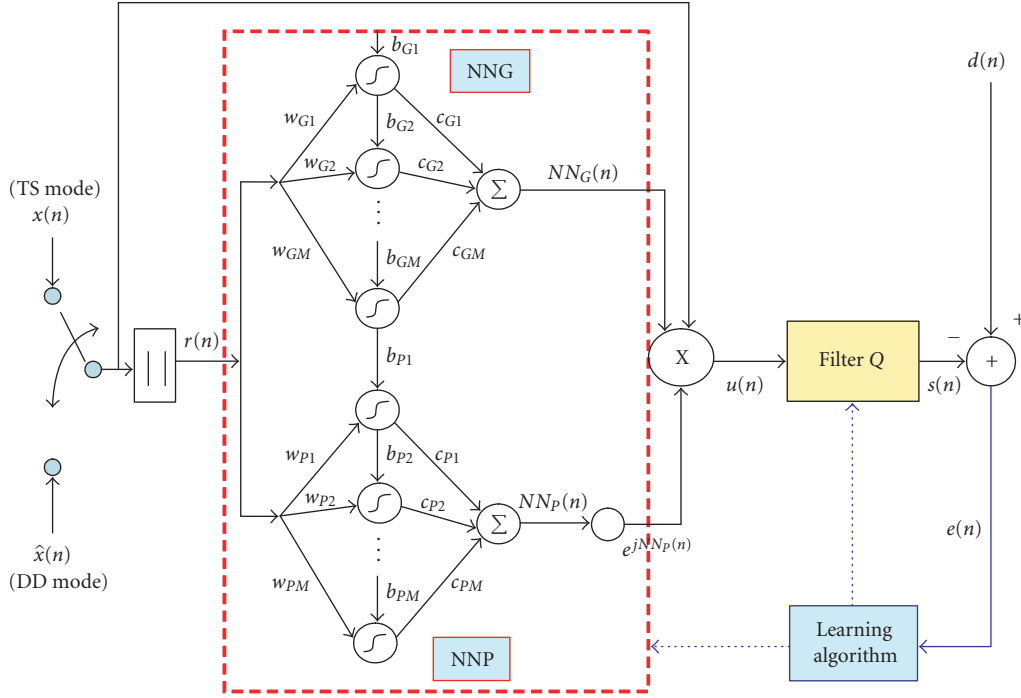


FIGURE 2: Neural network channel estimator (NNCE).

the case of training sequence (TS) mode; or the amplitude of the detected symbol (i.e., $\hat{r}(n) = |\hat{x}(n)|$), in the case of decision-directed (DD) mode.

In this paper, we derive the algorithm for the TS mode (for the DD mode, $\hat{x}(n)$ should be used as input).

The output of the neural network is expressed as

$$u(n) = x(n)NN_G(r(n))e^{jNN_P(r(n))}, \quad (3)$$

where

$$\begin{aligned} NN_G(r(n)) &= \sum_{i=1}^M c_{gi} f(w_{gi}r(n) + b_{gi}) \quad (\text{NNG output}), \\ NN_P(r(n)) &= \sum_{i=1}^M c_{pi} f(w_{pi}r(n) + b_{pi}) \quad (\text{NNP output}), \end{aligned} \quad (4)$$

where $f(\cdot)$ is the activation function which is taken here as the hyperbolic tangent function, w_{gi} , c_{gi} , b_{gi} (resp., w_{pi} , c_{pi} , b_{pi}) are the weights of subnetwork NNG (resp., NNP).

The adaptive FIR filter $Q = [q_0, q_1, \dots, q_{N_Q-1}]^t$, where N_Q is the size of filter Q . Finally, the output of Q is given by

$$s(n) = Q^t U(n), \quad (5)$$

where

$$U(n) = [u(n), u(n-1), \dots, u(n-N_Q+1)]^t. \quad (6)$$

The system parameter vector will be denoted by θ , which includes all parameters to be updated, that is, subnetwork NNG, subnetwork NNP, and filter Q weights:

$$\theta = [w_{g1}, \dots, w_{gM}, b_{g1}, \dots, b_{gM}, c_{g1}, \dots, c_{gM}, w_{p1}, \dots, w_{pM}, b_{p1}, \dots, b_{pM}, c_{p1}, \dots, c_{pM}, q_0, \dots, q_{N_Q-1}]^t. \quad (7)$$

2.3. Learning algorithm

The neural network is used to identify the channel by supervised learning. At each iteration, a pair of channel input¹-channel output signals is presented to the neural network. The NN parameters are then updated in order to minimize the squared error $J(n)$ between the channel output and the neural network output:

$$J(n) = \frac{1}{2} \|e(n)\|^2 = \frac{1}{2} [e_R^2(n) + e_I^2(n)], \quad (8)$$

where

$$e(n) = d(n) - s(n) = e_R(n) + je_I(n). \quad (9)$$

¹In the derivation of the algorithm we assume that a training input set is available (TS mode), this is the case for example of GSM frames where a number of known bits are used for supervised learning. If this set is not available, then the estimated symbol at the MLSE receiver output is used for training (DD mode).

