Blind DFE With Maximum-Entropy Feedback

Vladimir R. Krstić, *Member, IEEE*, and Miroslav L. Dukić, *Member, IEEE*

***Abstract*—This letter proposes an entropy-gradient adaptive feedback filter specially derived for the blind decision feedback equalizer with a self-optimized configuration. Using software simulations, the parametric nonlinearity of the feedback filter is aligned through two operation modes—blind acquisition and soft transition—to respond with a maximum entropy signal. As a result, a simple stochastic gradient algorithm is obtained, and it can be easily optimized for an applied signal constellation.**

***Index Terms*—Blind decision feedback equalizer (DFE), maximum-entropy feedback filter.**

I. Introduction

THIS letter addresses the blind equalization of linear time-invariant channels by a decision feedback equalizer (DFE) in quadrature amplitude modulation (QAM) systems. The acute drawback of a classical DFE scheme using hard decision feedback loop is the possibility of a disastrous error propagation that emerges immediately at the presence of data and prevents its activation. To "skip" this burden of errors, Labat *et al.* [1] proposed a self-optimized DFE (SO-DFE) scheme that optimizes both the structure and the adaptation criteria. At the start, the SO-DFE decomposes the structure into a cascade of linear adaptive devices performing a difficult blind equalization task step by step, and then, when a signal eye is open enough, switches the structure back so that the decision-directed minimum mean-square error (DD MMSE) adaptation can commence. Although the SO-DFE works reasonably well, its overall convergence achievements are not very impressive in severe fading environments since its least mean square (LMS)-based feedback filter cannot meet the learning ability to compensate for deep spectral nulls in the transmission passband.

The noted weakness of the feedback filter can be remarkably mitigated using Shannon's entropy measure [2] suitably transformed into the joint entropy maximization criterion (JEM) [3]. Based on this concept, in [4], the complex stochastic-entropy-gradient algorithm has been derived for the recursive part of the SO-DFE. As a continuation of that research, in this letter, we explore how the shape parameter β of the complex-valued nonlinearity $g(z) = z(1+\beta|z|^2)$, which is involved into the feedback filter, influences the convergence of the SO-DFE. It is shown via simulations that the practicable range of values of the shape parameter can be determined for an applied signal. This design approach leads to stochastic gradient algorithms that evade the adaptive shape-fitting of the nonlinearity.

Manuscript received July 17, 2008; revised September 18, 2008. Current version published December 12, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tongtong Li.

V. R. Krstić is with the Institute Mihailo Pupin, 11 060 Belgrade, Serbia (e-mail: vladak@kondor.imp.bg.ac.yu).

M. L. Dukić is with the Faculty of Electrical Engineering, University of Belgrade, 11 000 Belgrade, Serbia (e-mail: dukic@etf.bg.ac.yu).

Digital Object Identifier 10.1109/LSP.2008.2008566

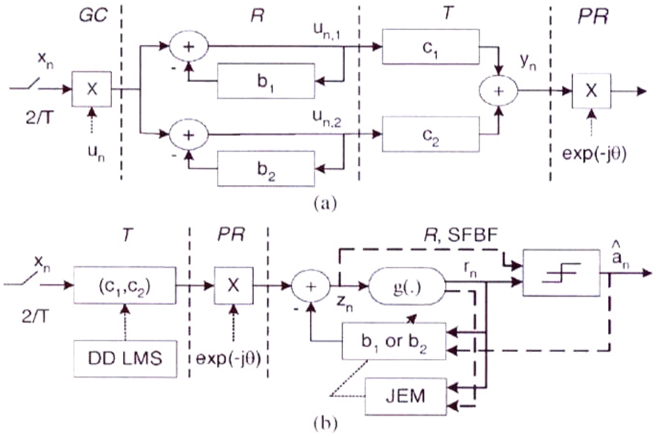

Fig. 1. Soft-DFE configurations in (a) blind acquisition and (b) soft transition mode. SFBF is given by basic (solid bold lines) and modified (dashed bold lines) schemes.

