DESIGN OF DECISION DEVICE FOR THE ADAPTATION OF DECISION DIRECTED EQUALIZERS

Atulya Yellepeddi

Electrical Eng./Applied Ocean Physics and Eng.
MIT-WHOI Joint Program
email: atulya@mit.edu

James C. Preisig

Applied Ocean Physics and Engineering
Woods Hole Oceanographic Institution
email: jpreisig@whoi.edu

ABSTRACT

Adaptation of the coefficients of an equalizer usually takes place in the decision directed mode. The decision used for adaptation has a profound influence on the performance of the equalizer. In this work, we consider the influence of the probabilistic model used by a soft information decision device on the performance of the equalizer. It is shown that the equalizer output is distributed as a Gaussian Mixture, and these statistics are used to form a soft-decision device. We show that this is very near optimal in the sense of the Recursive Expected Least Squares algorithm introduced in prior work. Sub-optimal decision devices that are often more practical are also introduced and the loss of performance as a result of using these is considered.

***Index Terms*—** Equalizers, adaptive signal processing, adaptive equalizers, time-varying channels

1. INTRODUCTION AND REVIEW

Adaptive equalizers are one of the key components of communication systems for time-varying channels. A number of algorithms may be used to adapt the coefficients of an equalizer- typical examples include the LMS and RLS algorithms [1]. However, adaptation algorithms typically require knowledge of the transmitted symbols in order to operate reliably. Thus, in practice, equalizers operate in the hard-decision directed mode after an initial training period.

Prior work [2] and experience have shown that hard-decision directed equalizers perform poorly, especially at lower SNRs or with rapidly varying channels. Therefore, other approaches have been sought. In particular, various soft-decision devices were designed in [3–7] among others. These techniques utilize algorithms such as the constant modulus algorithm, or heuristically designed soft-decision devices.

In [8], an algorithm was derived for adaptation that was shown to be optimal in an EM sense. This algorithm, termed the Recursive Expected Least Squares (RELS) algorithm, essentially involves replacing the hard decision in the RLS algorithm with a soft decision that depends on the statistics of the output of the equalizer.

The equalizer output statistics used to form the decision device in [8] were chosen for simplicity of implementation and to develop connections to the blind equalization techniques of [3] and [5]. In this work, we consider a soft decision which reflects the actual statistics of the equalizer output, and suboptimal approximations to this. We compare and contrast the performances of the adaptation process when driven by these different decisions.

The output statistics of the equalizer are useful in a variety of other contexts. For instance, in [9] the statistics (approximated by Gaussians) are used in Bit-Interleaved Coded Modulation for CDMA. Output statistics are also used in computing symbol probabilities in turbo equalization, and Gaussian distributed statistics are often assumed for this purpose [10]. Hence it is useful in such applications to see how much is lost by such Gaussianity approximations. In this work, we consider the effect on adaptation, but it may also provide insights into what might be expected in other scenarios.

Notation: Boldface lowercase math symbols (e.g., $\boldsymbol{x}$) are vectors and boldface uppercase math symbols (e.g., $\boldsymbol{X}$) are matrices. † is Hermitian of a matrix or vector, ^{T} is transpose and * is complex conjugation. $\mathrm{p}(x \mid y = y_0; z)$ is the probability distribution of x, *conditioned* on $y = y_0$, and *parametrized* by z. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the PDF of a circularly symmetric complex normal distribution.

2. SYSTEM MODEL

A block diagram of the system is shown in Figure 1. The transmitted signals $s(n)$ are drawn independently and uniformly from a finite, zero-mean, unit-variance constellation $\mathcal{S}$. They are passed through a causal, finite ISI channel $\boldsymbol{h}$, which may be time varying in general. The channel outputs are corrupted by additive noise $v(n)$ and passed to a serial-parallel converter. The converter also receives the past transmitted symbols and produces the overall input vector to the equalizer, denoted $\boldsymbol{u}(n)$, of length N. The past symbols are included in $\boldsymbol{u}(n)$ to account for the possibility that the equalizer may be a DFE. It is assumed that the feedback filter receives the correct transmitted symbols. The effect of errors in the feedback filter has been investigated in, for instance [11–13].

