Neural Equalizer for Time Varying Channel Using Gauss-Newton Training Algorithm

Claudio J. C. Santos, O. Ludwig Jr., Pablo C. Gonzalez*, and Antonio C. de C. Lima, *Member IEEE*
Universidade Federal da Bahia, and Universidad Miguel Hernández*

Abstract—**Artificial Neural Networks (ANN) techniques have become very common as equalization solutions in several types of communication channels. These Neural Networks are presented in many topologies. The suitable choice of a topology for equalization purpose depends on different criteria such as: convergence rate, Bit Error Rate (BER), computational complexity, among many others. In this paper, it is investigated the behavior of a structure similar to a Decision Feedback Equalizer (DFE) employed to equalize time varying channels. The structure, a single recurrent perceptron, is based on a simplified Recurrent Neural Network (RNN). The Gauss-Newton algorithm has been used to estimate the synaptic weights of the perceptron during the training and testing phases. Despite the simplicity of implementation and low computational cost, it has been shown that the proposed topology presents some good comparative performance related to more complex structures based on Recurrent Neural Networks (RNN) and Multilayer Perceptrons (MLP) using Kalman Filters.**

Index Terms— **Equalization, Gauss-Newton, Time Varying Channels.**

I. Introduction

It is well known that in digital communications noise and Intersymbolic Interference (ISI) are the most common source of error in the detection process of symbols at the receiver. Typically the channel is dispersive and its frequency response is far from ideal. As a result of it, the transmitted pulses suffer shape distortion which affects its adjacent pulses, giving origin to ISI. Equalizers are used at the receiver in order to mitigate ISI effects caused by channel dispersion and multipath. Noise is frequently present in the received signal, and in general it has thermal origin, which is assumed to be Additive White Gaussian Noise (AWGN). AWGN effects can be minimized if a matched filter is implemented at the receiver and, in order to combat ISI effects, the receiver must also incorporate an equalizer. In practical applications, an optimized receiver is designed to deal with noise and channel's dispersive effects [1].

Equalization can be basically divided in two categories: Maximum Likelihood Sequence Estimation (MLSE) and filtering equalization. MLSE techniques use statistics from data sequences send through the channel to estimate their impulse response, computational complexity in MLSE methods increase exponential with channel dispersion. If the size of the symbol alphabet is M and the number of symbols which contributes to ISI is L, then the Viterbi's algorithm, for instance, involves a computational metric given by $M^{(L+1)}$ for each received symbol [2].

Filtering equalization methods use digital filters to compensate symbol distortions. The detector receives the symbols after demodulation and modifies them by an equalizer whose goal is mitigate ISI effects [2]. The filters employed in the equalization process can be either a linear filter such as Finite Impulsive Response (FIR) and Infinite Impulsive Response (IIR), or nonlinear filters in which a nonlinear function is added at the equalizer output, as shown in Fig. 1-b. Equalizers that use IIR topology with hard limit as nonlinearity in their output, are called Decision Feedback Equalizers (DFE) [3].

Due to their nonlinear characteristics, DFEs have a better performance in channels where spectral nulls are present [2]. DFE can be specified using three parameters: *m, n* and *d*, which correspond to, respectively, the number of coefficients of the direct filter, number of coefficients of the feedback filter, and output delay [4].

It is widely known that Artificial Neural Networks (ANN) perform very well in tasks involving function approximation and in nonlinear mapping [5]. It is also know they are very useful in pattern recognition. For these reasons, neural network have become very popular as a solution for equalization problems [6] and [7]. Others works have also shown that recurrent ANN such as RNN outperform Feed Forward Neural Networks FNN, such as MLP [8]. This can be credited to the ability of RNN to use previous detected symbols to improve the estimation of present ones. Simple topologies, built with a single layer comprising one or more percerptrons, have been tested and proved to outperform more complex structures such as Multilayer Perceptrons (MLP) and Recursive Multilayer Perceptrons (RMLP) [9][10].