II. Soft-DFE Structure

The innovated SO-DFE, termed Soft-DFE, combines several adaptation criteria through three operation modes. In the blind acquisition, the Soft-DFE acts as a linear 2/T-spaced equalizer (T is the symbol interval) including four adaptive transformers in the cascade—gain control (GC), decorrelator (R), fractionally-spaced equalizer (T), and phase rotator (PR)—where R and T perform the most critical subtasks; see Fig. 1(a). The decorrelator R, which consists of two independent purely recursive filters (whiteners) with coefficient vectors $\mathbf{b}_i = [b_{i,1},\ldots,b_{i,N}]^T$, $i = 1, 2$, compensates the channel attenuation distortion while the equalizer T, which is defined by coefficient vectors $\mathbf{c}_i = [c_{i,1},\ldots,c_{i,L}]^T$, compensates the phase distortion introduced by the channel-decorrelator combination. The adaptation of R and T is decoupled and based on the JEM [4] and the constant modulus algorithm (CMA) [5], respectively. In the next stage, named the soft transition mode, one of the two whiteners—selected according to energy criterion—becomes the soft feedback filter (SFBF) keeping on the JEM adaptation and the equalizer T switches the algorithm from CMA to DD LMS; see Fig 1(b). Finally, when the signal eye is open enough, the SFBF switches into the DD LMS adaptation (tracking mode). In the described scheme, the GC and PR are implemented in a similar way as in [1].

III. Entropy-Gradient Algorithms for Feedback Filter

Let us recall the JEM criterion $J_{EM} = E\{\ln|\partial r_n/\partial z_n|\}$ [3], where z_n is the input of the monotonically increasing function $g(z_n)$ and E denotes expectation. Based on the information-theoretic principles [2], a suitably selected nonlinearity in the given structure [basic scheme in Fig. 1(b)] maximizes the joint entropy of the equalizer output $r_n = g(z_n)$, and by doing so, the intersymbol interference (ISI) removing will result. The JEM criterion was derived assuming the following real-valued system model: 1) a zero-mean i.i.d. non-Gaussian sequence a_n is applied to a linear noiseless channel; 2) the probability density function (pdf) of the ISI is unknown; and 3) the previous outputs of the SFBF are correct, i.e., $r_{n-j} = a_{n-j}, j = 1, \ldots, N$. Extending the described model into the complex domain, we have derived in [4] the JEM-type stochastic gradient algorithms for the recursive part of the Soft-DFE.

In the particular case of the complex-valued nonlinearity $g(z_n) = z_n(1 + \beta|z_n|^2)$, which is continuously differentiable and bounded-input bounded-output (BIBO) stable [4], the JEM-type stochastic gradient algorithm reads

$$b_{j,n+1} = b_{j,n} - \mu z_n \left(1 - \beta|z_n|^2\right) r^*_{n-j}, \quad j = 1, \ldots, N \quad (1)$$

where μ denotes a step size and β is a real positive constant. The algorithm (1) represents the basic learning rule for the SFBF where $e_n = z_n(1 - \beta|z_n|^2)$ can be viewed as a soft error signal, the quality of which depends on the shape parameter β. Having in mind that R changes its position with respect to T during the adaptation process (see Fig. 1), the selection of β has to take into consideration the two operation modes: blind acquisition and soft transition.

In the first case, the soft decisions r_{n-j} in (1) are replaced by their linearized counterparts z_{n-j}. This approximation has a twofold effect. First, the basic soft feedback filter becomes a purely recursive whitener. Second, the JEM algorithm is reduced to the corresponding whitening algorithm (JEM-W)

$$b_{j,n+1} = b_{j,n} - \mu_W u_n \left(1 - \beta_W|u_n|^2\right) u^*_{n-j}, \quad j = 1, \ldots, N \quad (2)$$

where μ_W denotes a step size and β_W stands for a whitening parameter. Note that the stochastic gradient in (2) does not depend on the previous innovations u_{n-j} but only on the current one. In fact, the gradient in (2) is calculated as if the JEM-whitener was nonrecursive. Let us recall that the same situation holds for the extended LMS algorithm [1], which can be seen as a special case of the JEM-W for $\beta_W = 2/|u_n|^2$. A major feature of the JEM-W is its prediction error $e_n(\beta_W)$ that gains the information content of the most faded components of input x_n. At the same time, it should be noted that the potential risk of instability is much higher than in the case of LMS whitening. Although it is still an open question currently under study, the BIBO stability of JEM-whitener has been proved for QAM signals [6].