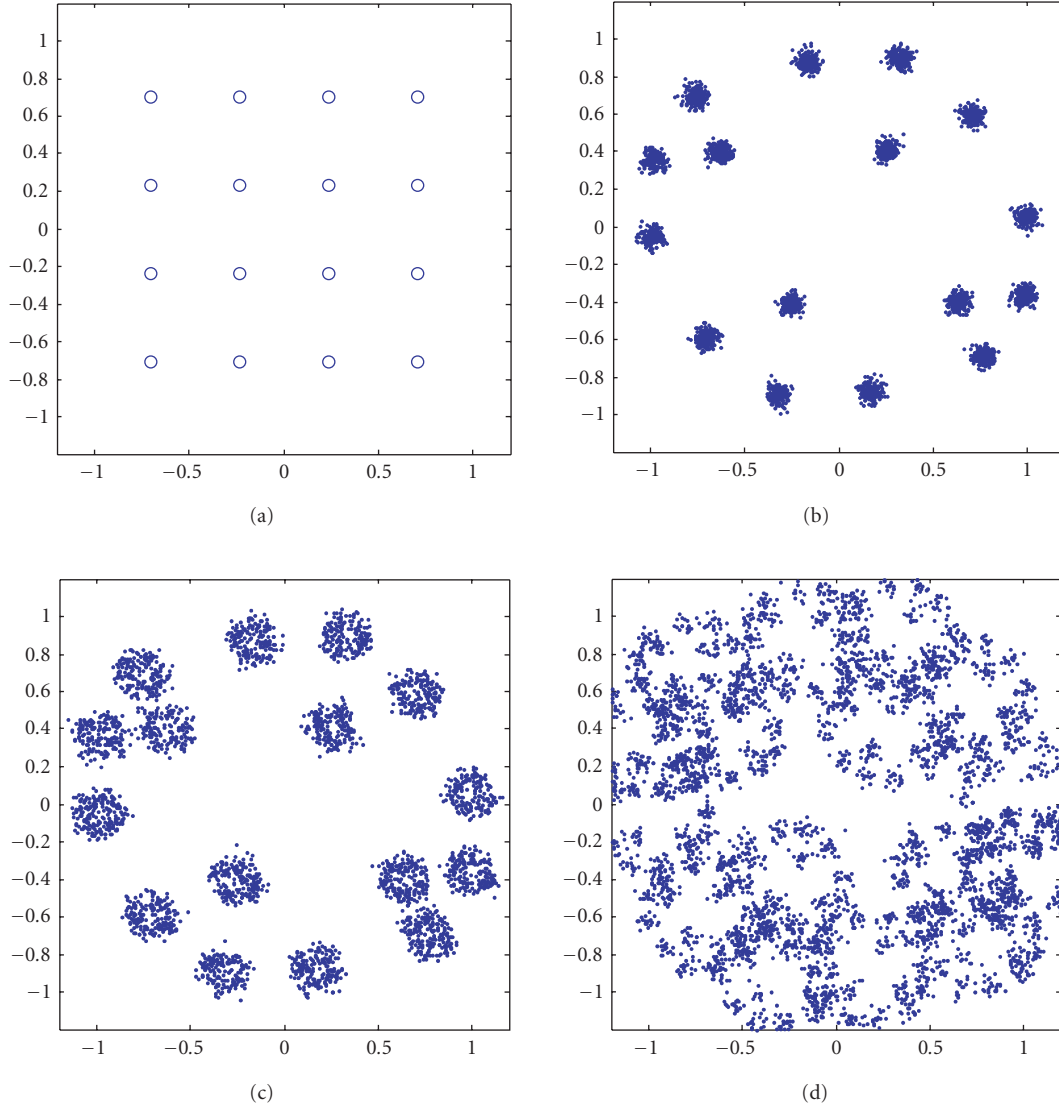


FIGURE 3: (a) Transmitted 16-QAM constellation. (b) Signal constellation at the channel output ($H = 1$, BO = 2.55 dB). (c) Signal constellation at the channel output ($H = [1 \ 0.1]^t$, BO = 2.55 dB). (d) Signal constellation at the channel output ($H = [1 \ 0.3]^t$, BO = 2.55 dB).

Indexes R and I refer to the real and imaginary parts, respectively.

We use a gradient descent algorithm to minimize this cost function. The ordinary gradient is the steepest descent direction of a cost function if the space of parameters is an orthonormal coordinate system. It has been shown [14] that, in the case of multilayer neural nets, the steepest descent direction (or the NG) of the loss function is actually given by $-\tilde{\nabla}_{\theta(n)}J(n) = -G^{-1}\nabla_{\theta(n)}J(n)$, where G^{-1} is the inverse of the Fisher information matrix (FIM), $G^{-1} = [g_{i,j}]^{-1}$, $g_{i,j} = E[(\partial J(n)/\partial \theta_i(n))(\partial J(n)/\partial \theta_j(n))]$.

Therefore, the neural network weights will be updated as follows:

$$\theta(n+1) = \theta(n) - \mu \tilde{\nabla}_{\theta(n)}J(n), \quad (10)$$

where μ is a small positive constant, and

$$\tilde{\nabla}_{\theta(n)}J(n) = G^{-1}(n)\nabla_{\theta(n)}J(n), \quad (11)$$

where $\nabla_{\theta(n)}J(n) = e_R(n)\nabla_{\theta(n)}e_R(n) + e_I(n)\nabla_{\theta(n)}e_I(n)$ represents the ordinary gradient of $J(n)$ with respect to θ (see the appendix).

Note that the classical (ordinary gradient descent) back-propagation (BP) [17] algorithm corresponds to the case where G equals the identity matrix.

The calculation of the expectation in the expression of G requires the probability distribution of the input $x(n)$, which is unknown in most cases. Moreover, the inversion of G is computationally costly when the number of neurons is large.

To obtain directly G^{-1} , we use a Kalman filter technique [15]:

$$\hat{G}^{-1}(n+1) = \frac{1}{(1-\varepsilon_n)}\hat{G}^{-1}(n) - \frac{\varepsilon_n}{(1-\varepsilon_n)} \times \frac{\hat{G}^{-1}(n)\nabla_{\theta(n)}s(n)(\nabla_{\theta(n)}s(n))^t\hat{G}^{-1}(n)}{(1-\varepsilon_n) + \varepsilon_n(\nabla_{\theta(n)}s(n))^t\hat{G}^{-1}(n)\nabla_{\theta(n)}s(n)}, \quad (12)$$

where $\nabla_{\theta(n)}s(n)$ is the ordinary gradient of $s(n)$ with respect to vector $\theta(n)$.

