# On the Learning Behavior of Decision Feedback Equalizers

Markus Rupp
Bell-Labs, Lucent Technologies, Wireless Research Lab.
791 Holmdel-Keyport Rd., Holmdel, NJ 07733-0400, USA
Fax: (732) 888 7074, Tel: (732) 888 7104
e-mail:rupp@lucent.com

## Abstract

The performance analysis of DFE equalizers, either fractional or symbol-spaced, is commonly based on the Wiener solution. This solution can only be computed once characteristic values of the channel and noise variance are known. In the "real" world, the solution needs to be estimated. In low-bit rate systems, complexity of algorithms is usually not an issue and Least-squares solutions approximating the Wiener solution with high accuracy are possible. For higher bit rate systems however, a gradient-type procedure like the LMS algorithm seems unavoidable. Given a training sequence of limited length, the learning behavior of LMS can considerably worsen the performance of the DFE. This paper shows some insight of undesired effects.

## 1 Introduction

Performance analyses of equalizers [1]-[3] compare the equalizer behavior based on their optimal Wiener solution. In receiver applications this solution is, however only achieved by an estimation. Experiments with higher digital modulation schemes show that the higher the symbol alphabet, the more precise the DFE solution needs to be.

The following paper deals with a fixed wireless loop scenario. The channel is assumed to be Ricean with an exponential power profile. Data rates are about 4MSPS for digital modulation schemes varying from QPSK to 64QAM. With a time dispersion of $\tau = 125ns$ assumed, a T/2-fractionally spaced decision-feedback equalizer with four feed-forward and three feedback taps shows promising results when looked at the Wiener solution. A relatively short training sequence of 13 symbols with data reuse of a few retraining periods showed almost Wiener solution properties at QPSK in the operating range of low Signal-to-Noise-Ratios (SNR). For higher modulation schemes, however, a constant Bit Error Rate (BER) shows up in their operating SNR ranges. Figure 1 displays this behavior for various modulation schemes. The continues lines are 64QAM,16QAM,8PSK and QPSK in the order from the upper to the lower lines. The corresponding dotted lines are the results when using the Wiener solution.

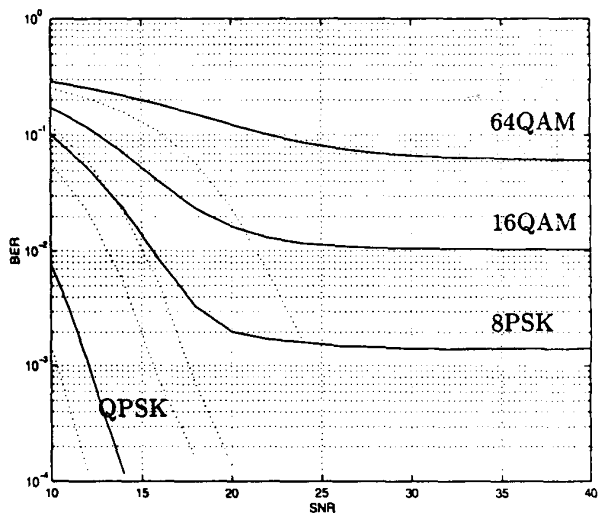


Figure 1: *BER Performance for various modulation schemes on 8dB Ricean channels and 13 symbol training sequence.*

## 2 Channel Model

The following channel model was inspired by [4]. The performance of such simulations is dependent on the chosen channel models. A Ricean channel model with exponential profile

$$c(t) = e^{-t/\tau} \tag{1}$$

the time dispersion being $\tau = 125$ns is used throughout this paper. The specular component $K$ is 8dB (3dB if mentioned). The simulations are run oversampled in $T/4$; the channels are normalized to gain one so that the SNR can be selected by choosing a specific noise variance. The channel remains constant for each run.



0-7803-5700-0/99/$10.00©1999 IEEE

The Ricean channel is generated with a random generator and a threshold variable $t_h$. The latter defines how many rays $N_r$ exist:

$$N_r = \lceil \frac{\tau}{\ln \frac{1}{t_h}} \rceil .$$

A power profile is then normalized to one:

$$p_i = \frac{e^{i/\tau}}{\sum_{i=0}^{N_r-1} e^{i/\tau}}, \quad i = 0..N_r - 1$$

and the channel impulse response generated by

$$\begin{aligned} c_0 &= \sqrt{K} + \sqrt{p_0} v(i) & (2) \\ c_i &= \sqrt{p_i} v(i), \quad i = 1..N_r - 1 & (3) \end{aligned}$$

using the random complex valued Gaussian numbers $v(i)$ with unit variance. In a final step, all coefficients are normalized by $1 + \sqrt{K}$.