II. System Modeling

This paper presents a DFE implemented with a hyperbolic tangent function as a nonlinear decision block during the learning or training step, as shown in Fig. 1a. This evolution is equivalent to a RNN with only one recurrent perceptron. Such architecture has been investigated by [9] and [11] in fixed channels equalization problems. The feedforward and feedback coefficients constitute the synaptic weights of the recurrent perceptron. The results presented in [9] seem to prove that the insertion of a soft decision function (hyperbolic tangent function) at the forward filter output, during the training phase, improves the DFE performance. While during the test phase that decision function is changed to a hardlimit or sign function (hard decision). Fig. 1 shows the schematic for both phases of the equalization process using this kind neural equalizer or modified DFE structure. Based on [9], one infers that simple ANN structures can perform signal equalization with low computational effort. In order to testing the equalization performance of this model, we propose to use three types of discrete time varying channel models each one with three coefficients. The DFE parameters employed in all simulations in this works are: m = 3, n = 2 and d = 2.

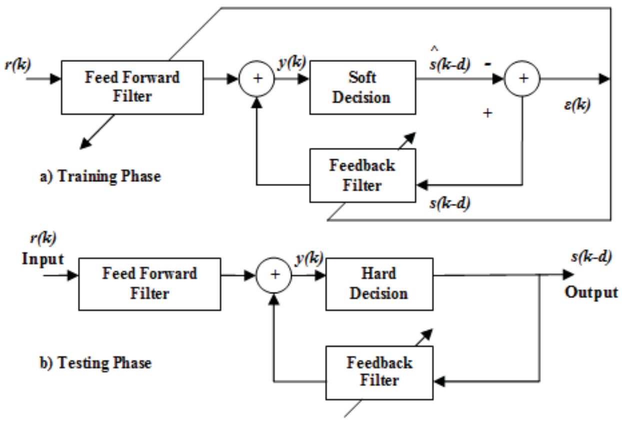

Fig. 1. DFE schematics: (a) training; (b) testing

The signal produced by an information source, s(k), is a complex baseband 4-QAM. The generated signal s(k) goes through a discrete channel and then it is contaminated by AWGN. After being deteriorated by the channel characteristics the signal at the receiver, r(k), is then feed to the equalizer. The input signal r(k) is convolved with the feed-forward filter coefficients and presented at its output, after the soft-decision, as a estimative now called u(k). This estimative is feedback and then convolved with the feedback filter coefficients. Thus, the signal at the input of the perceptron, or equalizer, can be expressed as

$$x(k) = [r(k), r(k-1), r(k-2), u(k-1), u(k-2)]^T \qquad (1)$$

and the output, before the soft-decision, is provided by

$$y(k) = Wx(k) \qquad (2)$$

where W is the weight coefficients matrix.

During then training phase, the desired response is known a priori. Therefore the feedback signal u(k) must be equals to the signal of the desired output, in this case, for a delay d = 2, $u(k) = \hat{s}(k-2)$. Therefore

$$x(k) = [r(k), r(k-1), r(k-2), s(k-3), s(k-4)]^T \qquad (3)$$

The output y(k) passes through the decision device and its hyperbolic tangent function, resulting in the estimative of the desired signal. Once more, as d = 2, the estimative of the output is given by

$$\hat{s}(k-2) = \tanh[y(k)] = \tanh[Wx(k)] \qquad (4)$$

The weight coefficients matrix W of the neural-DFE is updated by a training algorithm in order to minimize the error signal ε(k), by means of

$$\varepsilon(k) = \hat{s}(k-2) - s(k-2) \qquad (5)$$

In the testing phase, the output data are not known a priori, therefore the feedback signal u(k) must be equal to signal after the decision device, considering the delay d = 2. Thus, $u(k) = \hat{s}(k-2)$, obtained from (4) using

$$x(k) = [r(k), r(k-1), r(k-2), \hat{s}(k-3), \hat{s}(k-4)]^T \qquad (6)$$

The output y(k) passes through the decision device which in the testing phase is a hard-decision formed by the sign function. The result is the estimated signal

$$\hat{s}(k-2) = sig[y(k)] = sig[Wx(k)] \qquad (7)$$

III. Learning Technique

Methods based on the gradient present convergence rates of first order and can be very slow in equalization applications. Newton's methods guarantee convergence rates of second order; however they present the inconvenience of calculating the Hessian and its inverse [12].