This equation involves an updating rate ε_n . When ε_n is small, this equation can be approximated by

$$\hat{G}^{-1}(n+1) = (1+\varepsilon_n)\hat{G}^{-1}(n) - \varepsilon_n\hat{G}^{-1}(n)\nabla_{\theta(n)}s(n)(\nabla_{\theta(n)}s(n))^t\hat{G}^{-1}(n). \quad (13)$$

A search-and-converge schedule will be used for ε_n in order to obtain a good tradeoff between convergence speed and stability:

$$\varepsilon_n = \frac{\varepsilon_0 + c_\varepsilon n/\tau}{1 + c_\varepsilon n/\tau\varepsilon_0 + n^2/\tau}, \quad (14)$$

such that small n corresponds to a “search” phase (ε_n is close to ε_0), and large n corresponds to a “converge” phase (ε_n is equivalent to c_ε/n for large n). ε_0 , c_ε , and τ are positive real constants. As can be seen in these equations, the NG descent is applied to the adaptive filter Q and to the subnetworks, since vector θ includes all adaptive parameters.

Interesting discussions on the use of the NG descent for adaptive filtering and system inversion can be found in [18, 19].

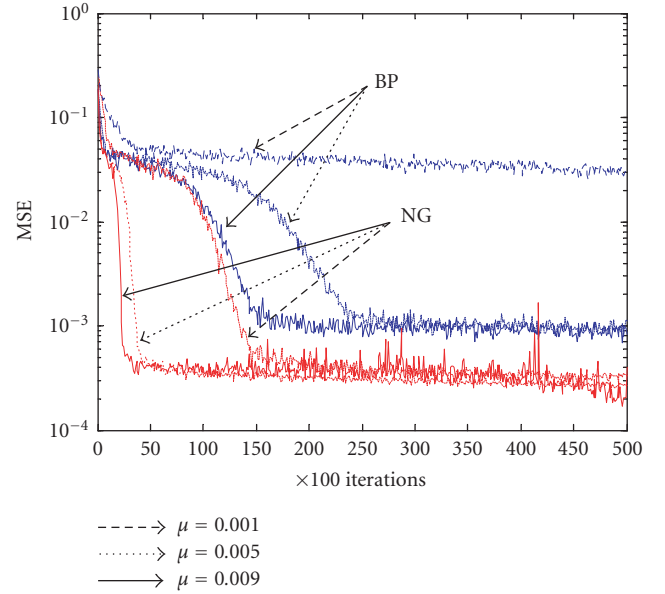
3. SIMULATION RESULTS AND DISCUSSIONS

This section presents computer simulations to illustrate the performance of the adaptive NN MLSE receiver. The transmitted signal was 16-QAM modulated. The amplifier BO was fixed to 2.55 dB. Figure 3 illustrates the effect of the satellite channel on the rectangular 16-QAM transmitted constellation. The transmitted constellation is illustrated in Figure 3a. Figure 3b shows the output constellation when filter $H = 1$, that is, the signal is affected only by the TWT nonlinearity and additive noise. It can be seen that the constellation is rotated because of the phase conversion, and the symbols are closer to each other because of the amplitude nonlinearity.

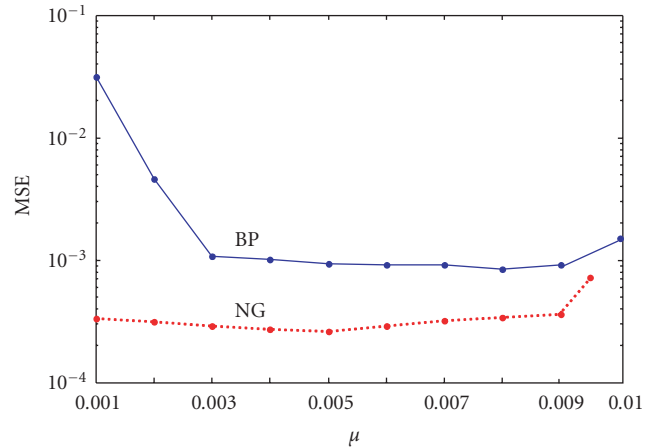
Figure 3c shows the output signal constellation when $H = [1 \ 0.1]^t$. ISI interferences (caused by the 0.1 reflected path) are illustrated by larger and overlapping clouds. Finally, Figure 3d shows the case where $H = [1 \ 0.3]^t$. The constellation is highly distorted.

In all these cases, an efficient receiver is needed to overcome the problems of nonlinearity and ISI.

In the simulations below, the unknown propagation channel was assumed to have two paths: $H = [1 \ 0.3]^t$ (corresponding to the case of a frequency-selective slow fading channel).



(a)



(b)

FIGURE 4: (a) Learning curves of BP and NG with different μ ($H = [1 \ 0.3]^t$, BO = 2.55 dB). (b) MSE versus μ .

The following parameters have been taken for the NG algorithm: $\varepsilon_0 = 0.005$, $c_\varepsilon = 1$, and $\tau = 70,000$. Each subnetwork was composed of $M = 5$ neurons. We have taken this number of neurons because a lower number decreases the performance and a higher one does not significantly improve the system performance. Viterbi decoding block contained $N_1 = 1$ training symbol and $N_2 = 9$ information symbols. The receiver was trained using a TS of 3000 transmitted symbols, after which the decision-directed mode was activated.