In the second case, the heuristic modification of the basic SFBF scheme is done in a way that the hard decision device is turned back into the SFBF [see the modified scheme in Fig. 1(b)]. Now the hard decision estimates of the transmitted symbols $\widehat{a}_{n-j}$ feed the feedback filter instead of the soft decisions r_{n-j}. Consequently, the decision-directed variant of the JEM algorithm (DD-JEM) is obtained

$$b_{j,n+1} = b_{j,n} - \mu_D z_n \left(1 - \beta_D|z_n|^2\right) \widehat{a}^*_{n-j}, \quad j = 1, \ldots, N. \quad (3)$$

The above algorithm combines the soft errors $e_n(\beta_D)$ and hard symbol estimates $\widehat{a}_{n-j}$ that may be unreliable at the moment of structure switching. For a reasonably selected switching threshold level between blind and soft transition mode, the quality of the symbol estimates mostly depends on convergence capabilities of the JEM-W and CMA algorithms (previous equalization subtasks) as well as on the current error $e_n(\beta_D)$. Thus, by using the improvement opportunity of the parameters β_W and β_D in JEM-W and DD-JEM, respectively, the above heuristic modification can be justified.

To assess the effects of $\beta = \{\beta_W, \beta_D\}$ on the JEM algorithms and to estimate their useful ranges of values for an applied signal constellation, the convergence features of the algorithms are evaluated via software simulations. The influence of β_W is observed at the output of T that maximizes the absolute kurtosis of y_n. In this context, the CMA is not seen as a constant modulus criterion [7] but rather as a special case of the kurtosis maximization criterion in the case of sub-Gaussian inputs [8]. In that sense, the absolute normalized kurtosis can be estimated by formulas $K_s = (1/2)\sum_{i=1}^{2} K_i$ and $K_i = K(y_{i,n})/K(a_n) = [\|\mathbf{c}_i\|_4/\|\mathbf{c}_i\|_2]^4$, where $\|\mathbf{c}_i\|_q = \left[\sum_{k=1}^{L} |c_{i,k}|^q\right]^{1/q}$. This kurtosis estimation method relies on a one-by-one correspondence between the stationary points in system (channel-equalizer) and equalizer domains [9]. On the other hand, the influence of β_D is assessed by means of symbol error rate (SER) that is measured during the soft-transition mode. In fact, the SER has been considered as a figure of merit for the error propagation effects.

IV. Simulation Results

The Soft-DFE features are characterized in terms of the kurtosis K_s, SER, and overall MSE convergence. Also, the MSE convergence of the Soft-DFE and SO-DFE (termed Hard-DFE) are compared. The only difference between these two solutions is the adaptation method applied to their recursive parts: the Soft-DFE relies on JEM and the Hard-DFE on LMS-type algorithms [1]. The simulations are carried out using 16- and 32-QAM signals and fading channels under a signal-to-noise ratio of 25 dB. The three-ray channel model [7] is involved into the transmitter filter; Fig. 2 depicts the attenuation response of filter-channel combinations Mp-(A,B,C,D,E) for different propagation parameters selected to gradually increase the level of ISI. The length of both equalizers is L = 22 and N = 6 in their T and R parts, respectively. The initial vectors of T are with zero components except the centered (reference) ones $c_{1,r} = c_{2,r} = 1$. The presented results are averaged over 200 Monte Carlo runs.

The equalizers switching threshold level is defined in term of the output MSE which is estimated in a similar way as in [1]. The Soft-DFE switches from blind to soft transition mode at the threshold level of 1.5 dB and then the soft transition continues during the next 2000 symbol intervals. This time period has been decided to be long enough for a signal eye opening

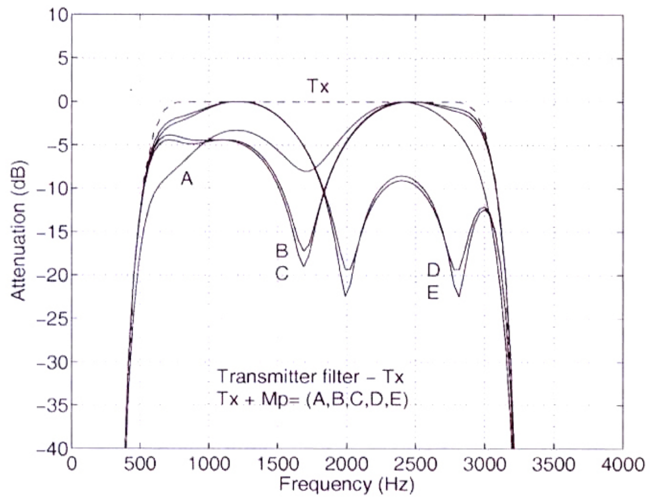

Fig. 2. Attenuation response of the MP channels.