Gauss-Newton method (GN) is used as a way to achieve convergence rates closed to second order with lower computation efforts when compared to Newton's methods.

The cost function in this work is the error between the corrected transmitted symbol and its estimative at the equalizer output, i.e.

$$e[\omega(k)] = \hat{s}(k) - s(k-d) \qquad (8)$$

where $\hat{s}(k)$ the estimative of the transmitted signal at the equalizer output when it is operating in soft decision mode with hyperbolic tangent function. Thus, the error expression can be rewritten as

$$e[\omega(k)] = \tanh[Wx(k)] - s(k-d) \quad (9)$$

The error function presented in (9) is used to form the Jacobian matrix employed in the Gauss-Newton algorithm. Each line of the Jacobian matrix corresponds to each training symbol m, for the weight vector at iteration k. Therefore, in the Jacobian matrix, the line number corresponds to the number of training symbols, and its column number is equal to the number of the equalizer coefficients. Each line of Jacobian matrix is computed using

$$\frac{\partial e(m)}{\partial \vec{\omega}} = \{1 - \tanh^2[Wx(k)]\}x(k) \quad (10)$$

Considering that in this paper, the DFE has five coefficients, then the Jacobian matrix at the time k, $J(k)$, can be represented by

$$J(k) = \begin{pmatrix} \frac{\partial e(1)}{\partial \omega_1} & \frac{\partial e(1)}{\partial \omega_2} & \cdots & \frac{\partial e(1)}{\partial \omega_5} \\ \frac{\partial e(2)}{\partial \omega_1} & \frac{\partial e(2)}{\partial \omega_2} & \cdots & \frac{\partial e(2)}{\partial \omega_5} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e(5)}{\partial \omega_1} & \frac{\partial e(5)}{\partial \omega_2} & \cdots & \frac{\partial e(5)}{\partial \omega_5} \end{pmatrix}_{\omega=\omega(k)} \quad (11)$$

Finally, the equalizer coefficients are updated according to Gauss-Newton iteration,

$$\begin{aligned} &\omega(k+1) - \omega(k) = \Delta\omega = \\ &-\{[J(k)^T J(k)]^{-1} J(k)^T\}\{\hat{s}(k) - s(k-d)\} \end{aligned} \quad (12)$$

It has been shown in [11] that the asymptotic analysis of the algorithm leads to computational complexity of the order $\boldsymbol{O}(l)$, where l is the number of symbols used for the algorithm. One can verify, in this case, that only two iterations are necessary to achieve a good estimative of the coefficients.

IV. TIME VARIANTS CHANNEL MODELS

As it was said before, in this paper it is evaluated the performance of the DFE presented in Fig. 1, training with GN algorithm. The DFE-GN is used to equalize 4-QAM signals which suffer the distortion imposed by three different kinds of time varying channels. Their models are presented in the following paragraphs.

Channel Model 1 - This channel model is described by the following transfer function:

$$H_1(z) = a_0(t) + a_1(t)z^{-1} + a_2(t)z^{-2} \quad (13)$$

where the coefficients $a_i(t)$ are varying in time. They are produced by a Markov chain process in which a white gaussian noise passes through a 2nd order Butterworth filter with cutoff frequency normalized according to the desired Doppler spread of the channel. This channel model has been proposed by [13], and there the authors used channels with the following features: bandwidth between 2 and 3kHz, transmission rate at 2400 bauds, and Doppler spreads of 0.5Hz and 1Hz.

