On the Convergence of a New Joint DFE & Decoding Procedure for Blind Decision Directed LMS Equalization

C. M. Panazio and J. M. T. Romano
DSPCom, Faculty of Electrical and Computer Engineering, State University of Campinas, SP, Brazil,
DECOM – FEEC – UNICAMP, CP 6101, CEP 13083-970

Abstract- **Adaptive Equalization is a classical technique for mitigating ISI in unknown or time varying channels. Decision-feedback equalizer (DFE) is considered to be an efficient approach in many types of channels where linear equalizers fail. Unfortunately, it suffers from error propagation phenomenon. In order to reduce such effect, the present paper deals with the joint application of both equalization and decoding process in the receiver. The main contributions are the proposal of a new joint DFE and decoding configuration which works in a blind decision directed (DD) context and the study of the corresponding error surfaces behavior, and the assessment of the convergence rate. The proposed technique has shown to be effective for the so-called bad channels class, which makes the conventional DD-DFE ill-converge when its weights are initialized with zeros.**

1. Introduction

It is well known that intersymbol interference (ISI) is one of the major impairments to achieve a higher capacity or a data rate improvement in communication systems. Adaptive equalizer is a classical and efficient technique for mitigating ISI in unknown or time varying channels. The most conventional approach employs a training sequence to adapt the equalizer weights into an opened-eye condition, normally using the LMS adaptive algorithm. Then the equalizer is changed to the so-called decision-directed (DD) mode, in which the effective information is transmitted. However, in some specific systems, the use of a training sequence may not be practical. The adaptation process is then said to be unsupervised and some more robust (blind) algorithms are used. In this case, only a few statistics characteristics of the transmitted data symbols are known a priori.

Usually, the equalization can be achieved by a linear transversal filter. However, there are some channels that it fails to equalize due to a long channel impulse response or spectral nulls. Normally, for these channels, a non-linear structure called decision-feedback equalizer is used. Unfortunately, the DFE suffers from the error propagation phenomenon, which limits its performance.

Moreover, in modern digital communication systems, error correction codes are used to counteract the effects of noise and interference, aiming to achieve a more reliable transmission. With the advent of the Viterbi decoder, the use of convolutional codes has become feasible. In 1982, Ungerboeck has proposed a technique called Trellis-Coded Modulation (TCM) where it is possible to achieve a coding gain over an uncoded modulation, without increasing the transmission bandwidth [9].

The present paper deals with the joint application of both equalization and decoding process in the receiver, by taking into account that a TCM scheme without interleaving is used in transmission. Reference [5] has proposed a first joint DFE and decoding procedure for the case of supervised equalization. Such work has not provided any results concerning convergence analysis. The DFE equalizer was employed since its recursive nature is particularly suitable for dealing with the corrected symbols of a TCM code [5].

The main contribution of the present work are posed on two steps: firstly, we propose a new joint DFE and decoding configuration where the decided symbols of the survivor path of the node with smallest metric are fed back directly from the TCM decoder, instead of a memoryless decision device. Secondly, we study the convergence of both joint procedures by means of an error surface analysis, as well as in term of the convergence rate.

Finally, we show that our new approach is even more effective than the other alternative solutions proposed in order to improve convergence in DFE equalizers, such as the techniques as [5] or the use of soft decision devices [1,7].

In order to assess such results, this paper is organized as follows. In section 2, we make a brief explanation of the DFE equalizer and pose the assumptions to be used in the simulations. The joint DFE & decoding procedure is introduced in section 3. In sections 4 and 5 we present and discuss the results. The conclusions and perspectives are briefly posed in section 6.

2. The DD-DFE Equalizers

In spite of its simplicity and suitability in several applications, it is well known that linear equalizers suffer from important limitations, among which a critical one is the noise enhancement problem, in cases when the zeros of the channel are close to the unit circle. Due to its non-linear nature, DFE equalizers are an interesting alternative in such cases. Its recursive structure is also appropriate in other contexts, for instance when the channel presents a long impulse response. On the other hand, the performance of DFE structure can be affected by the phenomenon of error propagation. Fig. 2 illustrates the DFE structure. The DD-LMS algorithm to be used in this case is given by:

$$\begin{aligned} y(n) &= x(n) + \mathbf{b}^H(n)\hat{\mathbf{a}}(n) \\ e(n) &= \hat{a}(n) - y(n) \\ \mathbf{b}(n+1) &= \mathbf{b}(n) + \mu e^*(n)\hat{\mathbf{a}}(n) \end{aligned} \tag{1}$$

where $x(n)$ and $y(n)$ are the equalizer input and output respectively; $\hat{a}(n)$ is the decided symbol, so that its past values compose the vector $\hat{\mathbf{a}}(n)=[\hat{a}(n-1)\ \hat{a}(n-2)\ \ldots\ \hat{a}(n-N)]^T$, which feeds the recursive and adaptive filter defined, at instant n, by the weight vector $\mathbf{b}(n)=[b_1(n)\ b_2(n)\ \ldots\ b_N(n)]^T$. The weights are updated by means of the decision error $e(n)$, as given in (1).