The equalizer filter is denoted $\mathbf{g}(n|n-1)$, which represents the coefficients used to filter the input at time n, computed given all the data upto time $n-1$ [14] . Define P as the total number of transmitted symbols on which the equalizer input depends, and let

$$\boldsymbol{s}_P(n) = \big[s(n),\, \dots\, , s(n-(P-1))\big]^T \tag{1}$$

Then the overall input to the equalizer at time n may be written as:

$$\boldsymbol{u}(n) = \boldsymbol{H}^\dagger \boldsymbol{s}_P(n) + \boldsymbol{v}(n) \tag{2}$$

where $\boldsymbol{H}$ is the $P \times N$ channel transfer matrix, which may vary with time in general. In this work, we will assume that $\boldsymbol{H}$ is fixed, but the equalizer does not have this information. Finally, $\boldsymbol{v}(n)$ is assumed to be distributed as $\mathcal{CN}(\mathbf{0}, \boldsymbol{S})$.

Work supported by ONR Grants #N00014-09-10540, N00014-07-10738, N00014-11-10426, NSF Grant #ECCS-1102156, and the WHOI Academic Programs office

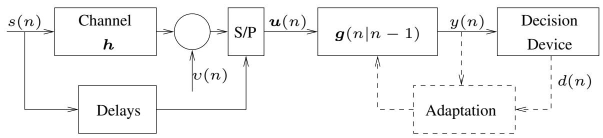


Fig. 1. System Model: Block Diagram

2.1. Adaptation

The RLS algorithm [1] is used as a starting point. With $\boldsymbol{k}(n)$ defined as the Kalman gain vector (which can be computed from the data), the RLS algorithm updates the coefficient vector by the following equation:

$$\boldsymbol{g}(n+1|n) = \boldsymbol{g}(n|n-1) + \boldsymbol{k}(n)(d(n) - \boldsymbol{g}^{\dagger}(n|n-1)\boldsymbol{u}(n))^{*} \quad (3)$$

$d(n)$ is the so-called *desired symbol* that drives adaptation. Optimally, therefore, $d(n) = s(n)$. The equalizer is said to be in "training mode" when $s(n)$ is known. Typically, however, the decision device of Figure 1 computes a value for $d(n)$ from $y(n)$. We say then that the adaptation is decision directed.

The decision device that is most commonly used is the hard-decision device:

$$d(n) = \arg\min_{s \in \mathcal{S}} |s - y(n)|^2 \quad (4)$$

When $d(n)$ is given by (4), the adaptation is said to be hard-decision directed.

In [8], an improved algorithm for adaptation in decision directed mode was derived, termed the *Recursive Expected Least Squares* algorithm. The RELS update step is (3), with $d(n)$ given by:

$$d(n) = \mathbb{E}[s(n) \mid y(n), \ldots, y(1) \; ; \; \boldsymbol{g}(n|n-1)] \quad (5a)$$

$$= \frac{\sum_{s(n)\in\mathcal{S}} s(n) \mathrm{p}(y(n) \mid \boldsymbol{y}(n-1), s(n) \; ; \; \boldsymbol{g}(n|n-1))}{\sum_{s(n)\in\mathcal{S}} \mathrm{p}(y(n) \mid \boldsymbol{y}(n-1), s(n) \; ; \; \boldsymbol{g}(n|n-1))} \quad (5b)$$

where $\boldsymbol{y}(n-1) = \left[y(n-1), \ldots, y(1)\right]$ and to obtain (5b) from (5a) we use the fact that symbols are equiprobable (extensions to non-equiprobable cases are considered in [14] in the context of turbo equalization). It is necessary to choose a suitable model for $\mathrm{p}(y(n) \mid \boldsymbol{y}(n-1), s(n) \; ; \; \boldsymbol{g}(n|n-1))$. There are several considerations here.