Fig. 2 shows the variation of the channel coefficients along the time, for Doppler scattering of 1Hz. While Fig. 3 presents the scatter plots before the equalization takes place. The signal-to-noise ratios (SNR) are 12dB and 15dB.

Channel Model 2 - This channel model presents the same characteristics of the channel model 1, except for the fact that the first coefficient is constant and unitary. The channel has been analyzed in [14] for a Doppler frequency of 0.5Hz and it is described by the following transfer function:

$$H_2(z) = 1 + a_1(t)z^{-1} + a_2(t)z^{-2} \quad (14)$$

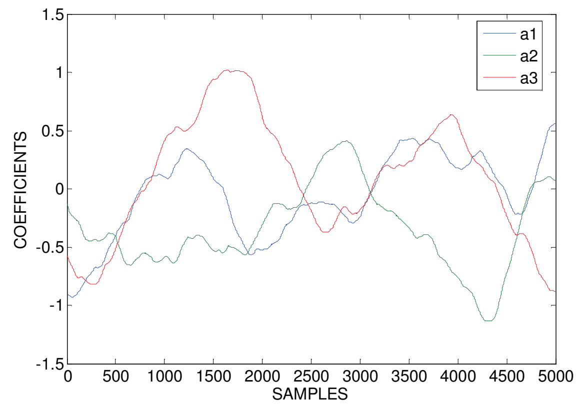

Fig. 2. Channel model 1, coefficient variations for Doppler spread of 1 Hz.

Channel Model 3 - The channel discrete model consists of three coefficients with delays equal to a symbol period, and it is

$$H_3(z) = [c_0 + a_0(t)] + [c_1 + a_1(t)]\, z^{-1} + [c_2 + a_2(t)]\, z^{-2} \quad (15)$$

where $c_0 = 0.3482$, $c_1 = 0.8704$ and $c_2 = 0.3482$.

The transfer function $H_3(z)$ is derived from a fixed channel model recommended by the ITU for testing of equalizers and it is frequently cited in literature [2]. In order to tranform it in a time varying channel, the $a_i(t)$ coefficients employed in Channel Model 1 are added to the fixed values c_i. However, in this case, the normalized cutoff frequency of the Butterworth filter is 0.1. This approach has been applied in [4] and proposed by [15].

Fig. 4 shows the variation of coefficients values for AWGN standard deviation of 0.3, while Fig. 5 presents the respective scatter plots for SNR equal to 6dB and 12dB.

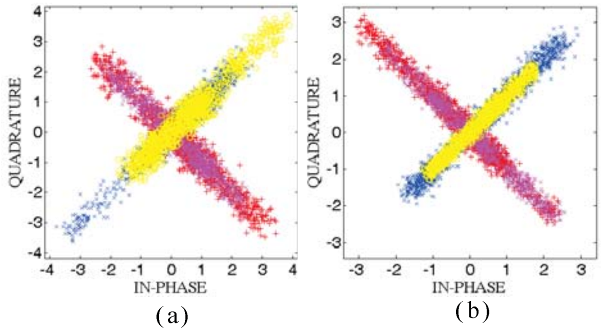

Fig. 3. Channel Model 1 - Scatter Diagram for Doppler spread of 1Hz: a) SNR 12 dB; b) SNR 15 dB.

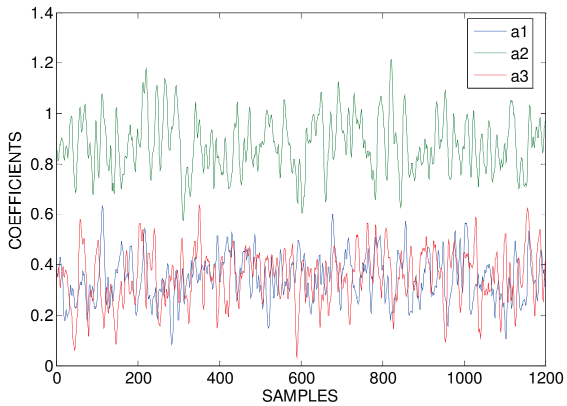

Fig. 4. Channel Model 3- Standard deviation of 0.3.