Figure 4a shows the learning curves of the NG and BP for different values of μ (the same initial weight values have

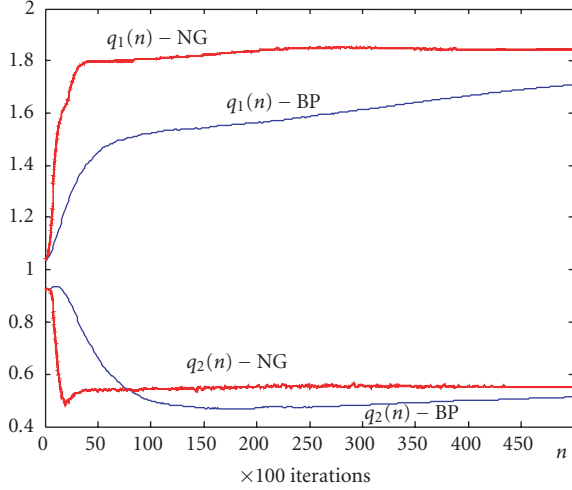


FIGURE 5: Evolution of adaptive filter Q weights (comparison between BP and NG), $\mu = 0.005$.

been taken for the two algorithms). It can be seen that the NG has better capabilities to escape from the plateau regions. It yields faster convergence speed and lower MSE than the BP algorithm. In Figure 4b, the MSE performance of each algorithm (obtained after 50,000 iterations in TS mode) is shown versus the learning rate μ . Note that for very small μ , the BP MSE is very high, which suggests that the algorithm could not escape from the plateau region. For high μ , the BP and NG MSEs increase, but the NG becomes quickly unstable (e.g., for $\mu = 0.01$).

In what follows, we will choose $\mu = 0.005$, which represents a good tradeoff between convergence speed and MSE for the two algorithms.

Figures 5 and 6 show that the different parts of the channel have been successfully identified: the linear filter (Figure 5), the TWT AM/AM conversion (Figure 6a), and the TWT AM/PM conversion (Figure 6b). Note that, concerning the identification of the channel filter by Q , the latter has converged to a scaled version of H . The scale factor is equal to 1.84 (resp., 1.71) for the NG algorithm (resp., BP algorithm). This scalar factor is compensated by the subnetwork NNG which controls the gain. In [16, 20], the convergence properties of adaptive identification of nonlinear systems are presented (for the ordinary gradient descent learning). Several structures are studied and it is shown, in particular, how the scale factor is distributed among the different parts of the adaptive system.

The NG algorithm yielded better AM/AM and AM/PM approximation than the BP algorithm. This is because the NG algorithm has better capabilities to quickly escape from plateau regions in the error surface [14]. It is worth to note that, since we used 16-QAM modulation, the TWT characteristics are expected to be better approximated around the 3 possible amplitudes of the 16-QAM constellation, as shown in Figure 6.

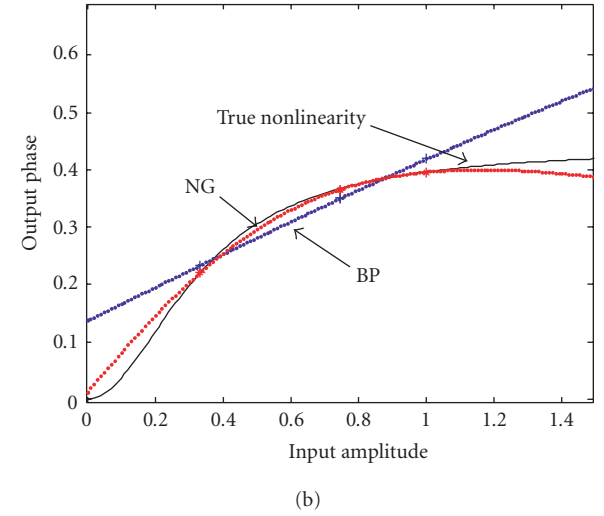
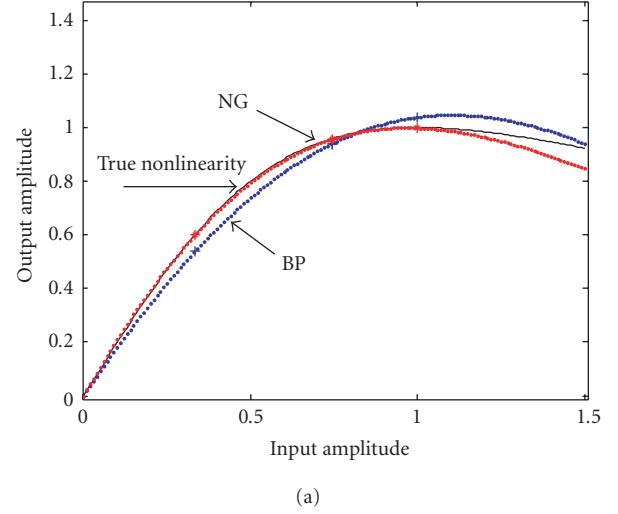


FIGURE 6: (a) TWT AM/AM characteristic. True curve and normalized neural network models, (+) and (*) represent the three 16-QAM amplitudes and their corresponding outputs for BP and NG, respectively. (b) TWT AM/PM characteristic. True curve and normalized neural network models, (+) and (*) represent the three 16-QAM amplitudes and their corresponding outputs for BP and NG, respectively.

The MLSE receiver has been compared to three equalizers which have been proposed previously in the literature. These are as follows.

(1) An LMS equalizer [21] composed of a tapped delay line (with 10 weights). The input to the LMS filter is

$$D(n) = [d(n) \ d(n-1) \ \cdots \ d(n-L+1)]^t, \quad L = 10. \quad (15)$$

The purpose of the LMS filter is to cancel out the ISI, but it is not able to mitigate the nonlinear effects of the HPA.

(2) A fully connected multilayer NN equalizer with memory trained with BP [12, 17] (Figure 7a). The input is $D(n)$.