2.1.1. *Choosing a Model*

Knowing $\boldsymbol{y}(n-1)$ provides some information about the past symbols $s(n-1), \ldots, s(1)$, and hence $y(n)$. But $\boldsymbol{y}(n-1) \in \mathbb{C}^{n-1}$, so the size of the conditional distribution is large, and grows with time. It is thus infeasible to work with this distribution- hence, we approximate $\mathrm{p}(y(n) \mid \boldsymbol{y}(n-1), s(n) \; ; \; \boldsymbol{g}(n|n-1)$ by $\mathrm{p}(y(n) \mid s(n) \; ; \; \boldsymbol{g}(n|n-1))$.

With this assumption, in [8], the model chosen was $y(n) \sim \mathcal{CN}(s(n), \sigma^2)$, where σ, in general, depends on $\boldsymbol{g}(n|n-1)$. If it were further assumed that the statistics are stationary, σ would be fixed, and can thus be estimated during a training period. In practice, we can choose the model as though the statistics are stationary and then track the parameters in the cases where we expect non-stationarity.

However, the true statistics of the equalizer output are not Gaussian. In the next section, we derive an expression for the output statistics, which will lead to a soft-decision device. We continue to assume that the statistics of $y(n)$ are stationary given $s(n)$, with the understanding that we can just track the relevant model parameters when they are not. Which parameters need to be tracked, of course, depends on the model.

3. STATISTICS OF EQUALIZER OUTPUT

To derive the expression for $\mathrm{p}(y(n) \mid s(n) \; ; \; \boldsymbol{g}(n|n-1))$, let $\boldsymbol{g} \equiv \boldsymbol{g}(n|n-1)$ and observe, for any $s_0 \in \mathcal{S}$:

$$\begin{aligned} &\mathrm{p}(y(n) \mid s(n) = s_0 \; ; \; \boldsymbol{g}) \\ &= \sum_{\boldsymbol{s}\in\mathcal{S}^{P-1}} \mathrm{p}(\boldsymbol{s}) \mathrm{p}\left(y(n) \,\middle|\, \boldsymbol{s}_P(n) = \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix} \; ; \; \boldsymbol{g}\right) \\ &= \frac{1}{|\mathcal{S}|^{P-1}} \sum_{\boldsymbol{s}\in\mathcal{S}^{P-1}} \mathrm{p}\left(y(n) \,\middle|\, \boldsymbol{s}_P(n) = \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix} \; ; \; \boldsymbol{g}\right) \end{aligned} \quad (6)$$

As $\boldsymbol{v}(n) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{S})$, conditioned on $\boldsymbol{s}_P(n)$, we have:

$$\mathrm{p}\left(\boldsymbol{u}(n) \,\middle|\, \boldsymbol{s}_P(n) = \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix}\right) = \mathcal{CN}\left(\boldsymbol{H}^{\dagger} \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix}, \boldsymbol{S}\right) \quad (7)$$

As linear combinations of complex Gaussians are complex Gaussian variables, and $y(n) = \boldsymbol{g}^{\dagger}(n|n-1)\boldsymbol{u}(n)$, we have:

$$\mathrm{p}\left(y(n) \,\middle|\, \boldsymbol{s}_P(n) = \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix} \; ; \; \boldsymbol{g}\right) = \mathcal{CN}\left(\boldsymbol{g}^{\dagger}\boldsymbol{H}^{\dagger} \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix}, \boldsymbol{g}^{\dagger}\boldsymbol{S}\boldsymbol{g}\right) \quad (8)$$

Using (8) in (6), we have the output statistics of the equalizer, given by:

$$\begin{aligned} &\mathrm{p}(y(n) \mid s(n) = s_0 \; ; \; \boldsymbol{g}) \\ &= \frac{1}{|\mathcal{S}|^{P-1}} \sum_{\boldsymbol{s}\in\mathcal{S}^{P-1}} \mathcal{CN}\left(\boldsymbol{g}^{\dagger}\boldsymbol{H}^{\dagger} \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix}, \boldsymbol{g}^{\dagger}\boldsymbol{S}\boldsymbol{g}\right) \end{aligned} \quad (9)$$