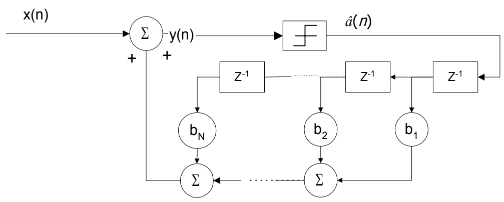

Fig. 1. DFE structure.

The analysis of convergence properties of the DD-DFE can be found in [1,4,6 and 7]. References [1,6] provide a class of channels that results in ill convergence when the feedback filter weights are initialized with zeros. Note that this is the most natural value to be used as initialization, since it guarantees convergence when the channel has an opened-eye condition. In this work, we deal with this class of bad channels to show that error correcting codes provide better decisions to the DD-DFE, so that it can converge to global minimum.

The following assumptions were made:

Assumption 1:

The source alphabet is QPSK {±1±j} obtained by the output of a convolutional encoder with rate R=½. This encoder is fed by an independent and identically distributed bit sequence with p(0)= ½.

Assumption 2:

The channel has a finite impulse response and the feedback filter matches the length of the channel postcursor response. The channel has no precursor and the leading tap dominates. Hence, we can define the vector **h**=[h_0 h_1 ... h_N], composed by N+1 elements of the channel impulse response, where h_0=1 and $|h_i|\leq 1$ for i=1,2,...N. Note that such condition does not imply in a minimum phase characteristic. Finally the channel is also considered to be noiseless.

Assumption 2 is very limiting for practical channels, where precursors are normally present. As a consequence of this assumption, the feedforward (FF) filter is not useful, and thereby it can be discarded. Therefore we can restrict our analysis to the feedback (FB) filter and to local minima associated with error propagation. Nevertheless, the assumption is justified since a full theoretical analysis was developed in [4], where both FF and FB filters were considered. However, this work did not take into account the impact of error correcting codes in the joint adaptation, which is the interest point of our work. Such analysis is not trivial and assumption 2 makes it more feasible. Thereby, further studies should be done in order to include precursor ISI and corresponding FF equalizer.

3. Joint DFE-Decoding Process

The first joint DFE and convolutional decoder structure to be used in this work is shown in Fig. 2. It can be seen that the feedback filter is divided in two parts, since the output of the decoder is expected to be more reliable then the output of the decision device. However, the decoder has an intrinsic delay. For that reason, its output can only be used as the input of the feedback filter if such delay is considered.

This scheme was firstly used in [5] with the objective of reducing the error propagation effect after a training period. In this work, we use the same idea but without a training period, i.e., in blind operation. The algorithm used for adaptation is the DD-LMS given by (1).

We also propose a new structure depicted in Fig. 3, where the decision device is no more employed and the survivor path of the node with the smallest metric is fed back to the FB filter. This procedure gives better performance, since the output of the decoder is more reliable than the decision device, even with no decision delay. Moreover, the new proposal can take advantage of delays larger than the tentative decision delay, which is used to feed back one part of the FB filter of [5].

Furthermore, the joint structure in Fig. 2 was originally proposed to mitigate error propagation after training [5]. However, since we are interested in a blind adaptation scheme, the feedback process in the adaptation algorithm is also important. The DD-LMS algorithm used in our structure has two types of feedback, the error calculation and the equalizer's input vector. We have seen that the feedback process over the algorithm is more important than the structure's feedback. Nevertheless, the best performance is achieved when both types of feedback are used, as pointed out in section 4.

We tested two different convolutional codes with R= ½. The first code has a polynomial generator [5 7] and the other has [64 74] (octal representation) [8]. The QPSK modulation uses the Gray code. This approach gives a simple trellis coded modulation scheme. The squared Euclidean distance is used as the metric of the Viterbi decoder.