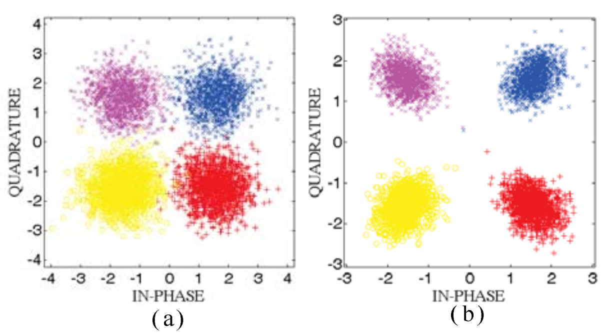

Fig. 5. Scatter plots for Channel Model 3. Standard deviation of 0.3: a) SNR 6dB; b) SNR 12dB.

V. Performance Evaluation

In order to compare the DFE-GN performance against other types of neural-equalizer proposed in the literature, two approaches were chosen during the testing phase.

In the first approach, the coefficients of the neural-equalizer proposed, DFE-GN, are frozen after a training period and then kept frozen during the testing phase. In the second approach, the coefficients are continuously updated by the GN algorithm in an attempt to tracking channel variations. For each sequence of testing symbols with identical amount of training symbols, a new Jacobiana matrix are computed and used for update the coefficients. The performance improvement, when achieved by the use of the second approach, is discussed separately for each channel model.

Channel Model 1 - For this channel model, the best results were obtained with 20 training symbols. In order to make the results comparable with [4], the ratio between training and testing symbols was fixed in 10. Therefore, for each SNR, 500 sequences were sent, each one with 20 symbols for training plus 200 symbols for testing, rendering in total 110.000 symbols tested for each value of SNR.

Simulations were performed considering both approaches during the testing phase. Fig. 6 shows that, for a Doppler spread of 1Hz, the use of GN algorithm to track channel variations provides a better result when compared to the approach of frozen coefficients after training. Therefore, for this kind of channel, we decide to employ the first approach in order to compare the DFE-GN with other topologies.

Fig. 7 shows the DFE-GN performance against the RNE-EKF, RNE-UKF e DFRNE neural equalizers proposed by [3]. The Doppler spread in this case was set to 1Hz.

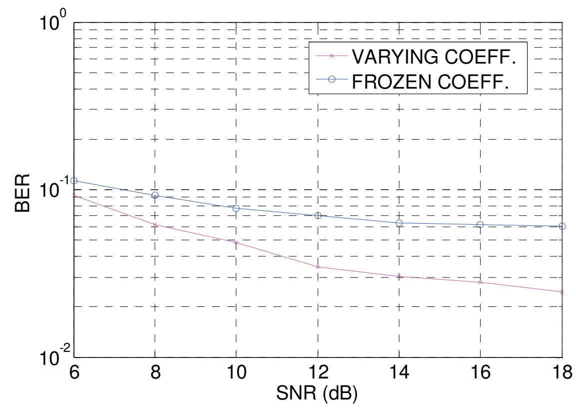

Fig.6. Channel Model 1 - BER comparison between frozen and varying coefficients for a Doppler spread of 1Hz.

Based on this result, we can verify for Channel Model 1 that the DFE-GN presents better performance when compared to RNE-EKF, RNE-UKF and DFRNE, for values of SNR lower than 11dB. But for SNR above 11dB, the proposed equalizer deteriorated its performance. That seems to indicate they are not suitable for fast fading channels.

Channel Model 2- For this channel model, the length of the training symbol sequence was also defined empirically. The best empirical result was once again 20 symbols. In order to make the results comparable with [14], the ratio between training and test symbols was fixed in 10. Therefore, for different SNR, 500 sequences were sent through the channel, each one with 20 symbols for training plus 200 symbols for testing. Again, the total amount symbols tested for each value of SNR was 110.000.