This is an exact expression for the output statistics. We consider an approximation to these statistics given by the "best-fit" Gaussian, i.e., the Gaussian with the mean and variance of the distribution of (9). These are given by:

$$\mathbb{E}[y(n) \mid s(n) = s_0 \; ; \; \boldsymbol{g}] = \boldsymbol{g}^{\dagger}\boldsymbol{H}^{\dagger} \begin{bmatrix} s_0 \\ \boldsymbol{0}_{(P-1)\times 1} \end{bmatrix} \quad (10a)$$

$$\begin{aligned} &\mathrm{var}(y(n) \mid s(n) = s_0 \; ; \; \boldsymbol{g}) \\ &= \boldsymbol{g}^{\dagger}\left(\boldsymbol{S} + \boldsymbol{H}^{\dagger} \begin{bmatrix} 0 & \boldsymbol{0}_{1\times(P-1)} \\ \boldsymbol{0}_{(P-1)\times 1} & \boldsymbol{I}_{P-1} \end{bmatrix} \boldsymbol{H}\right)\boldsymbol{g} \end{aligned} \quad (10b)$$

Observe from (10a) that the assumption used in the model of [8], that the mean of $y(n)$ conditioned on $s(n) = s_0$ is s_0 is sub-optimal. It is easy to show that, with high probability, the mean has a smaller magnitude than the symbol and tends to s_0 as the SNR goes to infinity. We do not include this proof for reasons of space.

In order to validate these statistics, consider an equalizer in the training mode (for a fixed channel). It is known [1] that in the steady state, the mean of the coefficient vector is the L-MMSE coefficient vector $\boldsymbol{g}_0$, and that the variance of the coefficient vector about its mean is small. Thus, in (9) and (10) we set $\boldsymbol{g} = \boldsymbol{g}_0$. For an equalizer with BPSK signaling, Figure 2 shows the simulated distribution (normalized histogram) of the output $y(n)$ corresponding to $s(n) = 1$ and the derived models.

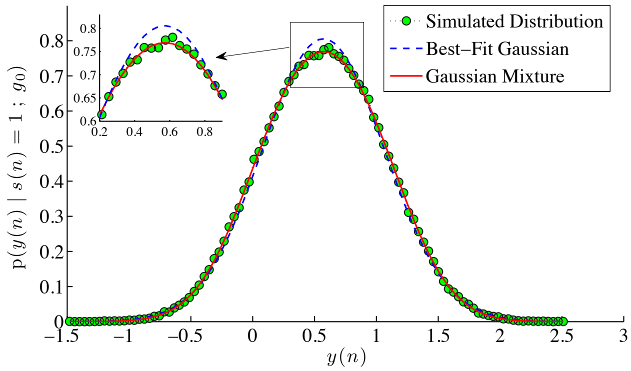


Fig. 2. Distributions of Equalizer Output in Training Mode, Validating the Output Statistics of the Equalizer

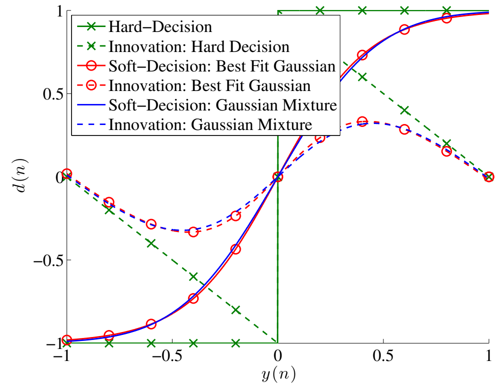


Fig. 3. Decision Devices and Corresponding Innovations for Various Models Considered

Evidently, the Gaussian Mixture distribution of (9) is nearly exact, even using $\boldsymbol{g} = \boldsymbol{g}_0$. This indicates, even in practice, that the "stationary output" assumption may be sufficient. We will demonstrate this again using simulation. In practice, of course, if we made the stationary statistics assumption, rather than assume that $\boldsymbol{g} = \boldsymbol{g}_0$, we would estimate the parameters during a training period. In this work, ML parameter estimation is employed for means and variances of Gaussian and the means and variances of the components in Gaussian Mixtures.