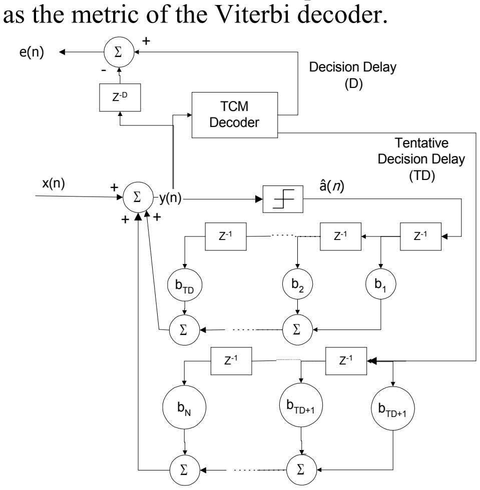

Fig. 2. Joint DFE and decoding procedure proposed in [5].

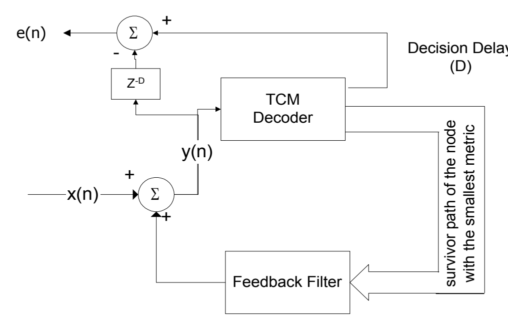

Fig. 3. New joint DFE and decoding procedure.

4. Effects on the Polytopes of the Performance Surface

4.1 Avoiding Local Minimum in the Performance Surface

The channel to be considered in our first simulation is given by $h(z)=1+z^{-1}-z^{-2}$. We have chosen this non-minimum phase channel because it is the worst case that we have found for the proposed technique and belongs to the class of bad channels defined in [1,6]. The error surfaces were obtained by fixing the two equalizer weights and transmitting a long sequence of about 3500 symbols. If ergodicity is assumed, a time average of the quadratic error can be calculated and used as an approximation of the cost function for each fixed pair of weights. The procedure is then repeated for a sufficient number of distinct weights, so that the surface contours are enough refined to render the analysis of the critical points. In Fig. 4, it is possible to observe the polytopes regions in the equalizer parameter space bounded by the hyperplanes defined by the nonlinear decision device.

It can also be observed that there is an undesired local minimum close to the point (0.0 0.0). Thus, the choice of such usual initialization results in ill convergence. This is illustrated by the trajectory of the DD-LMS algorithm from (0.0 0.0) to the local minimum in ≈ (-0.029,0.254), with a step-size of $\mu=1\text{x}10^{-4}$.

Fig. 5 shows that the joint technique in [5] is able to eliminate the undesirable local minimum that causes ill convergence. Note that no local minimum is observed in such figure. The error calculation of the DD-LMS is proceeded with a delay of 7 symbols, using the output of the decoder as desired signal. The input vector of the DD-LMS is the FB filter input delayed by 7 symbols.

In Fig. 6, we use only feedback of the decoder into the adaptive algorithm, in order to show its importance. In this case, the decoder's decision with a delay D is used to form the FB input vector. In comparison with Fig. 4 it can be seen that the local minimum which attracts the algorithm disappears. Nonetheless, there is another local minimum that was not observed in Fig. 5, but the algorithm is initialized outside its attraction domain. Another feature is that the attraction domain of the global minimum has larger influence over the region where the equalizer is initialized, when compared again with Fig. 5.

On the other hand, the new proposal (Fig. 7) with both algorithm and structure feedback, reveals the wideness of the attractive domain of the global minimum, which is expected to result in a better convergence, in comparison with Fig. 6. The improvement over the error surface of Fig. 6 occurs since the decoder can provide a smaller error rate than the case where only the decision device feeds back.

The channel with spectral null $h(z)=1+z^{-1}+z^{-2}$ also can be equalized without any problems, using the new joint equalization and decoding procedure, as can be seen in figure Fig. 9. In contrast, Fig. 8 confirms the convergence into the local minimum when the conventional DD-DFE is used in this kind of channel.

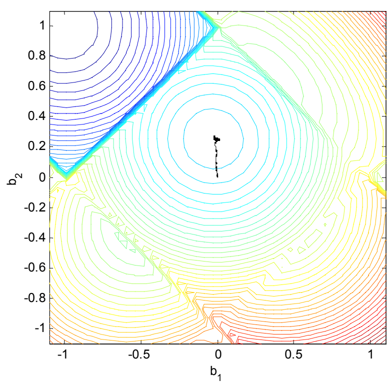

Fig. 4. DD-DFE error-performance surface for $h(z)=1+z^{-1}-z^{-2}$ and $\mu=1\text{x}10^{-4}$.