## 3 DFE and LMS

For such high data rates only a simple algorithm that can be implemented in fixed-point can be utilized. Since a Least-Squares realization in fixed-point does not seem possible without larger hardware complexity, it was therefore decided to realize an LMS algorithm with data reuse (UNDR-LMS in [5]) on the training sequence. It was favorable to decrease the step-size by a constant factor $\lambda$ after each run through the set of training sequence. Theory claims that for infinite runs on new data the impact of noise can be removed by applying a step-size sequence that decreases by $1/k$. Experiments showed however, that in these cases of limited updates with data reuse the exponential decaying step-size is better suited and easier to implement.

The question whether a symbol-spaced or fractionally spaced equalizer is to choose was decided many years ago in favor for the fractionally spaced type since it is very robust against timing estimation errors. The following section will however show that this structure does not come with an extra price attached. Also the selection of one of the two structures has an impact on the whole receiver. See [8] for more details.

## 4 Some Theory on Learning

From LMS theory [6] it is well known that the initial learning behavior is given in form of a linear system whose eigenvalues describe the speed of convergence. Each eigenvalue corresponds to a possible mode. Naturally, the largest eigenvalue describes the slowest learning rate, however, if only small amounts of the corresponding eigenvector exist in the final solution, this eigenvalue might show its slow rate at a much later point in time while other eigenvalues dominate the behavior first.

The initial values of the equalizer have a similar effect. By choosing different initial values, the impact of a particular eigenvector can be increased or decreased. Since the channel and with this the Wiener solution is not known a-priori, it does not seem possible to find an optimum initial value. Given a certain power profile, especially with a large specular component $K$, one might expect that there exists an average Wiener solution which is a good initial estimate. This is true if the transmitter and receiver maintain a constant phase relation, i.e., in a coherent receiver environment. Since such a setup is not expected, the initial estimates were set to zero in all simulations presented here.

Many parameters define the set of eigenvalues, among them the step-size. For a given channel $C$ and noise variance, it can be shown that there exists one optimal step-size $\mu$ leading to fastest learning rate. In general two eigenvalue spreads are distinguished:

$$\begin{aligned} \chi_0 &= f(M, J, C) & (4) \\ \chi(\text{SNR}) &= g(\text{SNR}, \chi_o) & (5) \end{aligned}$$

with $M$ and $J$ being the number of feed-forward and feed-back coefficients, respectively. The first value $\chi_0$ is the eigenvalue spread at the input of the receiver assuming zero noise. It is influenced by the structure of the receiver, in particular the choice of the parameters $M$ and $J$ and the channel $C$. The second value $\chi(\text{SNR})$ is the eigenvalue spread when noise is present. The noise adds on the feed-forward part of the receiver, however, not on the feedback part. In the Wiener solution this can be seen as additional noise terms on the autocorrelation matrix of the feed-forward part. Let us assume for reasons of simplification a diagonal autocorrelation matrix, with the eigenvalues ordered from smallest to largest ($\lambda_{\min}..\lambda_{\max}$). The eigenvalue spread $\chi_o$ is thus given by $\lambda_{\max}/\lambda_{\min}$. After adding the noise, it changes into

$$\chi(\text{SNR}) = \frac{\lambda_{\max} + 1/\text{SNR}}{\lambda_{\min} + 1/\text{SNR}} \qquad (6)$$

Given $\chi_0$, or more precisely the original largest and smallest eigenvalue, the eigenvalue spread $\chi(\text{SNR})$ can be computed. Since the spread is large, good approximations are $\lambda_{\max} = M + J$, $\lambda_{\min} = (M + J)/\chi_0$.

Figures 2 and 3 display values $\chi(\text{SNR})$ for fixed feedback parameter $J = 3$ and forward parameter $M = 4$, respectively. The results are obtained for 10 Ricean channels with 8dB K-factor. Both figures show the typical behavior that the eigenvalue spread increases with SNR to dramatically high values. In particular for varying the feed-forward order $M$, very distinct curved were obtained. Several conclusions can be drawn from these experiments:

1. For a fractionally spaced equalizer, the eigenvalue spreads $\chi_o$ and $\chi$ do not depend much on the actual channel but more on the number of parameters in the feed-forward path and the SNR. See also eqn. (13)-(16) for a more detailed discussion.
2. The relation of eigenvalue spread $\chi(\text{SNR})$ over the signal-to-noise ratio SNR is well described by (6), i.e., with increasing SNR the eigenvalue spread increases as well.
3. The results do not depend much on the actual channel. For smaller K-factors of the Ricean channels, the variations around the AWGN curves become larger, but do not increase in average or change the behavior described by (6).