Further, the Gaussian distribution is a reasonable approximation to the mixture distribution. Thus, we can use these 2 output distributions (with the stationary statistics assumption or with parameter tracking) to form improved decision devices.

3.1. Soft-Decision Devices Using Output Statistics

The derivation above gives us 2 choices of the decision function to use in adaptation, corresponding to the 2 models (the exact Gaussian Mixture distribution and the "best-fit" Gaussian distribution) for the equalizer output. We now consider the decision functions that arise as a result of using these models in (5b). It seems evident that the best, or optimal soft-decision directed device would be the one that uses the exact output statistics- i.e., the Gaussian Mixture model.

Consider a BPSK system with a fixed channel. Suppose the equalizer has been run in training mode for sufficiently long, so that the coefficient vector is in a steady state. Then the equalizer coefficient vector can be approximated by the L-MMSE coefficients $\boldsymbol{g}_0$. At this point, we switch into decision directed mode. Figure 3 demonstrates the operation of the decision devices for the hard-decision directed device and the soft decision devices using the best-fit Gaussian represented by (10) and the Gaussian Mixture model of (9) at this time. The Gaussian model of 2.1.1 is not included because for this channel and equalizer, this model and the best fit Gaussian model lead to nearly identical decision devices, and seeing the difference is hard. Also included on the same plot is the innovation, $d(n) - y(n)$, which represents the amount of change in the coefficient vector.

From Figure 3 it is clear why the RELS algorithm with the various models would perform better than hard-decision directed adaptation- the innovation when $y(n)$ is near 0 is small, which is logical, as such values correspond to the largest probability of error. Moreover, it should be noted that while the Gaussian Mixture distribution has a smaller innovation (as compared to the best-fit Gaussian distribution) when $y(n)$ is small, and a larger one when it is close to 1, the difference between the decision functions is not huge, at least for this channel/equalizer.

4. SIMULATION RESULTS

We now compare the performance of decision-directed equalizers using the different decision devices introduced. The channels used are drawn from [15], specifically, Figure 10-2-5 of the book, and are summarized in Table 1.

Figure 4a shows the performances of the soft-decision devices using the Gaussian Mixture of (9) and the best-fit Gaussian with the parameters of (10), with $\boldsymbol{g} = \boldsymbol{g}_0$ in both cases. The results show that the Gaussian Mixture distribution is indeed better, as expected than the Best-Fit Gaussian. Moreover, comparing with the training mode performance provides a justification for the stationary output statistics assumption, even in practice. Hence, the fact that assuming the statistics are stationary makes the system simple (as it elimates the need for parameter tracking) and seems to perform reasonably well.

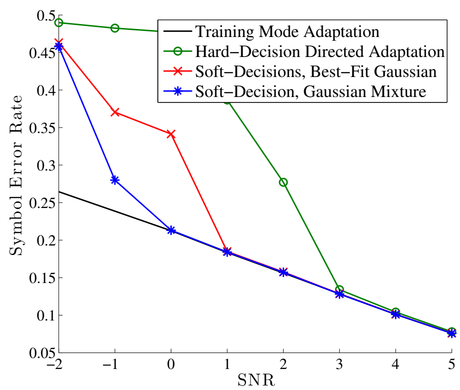
The results for the same system, but with the parameters estimated during training (when we say "estimated parameters" we mean the parameters of the distribution are estimated, as discussed previously), are shown in Figure 4b. In this figure we also consider the soft-decision directed device of [8], which uses the Gaussian distribution with means at symbols. It should be noted that all the soft-decision directed adaptation schemes outperform the hard-decision directed adaptation. Further the Best-Fit Gaussian (in which both the mean and variance are estimated) and the Gaussian Mixture model perform almost equally well, and their lines cannot be distinguished in the figure. The Gaussian with assumed means at symbols performs slightly worse that these at low SNRs, but their performance converges at high SNRs.