Fig. 8 presents the simulation results for a time varying coefficients with Doppler spread of 0.5Hz. There is an inversion of performance at SNRs around 10dB. For SNRs smaller than 10dB the improvement of performance with the tracking coefficients is not significant, on the other hand, for SNRs greater than 12dB, the improvement of the performance with the frozen coefficients is considerable. Thus, in this case, the choice is made in favor of keeping the coefficients frozen during the testing phase.

The topologies proposed in [14], trained by GEKF and RTRL learning algorithm, are used as benchmark for evaluation of Channel Model 2. In that article the authors set the Doppler frequency to 0.5Hz.

Fig. 9 reveals that DFE-GN outperforms in unquestionable way more complex RNN trained by GEKF and RTRL, for all range of SNR under test. Due to its extreme simplicity in implementation and significantly lower computational cost, the DFE-GN seems to be an outstand solution for the equalization of varying channels such as Channel Model 2.

Channel Model 3 - Initially was carried out simulation fixing SNR in 16dB and varying the standard deviation of AWGN source from the Markov chain. The values of standard deviation under testing were: 0.05, 0.1, 0.15, 0.2, 0.25, and 0.3, using 100 symbols for training and 1000 symbols for testing. Two hundred iterations were performed, like in [4].

Fig. 10 shows the simulation result for the coefficients behavior in the testing phase. One can verify that the DFE-GN does not have significant improvement in the performance when the GN algorithm attempts to track the variations of the channel. Considering that this small improvement does not worth to pay the price of computational cost, the decision was made to keep the coefficients frozen in testing phase, for this channel model.

Fig. 11 shows the DFE-GN performance, for Channel Model 3, when compared to other topologies proposed by [4], i.e.: DFRNE, RNE-EKF, and RNE-UKF. In this case, the SNR is set to 16dB for a variable standard deviation. It is important to emphasize that the standard deviation affects directly the channel variations.

Clearly the DFE-GN outperforms all other RNNs for the entire range of standard deviation under test.

The last evaluation is a comparison in terms of BER performance among different RNNs. The SNR, in this case, ranges from 4dB to 18dB, in steps of 2dB. The following parameters were considered during the simulations: standard deviation set to 0.1, 100 training symbols, and 100.000 testing symbols. The DFN-GN coefficients were frozen after the training phase. After 10 statistically independent experiments, the average value of the BER was obtained. This procedure is similar to the one carried out in [4]. Afterwards, the results were compared to those presented by [4]. Fig. 12 summarizes these results.

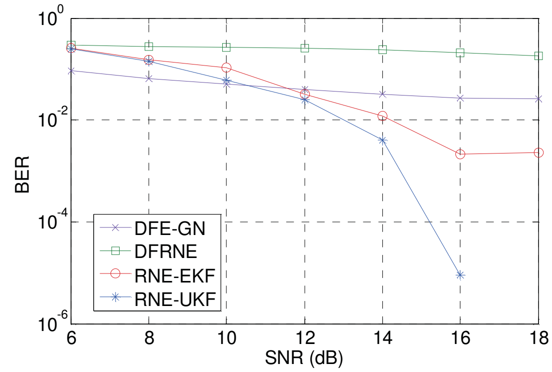

Fig. 7. Channel Model 1 - BER comparison for Doppler spread of 1Hz.

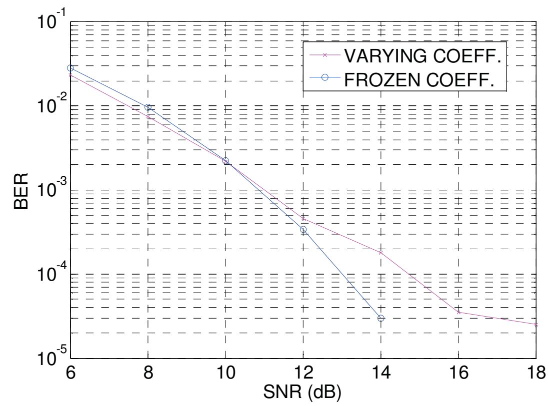

Fig. 8. Channel Model 2- BER for a Doppler spread of 0.5Hz.