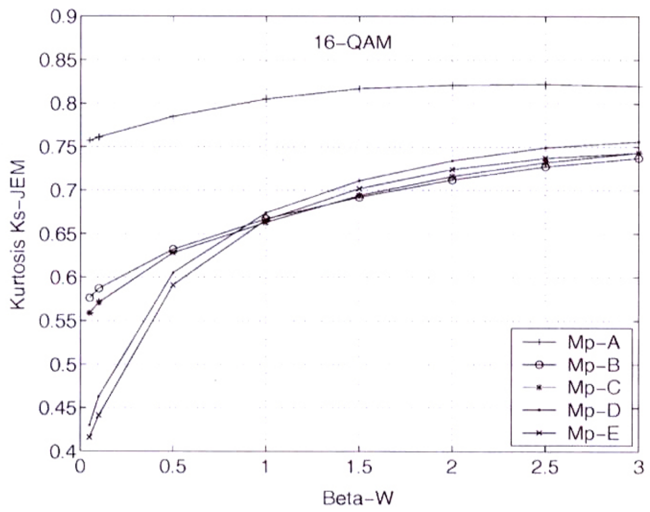

Fig. 3. Normalized kurtosis at the output of T.

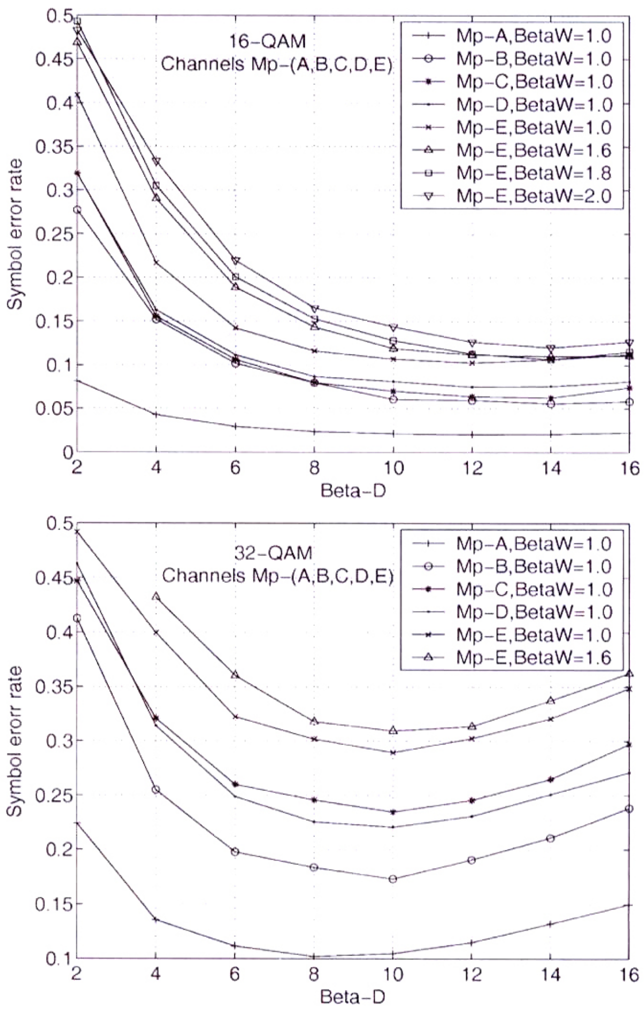

Fig. 4. Symbol error rate measured during the soft transition mode for 16-QAM and 32-QAM signals.

to be completed with a high probability of success. After that, the SFBF adaptation switches from DD-JEM to DD-LMS algorithm. On the other hand, the Hard-DFE directly switches, at the above threshold level, from blind acquisition to conventional DD LMS tracking.