System B shows similar trends, in general. The performances of the Gaussian Mixture and Best-Fit Gaussian based soft-decision directed equalizers is better than the one that uses Gaussian decisions with means at symbols. However, at low SNRs in this case we see that the Best-Fit Gaussian, which is an approximation to the Gaussian Mixture, actually outperforms the latter, contradicting Figure 4a!

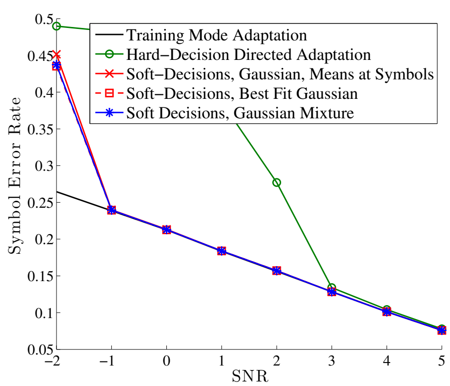
The reason is that, the more components there are in the mixture ($|\mathcal{S}|^{P-1}$), the more data is needed to estimate the parameters. Moreover, the lower the SNR, the more data is needed to get reliable estimates (intuitively, more noise needs to be averaged out). The

Channel	Impulse Response	FF Length	FB Length
A	$[0.81, 0.42, 0.42]$	2	2
B	$[0.69, 0.46, 0.46, 0.23, 0.23]$	3	2
C	$[0.72, -0.5, 0.36, 0.21, 0.21$ $0.07, 0.05, 0.03, 0.04, 0.07]$	10	5

Table 1. Impulse Responses of Channel and Parameters of Equalizer



(a) Fixed Parameters



(b) Estimated Parameters

Fig. 4. System A- Performance with Various Adaptation Schemes

performance of the soft-decision directed adaptation with the Gaussian Mixture model depends on the accuracy of the model. Thus, the inaccurate exact model does not do as well as the approximate model.

Finally, for the much longer System C, it is impractical to fit a Gaussian Mixture distribution, due to the computational complexity of finding the parameters. With time-varying channels, and when the channel length is unknown, this problem is even harder. The 2 different Gaussian approximations, which are practical, perform as before- the high SNR performance is nearly identical, and at low SNRs the Best-Fit procedure, which estimates the means, does better. However, the Best-Fit procedure requires the estimation of 2 parameters during training, and thus, requires more data to get accurate estimates. This might become an issue in non-stationary environments, or if not much training data can be used.

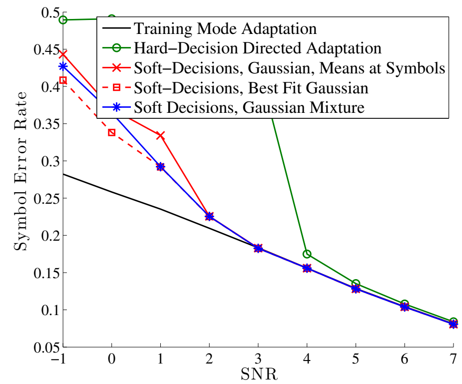


Fig. 5. System B- Performance with Various Adaptation Schemes, Estimated Parameters

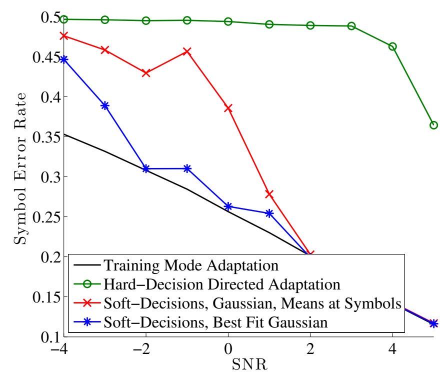


Fig. 6. System Cs- Performance with Gaussian Approximations-Best-Fit and Means at Symbols

5. CONCLUSIONS

The effect of the statistical model used by the decision device on the performance of adaptation algorithms was examined. The optimal decision device would depend on the distribution of the equalizer output, which is a Gaussian Mixture. The Best-Fit Gaussian statistics for the output statistics was derived. We then showed how the decision devices that use these models behave, and compared the performance of these, versus the hard-decision directed equalizer and the soft decision device of [8] that uses a Gaussian with means at symbols as the model.