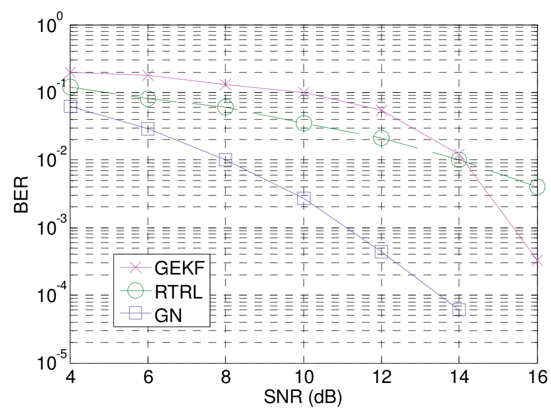

Fig. 9. Channel Model 2 - BER for a Doppler spread 0.5Hz

Again, the DFE-GN outperformed the topologies DFRNE,

RNE-EKF, and RNE-UKF, for all range of SNR under test. Once more, considering the simplicity of the structure and the involved computational cost, we can infer that the neural-equalizer with Gauss-Newton training algorithm is a promising equalization solution for channel such as Channel Model 3.

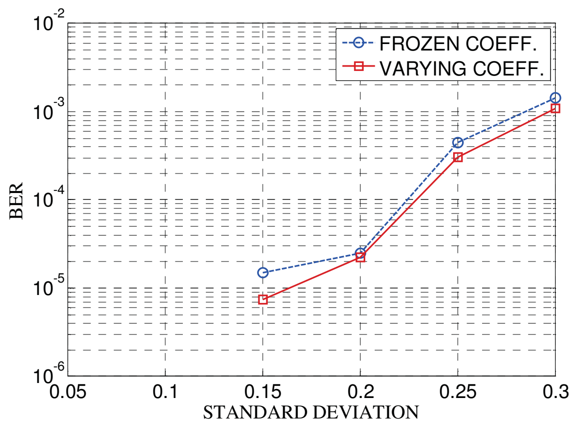

Fig. 10. Channel Model 3 - BER x Std. Deviation comparison for frozen and varying coefficients.

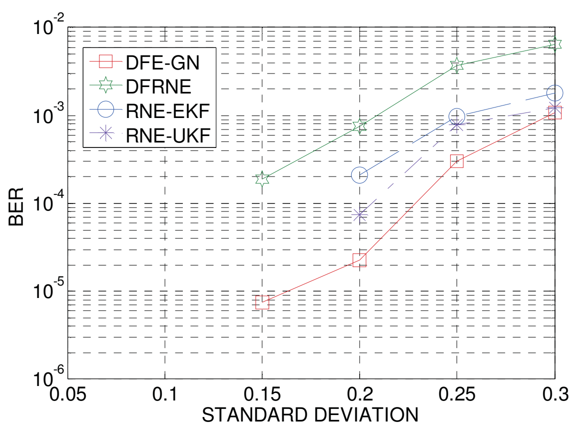

Fig.11. Channel Model 3 - BER x Std. Deviation, comparison among different topologies.

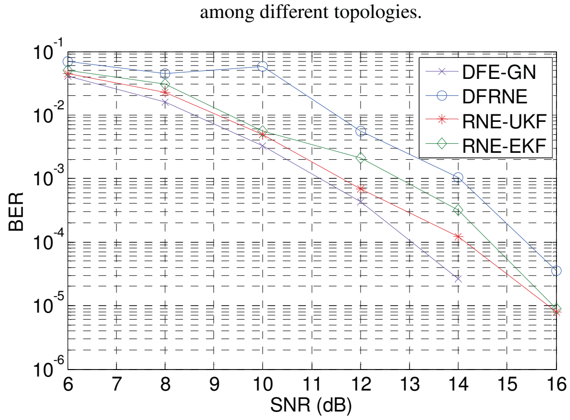

Fig. 12. Channel Model 3 - BER x SNR, comparison among different topologies.