The maximal kurtosis yields $K_{s,JEM}$ achieved by 16-QAM signal for different values of β_W are depicted in Fig. 3; the corresponding curves for 32-QAM closely follow the presented ones. The kurtosis growth, which is similar for all channels, shows the ability of the JEM-W to improve a channel conditioning. Similar results can also be expected for other QAM signals and channels. In the next steps, this feature of the JEM-W is taken as a design guideline to consider in more detail the JEM whitening in the β_W range from 0.8 to 2.0. For the sake of comparison, the corresponding kurtosis $K_{s,LMS}$, obtained by Hard-DFE and 16-QAM, is (0.76, 0.47, 0.57, 0.45, 0.59) for Mp-(A,B,C,D,E) channel, respectively. It is easy to see that these kurtosis values are approximately the same as the corresponding $K_{s,JEM}$ for $\beta_W = 0.15$.

The SER versus β_D is shown in Fig. 4. The provided curves indicate the capability of DD-JEM to mitigate the error propagation. Independently of channels, these curves present minima for the same optimal value of β_D that can be estimated as $\beta_{D,16} = 12$ and $\beta_{D,32} = 10$ for 16- and 32-QAM signals, respectively. In fact, selecting β_D in this way, we have fitted the shape of $g(z)$ which makes the pdf of its output to appear more flat [2]. Associated with this ability of the SFBF, the relation $\beta_{D,32} < \beta_{D,16}$ indicates that the intersymbol interference that comes from 16-QAM has a more "peaked" pdf than is the one of 32-QAM; recall that the optimal slope of a sigmoid maximizing entropy is inversely proportional to the variance of its input distribution [2], [10].

Now, when we have in hand the optimal values of β_D for given signals, we can refine the influence of parameter β_W on the Soft-DFE convergence. Fig. 5 depicts the MSE convergence characteristics for different values of β_W: they present a gradual improvement of both the speed and the residual error for 16-QAM and β_W in the range of {0.8, 2.0}, while in the case of 32-QAM, the corresponding range of β_W is a little smaller {0.8, 1.4}. Moreover, the positions of SER minima have

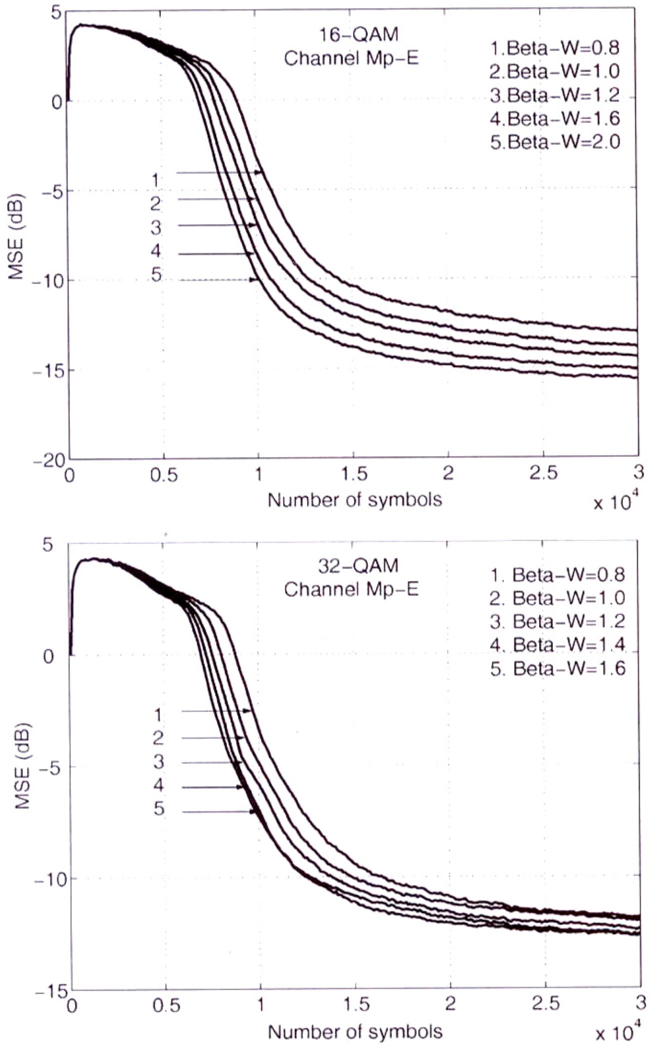

Fig. 5. MSE convergence of Soft-DFE ($\beta_{D,16} = 12, \beta_{D,32} = 10$).