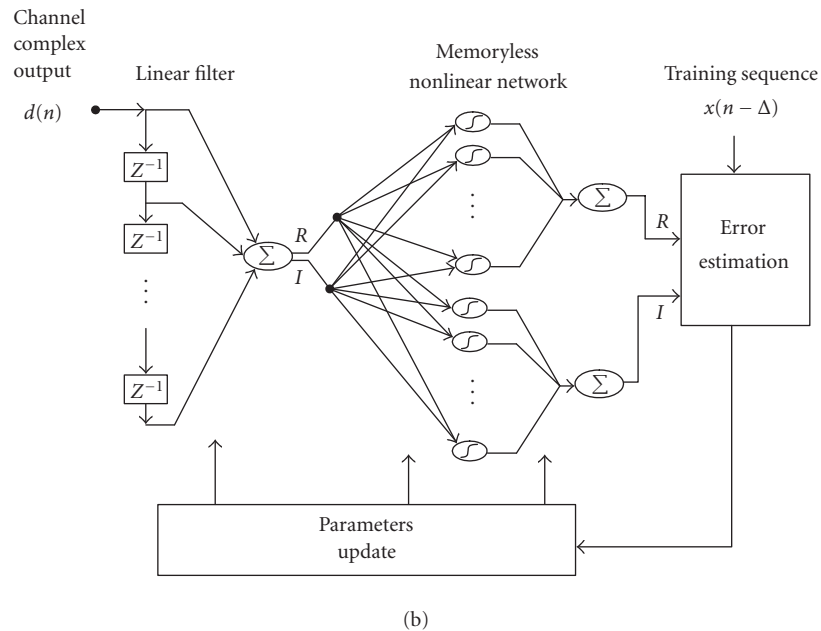
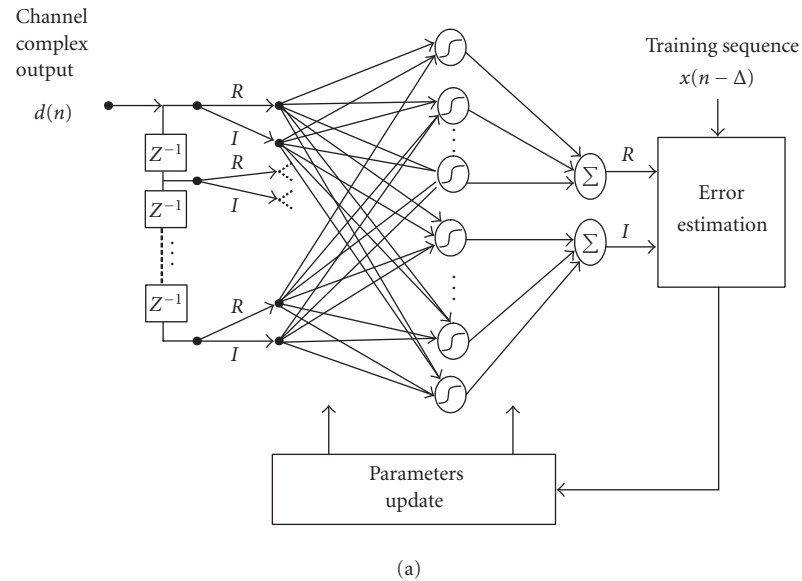


FIGURE 7: (a) Fully connected NN equalizer structure. (b) Filter-memoryless NN equalizer structure.

This input is connected to 10 neurons in the hidden layer (5 for the real part and 5 for the imaginary part). The output neuron is linear and complex valued. The fully connected NN aims at simultaneously mitigating both ISI and HPA nonlinear effects. This equalizer was trained by the BP algorithm.

(3) An LMS filter combined with a memoryless neural network (LMS-NN) equalizer (Figure 7b) [12, 17]. The LMS-NN equalizer is composed of a linear filter Q' (with 10 weights) followed by a two-layer memoryless neural network, with 5 neurons in the real (R) part and 5 neurons in

the imaginary (I) part, and a complex-valued output. The purpose of this adaptive filter-NN scheme is to cancel the ISI by the linear filter, and to mitigate the nonlinearities by the memoryless NN [12, 22]. These two tasks are split into the filter and the memoryless NN, respectively. This kind of NN equalizer has been shown to outperform classical nonlinear equalizers, such as Volterra series equalizers [9, 12]. Two algorithms have been used to train this equalizer: the NG algorithm and the BP algorithm. A comparative study of these two training algorithms for channel inversion can be found in [18].

TABLE 1: Performance comparison between the different receivers, $H = [1 \ 0.3]^t$, $BO = 2.55$ dB (see Figure 8).

Structure	IMLSE	NG MLSE	BP MLSE	NG LMS-NN	BP LMS-NN	Full-NN equalizer	LMS equalizer
SNR needed to reach 10^{-4} BER (dB)	25	25.2	28.2	31	31.5	31.7	33.1
NG MLSE gain in SNR with respect to other techniques (dB)	-0.2	—	3	5.8	6.3	6.5	7.8

References [12, 17] present extensive analysis and comparisons between the above equalizers and other NN-based equalizers, such as radial basis function (RBF) equalizers and self-organizing map (SOM) equalizers. The reader can find in references [22, 23] other complex-valued neural networks that have been successfully used for adaptive channel equalization.

The chosen number of neurons and size of filters gave a good tradeoff between computational complexity and performance (i.e., larger sizes did not improve the equalizers performances).