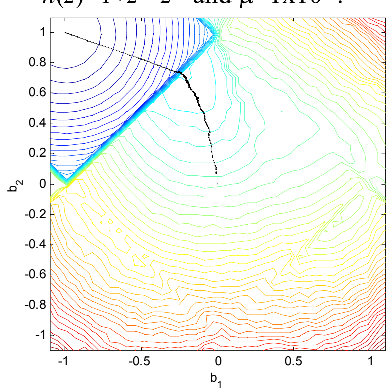

Fig. 5. Joint DFE & decoding procedure for $h(z)=1+z^{-1}-z^{-2}$ with convolutional code generated by polynomial [5 7], D=7, TD=1 and $\mu=1\text{x}10^{-4}$.

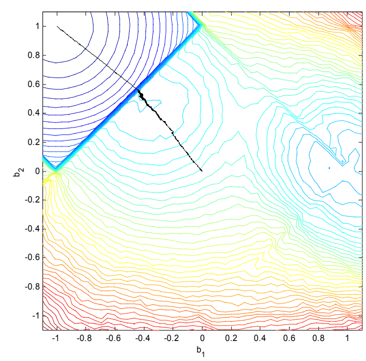

Fig. 6. DD-DFE with decoder feedback only into the algorithm. Convolutional code generated by polynomial [5 7], D=7, $h(z)=1+z^{-1}-z^{-2}$ and $\mu=1\text{x}10^{-4}$.

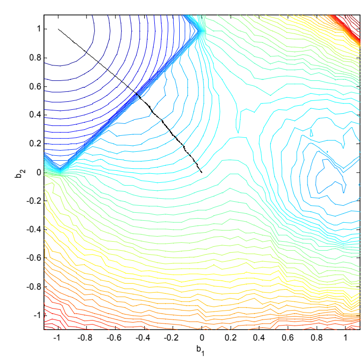

Fig. 7. New joint DFE & decoding procedure, with convolutional code generated by polynomial [5 7], D=7, $h(z)=1+z^{-1}-z^{-2}$ and $\mu=1\text{x}10^{-4}$.

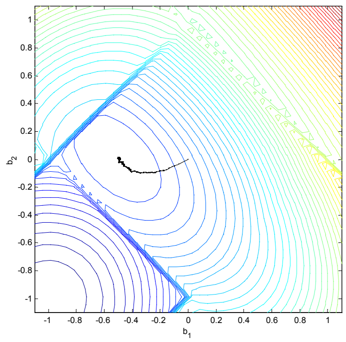

Fig. 8. DD-DFE error-performance surface for $h(z)=1+z^{-1}+z^{-2}$ and $\mu=1\text{x}10^{-4}$.

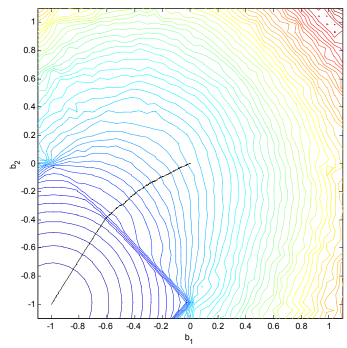

Fig. 9. New joint DFE & decoding procedure, with convolutional code generated by polynomial [5 7], $D=7$, $h(z)=1+z^{-1}+z^{-2}$ and $\mu=1\text{x}10^{-4}$.

In this channel, there is not a local minimum for the new joint procedure as it was observed in Figs. 6 and 7. By the way, only a few classes of channels present a local minimum. They are $h(z)=1+z^{-1}-z^{-2}$ and some similar channels such as $h(z)=1+0.8z^{-1}-0.8z^{-2}$. Further studies are in course in order to provide a more complete analysis of such behavior.