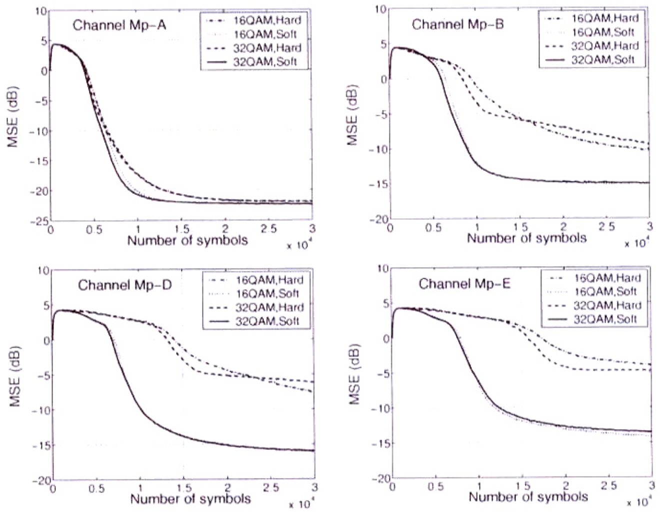

Fig. 6. MSE comparison of Soft-DFE ($\beta_W = 1.2, \beta_{D,16} = 12, \beta_{D,32} = 10$) and Hard-DFE for Mp-(A,B,D,E).

stayed unchanged over the observed ranges of β_W (see Fig. 4). Finally, the test of comparison of the Soft-DFE and Hard-DFE is presented in Fig. 6. For the moderate channel Mp-A, the convergence characteristics of both solutions are similar, but with the severe channels MP-(B,D,E) the Hard-DFE is inferior to the Soft-DFE. In the latter case, the Hard-DFE shows a slow blind acquisition and a high percentage of failed transitions from blind to tracking mode.

V. Conclusions

This letter presents the performance of the entropy-based stochastic gradient algorithms, which are derived for the recursive part of the self-optimized DFE. It is shown that the shape of the nonlinearity can be optimized for 16- and 32-QAM signals so that the soft feedback filter responds with minimal error propagation. Also, the optimal values of the shape parameters are practically channel-independent and can be verified via simulations for any other QAM signal. The overall performance tests have confirmed the superiority of the new DFE over the DFE solution based on the traditional LMS-type algorithms.

References

[1] J. Labat, O. Macchi, and C. Laot, "Adaptive decision feedback equalization: Can you skip the training period?," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 921–930, Jul. 1998.

[2] J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, pp. 1129–1159, 1996.

[3] Y. H. Kim and S. Shamsunder, "Adaptive algorithms for channel equalization with soft decision feedback," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 9, pp. 1660–1669, Dec. 1998.

[4] V. R. Krstic and Z. Petrovic, "Complex-valued maximum joint entropy algorithm for blind decision feedback equalizer," in *Proc. 8th Int. Conf. Telecommunications in Modern Satellite Cable and Broadcasting Services (IEEE-TELSIKS 2007)*, Niš, Serbia, pp. 601–604.

[5] C. R. Johnson, Jr. *et al.*, "The core of FSE-CMA behavior theory," in *Unsupervised Adaptive Filtering, Vol II Blind Deconvolution*, S. Haykin, Ed. New York: Wiley, 2000, pp. 13–112.

[6] V. R. Krstic and M. Dukic, "On complex domain dicision feedback equalizer based on Bell-Sejnowski neuron," in *Proc. 3rd Int. Symp. Communication, Control and Signal Processing (IEEE-ISCCSP 2008)*, St. Julians, Malta, pp. 1167–1172.

[7] Y. Li and Z. Ding, "Global convergence of fractionally spaced godard (CMA) adaptive equalizers," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 818–826, Apr. 1996.

[8] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of non-minimum phase systems (Channels)," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.

[9] G. J. Foschini, "Equalizing without altering or detected data," *AT&T Tech. J.*, vol. 64, pp. 1885–1911, Oct. 1985.

[10] S. Fiori, "Some properties of Bell-Sejnowski PDF-matching neuron," in *Proc. 3rd Int. Conf. Independent Component Analysis, and Signal Separation*, San Diego, CA, Dec. 2001, pp. 194–199.