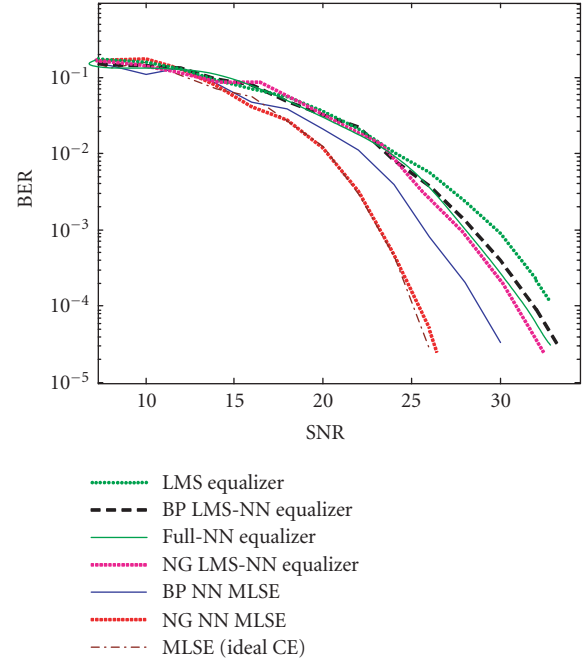
To ensure a good comparison between the different algorithms, the same learning rate ($\mu = 0.005$) has been used for the three equalizers. However, the performance evaluation has been made after final convergence of the different algorithms (i.e., when the values of the weights as well as the output MSE reach a steady state).

It should be noted that, since the criteria in training the above equalizers is minimizing the MSE error between the output sequence and the desired output, it is expected that these equalizers will have a lower performance than the MLSE receiver (which maximizes the likelihood of correct detection). We have also compared the results to the IMLSE receiver in which the channel is assumed to be perfectly known. The performance of our NN MLSE receiver is close to that of the IMLSE. This is justified since the different parts of the channel have been correctly identified, in particular at the 16-QAM constellation points (Figures 5–6).

Our NN MLSE receiver trained by the NG algorithm outperforms the other receivers (Figure 8) in terms of bit error rate (BER).

Table 1 shows the different SNR gains of our NG MLSE receiver over the other receivers, when $H = [1 \ 0.3]^t$ and $BO = 2.55$ dB, for a BER of 10^{-4} .

It is worth to note that the LMS-NN structure trained with NG allows a gain of 0.5 dB over the same structure trained with BP. This is because the NG allows the algorithm to quicker escape from the plateau regions in the MSE surface, yielding better inversion of the channel. On the other hand, the LMS-NN structure performs slightly better than the fully connected NN (when they are both trained with BP), with an important advantage that its computational complexity is much lower than the fully connected NN. This is due to the fact that the ISI (caused by the propagation channel with memory) and the HPA nonlinear distortions

FIGURE 8: BER versus SNR. Comparison between different receivers, $H = [1 \ 0.3]^t$, $BO = 2.55$ dB.

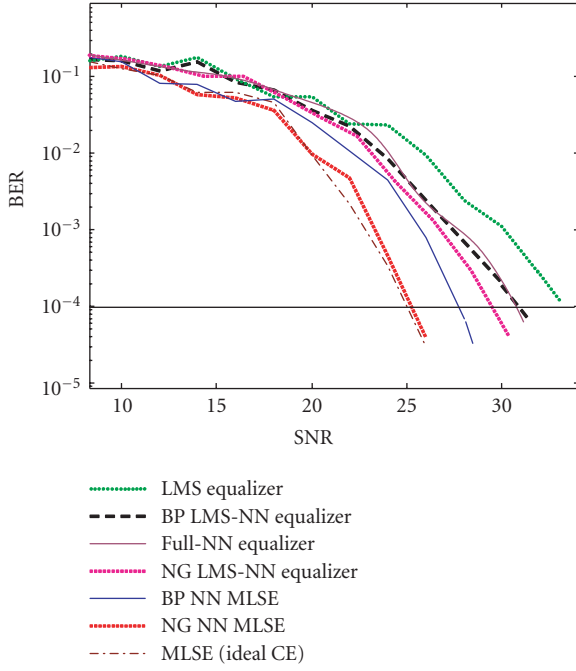
come from two physically separated sources. The LMS-NN tries to mitigate each of them by two separated tools (LMS filter to mitigate ISI and memoryless NN to invert the nonlinearity). The fully connected NN deals with these two problems as a whole and yields a multidimensional function with memory to reduce both ISI and nonlinear distortions. See [12, 18] for useful discussions about these structures.

Figure 9 shows the BER performance when $H = [1 \ 0.1]^t$ ($BO = 2.55$ dB). Here, the performance of the NG MLSE is close to the IMLSE. Table 2 shows the different SNR gains of our NG MLSE receiver over the other receivers, where $H = [1 \ 0.1]^t$ and $BO = 2.55$ dB, for a BER of 10^{-4} .

Note that the performance of the NG MLSE for this case is close to the case where there are higher interferences ($H = [1 \ 0.3]^t$, Figure 8), this is justified by the fact that the different parts of the channel have been well estimated,

TABLE 2: Performance comparison between the different receivers, $H = [1 \ 0.1]^t$, BO = 2.55 dB (see Figure 9).