4.2 The effects of the convolutional decoder

References [1,6] point out that the DD-DFE has a piecewise quadratic cost function, where there is only one possible local minimum inside each polytope. The DD-DFE cost function for a given polytope $\mathcal{P}$ can be written as:

$$J_{DD}(\mathcal{P}) = E\left\{\frac{1}{2}\left|\hat{a}(n) - \mathbf{h}^T\mathbf{a}(n) - \mathbf{b}^H\hat{\mathbf{a}}(n)\right|^2\right\} \qquad (2)$$

where $\mathbf{a}(n)=[a(n)\ \ a(n\text{-}1)\ \ \ldots\ \ a(n\text{-}N)]^T$. Thus, setting its gradient to zero, we find the possible local minimum:

$$\mathbf{b}_m(\mathcal{P}) = E\left\{\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^H(n)\right\}^{-1}\left[E\left\{\hat{a}^*(n)\hat{\mathbf{a}}(n)\right\} - E\left\{\hat{\mathbf{a}}(n)\mathbf{a}^H(n)\right\}\mathbf{h}^*\right] \quad (3)$$

The statistics $E\{\cdot\}$ are constants within a polytope, but they assume different values for different polytopes. Therefore, they are piecewise constant. Nonetheless, this is not true when we use, for instance, a soft decision device, which *smooths* the cost function [1]. Reference [1] states that these statistics vary with the equalizer parameters and they tend to approach perfect equalization condition as the algorithm moves closer to the solution, rather than remaining fixed over a large region as for the DD-DFE. We believe by analogy that the use of joint equalization and decoding procedure provides a similar behavior, despite the decoder returns transmitted symbols.

However, the joint DFE and decoding technique is more effective than using the soft decision device, since convergence to the global minimum in more critical cases can be attained (see Fig. 11 in the next section). This better performance appears to be related to the code correction capability, even in critical situations where there are a long sequence of errors, so that the code correction capability is exceeded. Even so, it does work!

Table 1. Spectral Null Channel, code=[5 7]

	Error Rate		
Decision Delay	**b**= (0,0)	**b**= (-0.29,-0.14)	**b**= (-0.57,-0.36)
0	0.6148	0.3914	0.3573
1	0.5756	0.3834	0.3500
7	0.6457	0.3770	0.2998

Table 2. Channel [1 1 –1], code=[5 7]

	Error Rate		
Decision Delay	**b**= (0,0)	**b**= (-0.22, 0.29)	**b**= (-0.41, 0.51)
0	0.4800	0.3457	0.3470
1	0.5062	0.3259	0.3484
7	0.4581	0.3433	0.3217

In tables 1 and 2, we have three configurations where the error rate is obtained, fixing the filter weights. The first configuration is set in the initialization, the third is set in near the border of the region that there is no more decision errors and the second is set in the middle of these two configurations. As we can observe, the decoder does not behave as expected: the error rate oscillates with the decision delay and such oscillations tend to be smaller as the delay increase.

Changing the code into [64 74] does not always bring a better performance as should be expected. Channels with less ISI have presented a slight increase in performance when using such code. We believe that such phenomena are related to the situation of these large error rates. The code was not designed to handle so much errors, specially burst errors. Further studies must be carried out with other codes and other modulations, in order to provide more general conclusions.

5. Convergence Rate Assessment

Another important aspect to be taken into account is the value of the adaptation step size μ. It is known that the algorithm may escape from the local minimum by increasing the value of the step size [3]. In order to compare the performance of conventional DD-DFE and the joint DFE and decoding in terms of the adaptation step size, an example is shown in Fig. 10. The convolutional code was generated by polynomial [5 7] and the channel was $h(z)=1+0.4z^{-1}-0.2z^{-2}+0.8z^{-3}-0.7z^{-4}+0.1z^{-5}$. The same step-size was used for both structures in each case. We have made an ensemble average over 40 trials.

As it can be observed, the step size in Fig. 10a is too small, so that the conventional DD-DFE remains in the local minimum, while the joint structure converges toward the optimal solution. On the other hand, Fig. 10b shows a case where conventional DD-DFE escapes from the local minimum. Nevertheless, we note that the joint DFE and decoding procedure converges significantly faster. Additionally, Fig. 10c shows the worst case of gradient noise due to the higher value of step size, which leads the code to lose performance, so that it converge slower than the conventional DD-DFE. Note, however, that the value used for the adaptation step size is very high and close to the limit for which convergence is guaranteed and that the conventional DD-DFE remains slower than the joint procedure with the step size used in 10b.

Others solutions have been proposed in the literature in order to search for the convergence to the desired global minimum in DFE. An interesting one is the use of soft decision like a saturation device [1]. We tested such approach with many types of bad channels and observed that it converges to the desired global minimum for the majority of them. However, for non-minimum phase channels, specially channels with zeros located far from the unitary circle, it fails to converge.