VI. Conclusions and Remarks

This paper presented the behavior of a neural-equalizer in an attempt to mitigate the severe distortion imposed by time-varying channels to 4-QAM signal. The topology is a simple recurrent perceptron, derived from DFE, employing the well know Gauss-Newton algorithm during the training phase, and as an alternative for tracking in testing phase.

Its performance for different channel models revealed that this type of topology is a promising solution for channels with moderate time-varying conditions (slow fading). In this situation, it has been shown that the DFE-GN outperformed complex ANN topologies when comparison was made. However, for fast-fading channel, the neural-equalizer failed in part to achieve reasonable results.

Acknowledgment

Antonio C. de C. Lima would like to acknowledge the Brazilian Research Council (CNPq) for financial support.

References

[1] Haykin, S. (2001). Communications Systems, 4th edn, Willey.
[2] Proakis, J. G. (2001). Digital Communications, 4th edn, McGraw-Hill.
[3] Sklar, B. (2001). Digital Communications: Fundamentals and Applications, 2nd edn, Prentice Hall
[4] Choi, J., Lima, A. C. de C. and Haykin, S. (2005). Kalman Filter-Trained Recurrent Neural Equalizer for Time-Varying Channels, IEEE Trans. on Commun. 53(3): 472–480.
[5] Haykin, S. (1994). Neural Network: A Comprehensive Foundation, 1st edn, Prentice Hall.
[6] Ibnkahla, M. (2000). "Applications for Neural Networks to Digital Communications - A Survey", Signal Process., 80: 1185–1215.
[7] Parisi, R., Claudio, E. D. D., Orlandi, G. and Rao, B. D. (1997). Fast Adaptive Digital Equalization by Recurrent Neural Networks, IEEE Trans. Signal Process. 45: 2731–2739.
[8] Kechriotis, G., Zervas, E. and Manolakos, E. S. (1994). "Using Recurrent Neural Networks for Adaptive Communication Channel Equalizations," IEEE Trans. Neural Net., 5(2): 267–278.
[9] Chen, Z. and Lima, A. C. de C. (2004). A New Neural Equalizer for Decision-Feedback Equalization, IEEE Signal Processing Society Workshop, pp. 675– 684.
[10] Ong, S. Choi, S, You, C. and Hong, D. (1997). "A Complex Version of a Decision Feedback Recurrent Neural Equalizer as an Infinite Impulse Response Filter," in Proc. GLOBECOM'97, Nov. 3–8, pp. 57–61.
[11] Cortes, P.N., Junior, O.L. and Lima, A. C de C. (2006). Neuro-Equalizador com Algoritimo de Treinamento do Tipo Gauss-Newton, Anais do XVI CBA, n. 532, Salvador, Bahia,Brasil.
[12] Luenberguer, David G.(2003). Linear and Nonlinear Programming, 2nd edn, Springer.
[13] Ling, F., and Proakis, J. G. (1985). "Adaptive Lattice Decision-Feedback Equalizer - Their Performance and Application to Time-Variant Multipath Channels," IEEE Trans. Commun. vol. COM-33, pp. 348–356, Apr. 1985.
[14] Choi, J., Bouchard, M. and Yeap, T. (2005). Decision Feedback Recurrent Neural Equalization with Fast Convergence Rate, IEEE Trans. on Neural Net. 16(3): 699–708.
[15] Zerguine, A., Shafi, A., and Bettayeb, M. (2001). "Multilayer Perceptron-Based DFE with Lattice Structure", IEEE Trans. Neural Net., 12(4): 532–545.