Structure	IMLSE	NG MLSE	BP MLSE	NG LMS-NN	BP LMS-NN	Full-NN equalizer	LMS equalizer
SNR needed to reach 10^{-4} BER	25	25	27.5	29.5	31	31	33
NG MLSE gain in SNR with respect to other techniques	0	—	2.5	4.5	6	6	8.1

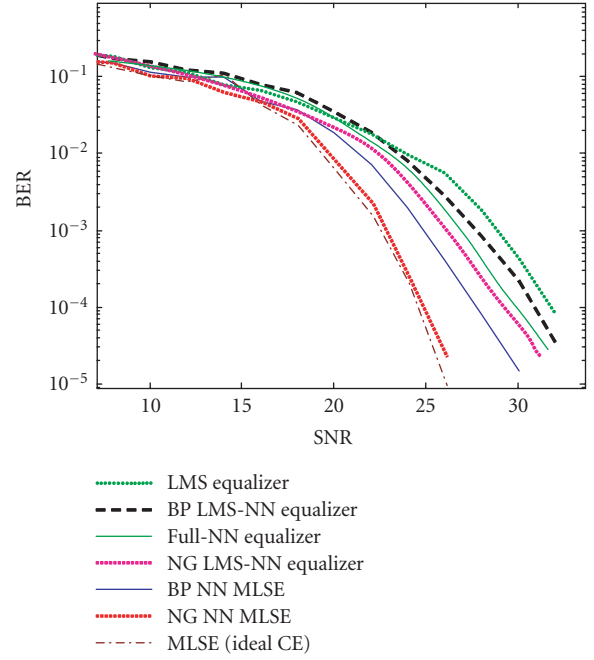
FIGURE 9: BER versus SNR. Comparison between different receivers, $H = [1 \ 0.1]^t$, BO = 2.55 dB.

regardless of the amount of interferences. Note that the performances of the BP MLSE and the equalizers degrade as the amount of interferences increases. For the BP MLSE, this is due to the fact that it is not able to give a very accurate approximation of the propagation channel. For the different equalizers, the degradation in performance is due to the fact that the increase in ISI makes it difficult to invert the channel, especially in the presence of the nonlinearity.

Finally, Figure 10 shows the BER results when the nonlinearity BO is reduced to 3 dB and the propagation channel is kept to $H = [1 \ 0.3]^t$. We notice that the BER performances of the different receivers are improved compared to Figure 8. This is because the amount of nonlinear distortions has been reduced.

4. CONCLUSION

In this paper we have proposed an adaptive MLSE receiver based on an NNCE and a Viterbi detector. This structure

FIGURE 10: BER versus SNR. Comparison between different receivers, $H = [1 \ 0.3]^t$, BO = 3 dB.

was applied to 16-QAM transmission over nonlinear satellite channels with memory. The NG descent has been used to update the neural network weights.

The proposed algorithm was shown to outperform the BP algorithm and classical equalizers such as the multi-layer neural network and the LMS equalizers. Simulation results have shown that the BER performance of our receiver is close to that of an IMLSE receiver in which the channel is perfectly known.

APPENDIX

COMPUTATION OF THE GRADIENTS

We substitute (5) in (9) to express the output error as function of the NN output, and therefore as function of the different weights (i.e., vector θ). The gradients are calculated by taking the derivatives of $e_R(n)$ (resp., $e_I(n)$) (5) with respect to each of the components of vector θ .

$$\begin{aligned}
\nabla_{\theta} e_R(n) = & \left[\begin{array}{c} \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{G1} f'(w_{G1}r(n-k) + b_{G1}) \\ \vdots \\ \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{GM} f'(w_{GM}r(n-k) + b_{GM}) \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{G1} f'(w_{G1}r(n-k) + b_{G1}) \\ \vdots \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{GM} f'(w_{GM}r(n-k) + b_{GM}) \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) f(w_{G1}r(n-k) + b_{G1}) \\ \vdots \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) f(w_{GM}r(n-k) + b_{GM}) \\ - \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{P1} f'(w_{P1}r(n-k) + b_{P1}) \\ \vdots \\ - \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{PM} f'(w_{PM}r(n-k) + b_{PM}) \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{P1} f'(w_{11}r(n-k) + b_{P1}) \\ \vdots \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{PM} f'(w_{PM}r(n-k) + b_{PM}) \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) f(w_{P1}r(n-k) + b_{P1}) \\ \vdots \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) f(w_{PM}r(n-k) + b_{PM}) \\ u_R(n) \\ \vdots \\ u_R(n - N_Q + 1) \end{array} \right] \quad (A.1)
\end{aligned}$$

$$\begin{aligned}
\nabla_{\theta} e_I(n) = & \begin{bmatrix} \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{G1} f'(w_{G1}r(n-k) + b_{G1}) \\ \vdots \\ \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{GM} f'(w_{GM}r(n-k) + b_{GM}) \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{G1} f'(w_{G1}r(n-k) + b_{G1}) \\ \vdots \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) c_{GM} f'(w_{GM}r(n-k) + b_{GM}) \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) f(w_{G1}r(n-k) + b_{G1}) \\ \vdots \\ \sum_{k=0}^{N_Q-1} q_k r(n-k) \sin(NN_P(r(n-k)) + \phi(n)) f(w_{GM}r(n-k) + b_{GM}) \\ - \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{P1} f'(w_{P1}r(n-k) + b_{P1}) \\ \vdots \\ - \sum_{k=0}^{N_Q-1} q_k r^2(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{PM} f'(w_{PM}r(n-k) + b_{PM}) \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{P1} f'(w_{P1}r(n-k) + b_{P1}) \\ \vdots \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) c_{PM} f'(w_{PM}r(n-k) + b_{PM}) \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) f(w_{P1}r(n-k) + b_{P1}) \\ \vdots \\ - \sum_{k=0}^{N_Q-1} q_k r(n-k) \cos(NN_P(r(n-k)) + \phi(n)) f(w_{PM}r(n-k) + b_{PM}) \\ u_I(n) \\ \vdots \\ u_I(n - N_Q + 1) \end{bmatrix}. \quad (\text{A.2})
\end{aligned}$$

REFERENCES

- [1] M. Ibnkahla, Q. M. Rahman, A. I. Sulyman, H. A. Al-Asady, J. Yuan, and A. Safwat, "High-speed satellite mobile communications: technologies and challenges," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 312–339, 2004, Special issue on Gigabit wireless communications. Technologies and challenges.
- [2] A. Jamalipour, "Broadband satellite networks—the global IT bridge," *Proceedings of the IEEE*, vol. 89, no. 1, pp. 88–104, 2001.
- [3] D. Boudreau, G. Caire, G. E. Corazza, et al., "Wide-band CDMA for the UMTS/IMT-2000 satellite component," *IEEE Trans. Vehicular Technology*, vol. 51, no. 2, pp. 306–331, 2002.