For $h(z)=1+0.9z^{-1}-0.8z^{-2}$ it converges, but it fails for $h(z)=1+0.95z^{-1}-0.9z^{-2}$. A comparison with the joint and the new joint DFE and decoding procedure is shown in Fig. 11.

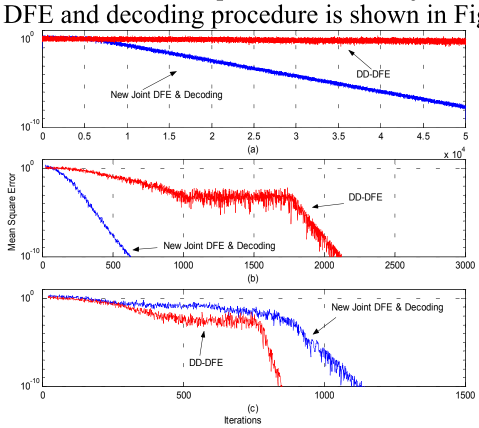

Fig. 10. Dynamical comparison between conventional DD-DFE and the joint DFE & decoding procedure: (a) μ=1x10^{-4}, (b) μ=0.01, (c) μ=0.05

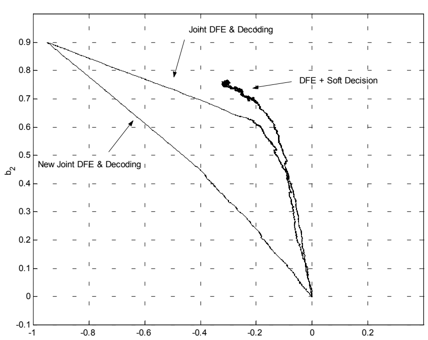

Fig. 11. Weights trajectories for the joint and the new joint DFE & decoding procedure vs. DFE + soft decision. Algorithm step size is equal to 1x10^{-4}

6. Conclusions & Perspectives

In this paper we have shown that it is possible to achieve convergence to the desirable global minimum with the use of the DD-LMS. Using the DFE structure together with DD-LMS, there is a class of channels that presents ill convergence when the feedback filter weights are initialized with zeros. We have shown that the use of error correcting codes renders the convergence to the desirable global minimum possible, even for channels belonging to this class.

Additionally, the proposed new joint DFE and decoding has shown to be more effective than the one in [5], since it deals with more reliable feedback symbols.

The convergence analysis was based on the error surface, in order to study the behavior of the critical points, as well as the convergence rate. The proposed technique has shown to be rather efficient for the class of bad channels.

Further studies concerns the achievement of more general result, like different modulation scheme. It is worth pointing out that complete analytical results are extremely difficult to obtain in such studied configuration due to the difficulties imposed by the decoding with error propagation and residual ISI. It is also in course a study with the constant modulus algorithm.

Acknowledgment

We would like to express our appreciation to the State of São Paulo Research Foundation (FAPESP) for supporting this work.

References

[1] R. A. Casas, C. R. Johnson, Jr., R. A. Kennedy, Z. Ding and R. Malamut, “Blind adaptive decision feedback equalization: a class of channels resulting in ill-convergence from a zero initialization,” Int. Journal of Adaptive Control Signal Proc., vol. 12, 173-193, 1998.

[2] O. Macchi and E. Eweda, “Convergence analysis of self-adaptive equalizers,” IEEE Trans. Inf. Theory, vol. 30, no. 2, 161-176, March 1984.

[3] J. E. Mazo, “Analysis of decision-directed equalizer convergence”, Bell System Technical Journal, vol. 59, no. 10, 1857-1876, December 1980.

[4] R. Kennedy, B. Anderson, R. Bitmead, “Blind adaptation of decision feedback equalizers: gross convergence properties,” Int. Journal of Adaptive Control Signal and Processing, vol. 7, 497-523, 1993.

[5] S. L. Ariyavisitakul and G. M. Durant; “A Broadband Wireless Packet Technique Based on Coding, Diversity, and Equalization,” IEEE Comm. Mag., 110-115, July 1998.

[6] R. A. Casas, “Blind adaptive decision feedback equalization: a class of bad channels’, MSc Thesis, Cornell University, 1996.

[7] S. Marcos, S. Cherif, M. Jaidane, “Blind cancellation of intersymbol interference in decision feedback equalizers,” Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1073-1076, May 1995.

[8] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[9] G. Ungerboeck, “Trellis-Coded Modulation with Redundant Signal Sets, Part I and II,” IEEE Commun. Mag., vol. 25, pp. 5-21, February.