While the Gaussian Mixture does better, it does not have a significant advantage over the Best-Fit Gaussian in practice. Further, estimating the parameters of the mixture is complicated, and may not even be possible with time-varying systems. Thus, the Best-Fit Gaussian does acceptably well in terms of both complexity and performance. The Gaussian with the "means at symbols" assumption is reasonable at high SNRs.

However, all of the work herein with practical systems assumed stationary output statistics. Performance could be improved by tracking the model parameters. Better estimators for the parameters and simple ways of taking advantage of the true output statistics could be explored. The field of adaptation with soft information and reliable adaptation algorithms with limited data is itself still nascent and much further exploration is possible.

6. REFERENCES

[1] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, 4 edition, Sept. 2001.

[2] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *Communications, IEEE Transactions on*, vol. 28, no. 11, pp. 1867–1875, 2002.

[3] Sandro Bellini, "Bussgang techniques for blind equalization," in *IEEE Global Telecommunications Conference, Conference record*, 1986, pp. 1634–1640.

[4] J.B.D. Filho, G. Favier, and J.M.T. Romano, "New Bussgang methods for blind equalization," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, 1997, vol. 3, pp. 2269–2272 vol.3.

[5] S. J Nowlan and G. E Hinton, "A soft decision-directed LMS algorithm for blind equalization," *Communications, IEEE Transactions on*, vol. 41, no. 2, pp. 275–279, 2002.

[6] J. Karaoguz and SH Ardalan, "A soft decision-directed blind equalization algorithm applied to equalization of mobile communication channels," in *Communications, IEEE International Conference on*. IEEE, 1992, pp. 1272–1276.

[7] S. Chen, "Low complexity concurrent constant modulus algorithm and soft decision directed scheme for blind equalisation," in *Vision, Image and Signal Processing, IEE Proceedings-*. IET, 2003, vol. 150, pp. 312–20.

[8] A. Yellepeddi and J.C. Preisig, "Direct-form adaptive equalization using soft information," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*. IEEE, 2011, pp. 461–468.

[9] D. Seethaler, G. Matz, and F. Hlawatsch, "An efficient mmse-based demodulator for mimo bit-interleaved coded modulation," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, 2004, vol. 4, pp. 2455–2459 Vol.4.

[10] R. Otnes and M. Tuchler, "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *Wireless Communications, IEEE Transactions on*, vol. 3, no. 6, pp. 1918–1923, 2004.

[11] J.E. Smee and N.C. Beaulieu, "Error-rate evaluation of linear equalization and decision feedback equalization with error propagation," *Communications, IEEE Transactions on*, vol. 46, no. 5, pp. 656–665, 1998.

[12] M. Reuter, J.C. Allen, J.R. Zeidler, and R.C. North, "Mitigating error propagation effects in a decision feedback equalizer," *Communications, IEEE Transactions on*, vol. 49, no. 11, pp. 2028–2041, 2001.

[13] A. Fertner, "Improvement of bit-error-rate in decision feedback equalizer by preventing decision-error propagation," *Signal Processing, IEEE Transactions on*, vol. 46, no. 7, pp. 1872–1877, 1998.

[14] A. Yellepeddi and J.C. Preisig, "Soft-Adaptive Turbo Equalization- Using soft information in adaptation," in *Signals, Systems and Computers, 2012 46th Annual Asilomar Conference on*. IEEE, 2012, To be published.

[15] J.G. Proakis, *Digital Communications*, vol. 4, McGraw-Hill New York, 2001.