- [4] L. Cimini Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Communications*, vol. 33, no. 7, pp. 665–675, 1985.
- [5] P. Chitre and F. Yegenoglu, "Next-generation satellite networks: architectures and implementations," *IEEE Communications Magazine*, vol. 37, no. 3, pp. 30–36, 1999.
- [6] G. Maral and M. Bousquet, *Satellite Communication Systems*, John Wiley & Sons, New York, NY, USA, 1996.
- [7] G. E. Corazza, R. Pedone, and A. Vanelli-Coralli, "Mobile satellite channels: statistical models and performance analysis," in *Signal Processing for Mobile Communications Handbook*, M. Ibnkahla, Ed., chapter 4, pp. 4.1–4.35, CRC Press, Boca Raton, Fla, USA, 2004.
- [8] F. J. Dietrich, P. Metzen, and P. Monte, "The Globalstar cellular satellite system," *IEEE Trans. Antennas and Propagation*, vol. 46, no. 6, pp. 935–942, 1998.
- [9] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*, Kluwer Academic, Boston, Mass, USA, 1999.
- [10] A. N. D'Andrea, V. Lottici, and R. Reggiannini, "RF power amplifier linearization through amplitude and phase predistortion," *IEEE Trans. Communications*, vol. 44, no. 11, pp. 1477–1484, 1996.
- [11] E. Costa and S. Pupolin, "M-QAM-OFDM system performance in the presence of a nonlinear amplifier and phase noise," *IEEE Trans. Communications*, vol. 50, no. 3, pp. 462–472, 2002.
- [12] M. Ibnkahla, "Applications of neural networks to digital communications: a survey," *Signal Processing*, vol. 80, no. 7, pp. 1185–1215, 2000.
- [13] U. Vilaipornsawai and M. R. Soleymani, "Trellis-based iterative decoding of block codes for satellite ATM," in *Proc. IEEE International Conference on Communications (ICC '02)*, vol. 5, pp. 2947–2951, New York, NY, USA, April–May 2002.
- [14] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.
- [15] S.-I. Amari, H. Park, and K. Fukumizu, "Adaptive method of realizing natural gradient learning for multilayer perceptrons," *Neural Computation*, vol. 12, no. 6, pp. 1399–1409, 2000.
- [16] M. Ibnkahla, N. J. Bershad, J. Sombrin, and F. Castanié, "Neural network modeling and identification of nonlinear channels with memory: algorithms, applications, and analytic models," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1208–1220, 1998.
- [17] S. Haykin, *Neural Networks: A Comprehensive Foundation*, IEEE Press, New York, NY, USA, 1997.
- [18] M. Ibnkahla, "Natural gradient learning neural networks for adaptive inversion of Hammerstein systems," *IEEE Signal Processing Letters*, vol. 9, no. 10, pp. 315–317, 2002.
- [19] L.-Q. Zhang, S. Amari, and A. Cichocki, "Semiparametric model and superefficiency in blind deconvolution," *Signal Processing*, vol. 81, no. 12, pp. 2535–2553, 2001.
- [20] N. J. Bershad, P. Celka, and J.-M. Vesin, "Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for Gaussian data and noisy input and output measurements," *IEEE Trans. Signal Processing*, vol. 47, no. 3, pp. 675–689, 1999.
- [21] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.
- [22] Q. Gan, P. Saratchandran, N. Sundararajan, and K. R. Subramanian, "A complex valued radial basis function network for equalization of fast time varying channels," *IEEE Transactions on Neural Networks*, vol. 10, no. 4, pp. 958–960, 1999.
- [23] A. Uncini, L. Vecci, P. Campolucci, and F. Piazza, "Complex-valued neural networks with adaptive spline activation function for digital-radio-links nonlinear equalization," *IEEE Trans. Signal Processing*, vol. 47, no. 2, pp. 505–514, 1999.

Mohamed Ibnkahla obtained an Engineering degree in electronics in 1992, an M.S. degree in signal and image processing in 1992 (first-class honors), a Ph.D. degree in signal processing in 1996 (first-class honors), and the Habilitation à Diriger des Recherches (HDR) in digital communications and signal processing in 1998, all from the National Polytechnic Institute of Toulouse (INPT), France. Dr. Ibnkahla has held an Assistant Professor position at INPT (1996–1996). In 2000, he has joined the Department of Electrical and Computer Engineering, Queen's University, Kingston, Canada, where he is now an Associate Professor. Dr. Ibnkahla is the Director of the Satellite and Mobile Communications Laboratory, Queen's University. He is the Editor of the *Signal Processing for Mobile Communications Handbook*, CRC Press, 2004. He has published 21 refereed journal papers and book chapters, and more than 60 conference papers. His research interests include cross-layer design, wireless communications, satellite communications, neural networks, and adaptive signal processing. Dr. Ibnkahla received the *INPT Leopold Escande Medal*, France, in 1997; the *Premier's Research Excellence Award (PREA)*, Ontario, Canada, in 2000; and the *Favorite Professor Award*, Department of Electrical and Computer Engineering, Queen's University, in 2004.



Jun Yuan received the B.S. degree in electrical engineering and applied mathematics from Shanghai Jiao Tong University, Shanghai, China, in 2001, and the M.S. degree in electrical and computer engineering from Queen's University, Kingston, Ontario, Canada, in 2003. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada. His research interests are in the areas of wireless communication, adaptive signal processing, and multiuser information theory.