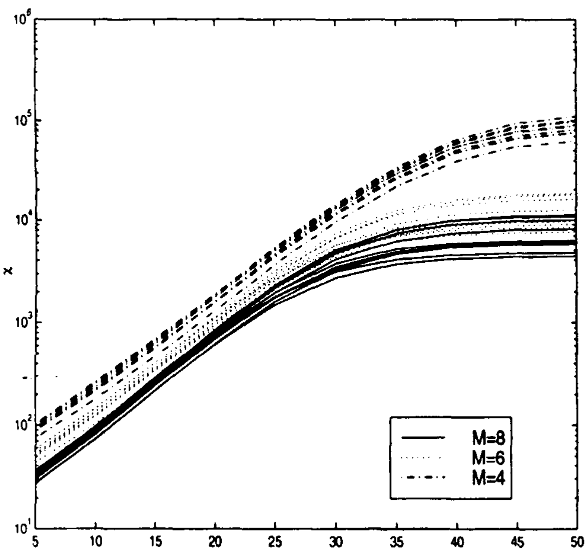

Figure 2: *Eigenvalue spread of the input process for fractional* $T/2$ *DFE* $(M/2, 3)$ *with values* $M = 4, 6, 8$.

How does the learning rate relate to the eigenvalue spread? Bounds exist in the literature [7] that can be used for gaining insight. The maximum eigenvalue $\lambda_{\max}^{(B)}$ of a specific matrix defines the slowest learning behavior. It lies between:

$$1-\frac{2\alpha}{M+J-1+\chi} \leq \lambda_{\max}^{(B)} \leq 1-\frac{2\alpha}{1+(M+J-1)\chi} \quad (7)$$

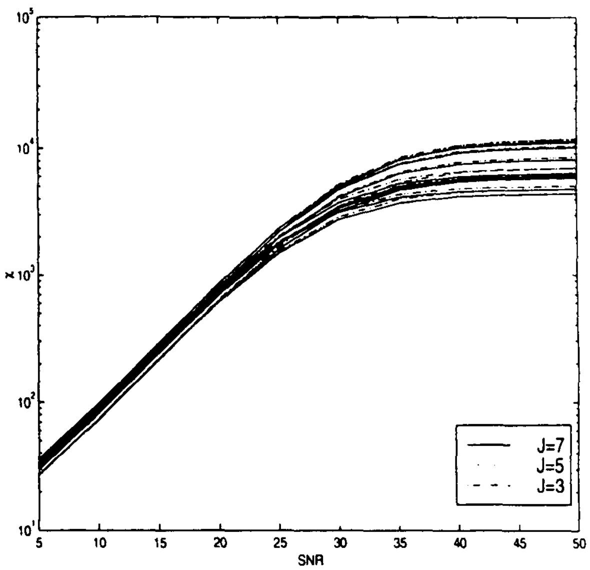

Figure 3: *Eigenvalue spread of the input process for fractional* $T/2$ *DFE* $(4/2, J)$ *with values* $J = 3, 5, 7$.

with $\alpha = \mu M \sigma_u^2$, the normalized step-size. To simplify matters, the average of the two bounds is taken and

$$\lambda_{\max}^{(B)} \approx 1-\frac{\alpha M(1+\chi)}{[M+J-1+\chi](1+[M+J-1]\chi)} \quad (8)$$

obtained which can further be approximated by

$$\lambda_{\max}^{(B)} \approx 1 - \frac{\mu}{\text{SNR}} \quad (9)$$

for large enough eigenvalue spread $\chi \gg M + J$ in the range of SNR from 10dB to 40dB, i.e. only valid for the ramp-part of the curve in Figures 2 and 3. An example shows the relations. Assume for $M = 4, J = 3, \mu = 1/7$. After 60 learning iterations and an SNR of 10dB an improvement (relative system mismatch) of 0.066 is obtained while for 30dB it is only 0.77, about 12 times worth. Figures 2 and 3 clearly show that for high SNR, it cannot be expected that the Wiener solution is obtained. Even after long training periods this would not be the case and in particular for high SNR the learning result is expected to be poor.

### 4.1 Non-Fractionally Spaced Equalizer Learning

Can a non-fractionally $T-$spaced equalizer have better learning performance?

The answer is yes, under certain conditions. Figures 4 and 5 display the corresponding eigenvalue spreads when run on a K=8dB Ricean channel. They are much smaller and not as distinguished in the parameters as is the case for a fractionally spaced equalizer. However, these small eigenvalue spreads are only obtained when the cursor position is close to an optimum position. Once this condition is violated, the eigenvalues can become as big as several thousand. Assuming perfect cursor positioning, it is thus to expect that $T-$spaced equalizers show better learning. Note, however, that better learning is **not** necessarily equivalent with more accurate estimates, smaller MMSE or lower SER(BER). It only means that an estimate to the (Wiener-) solution is found in a more rapid way, not necessarily more precise.

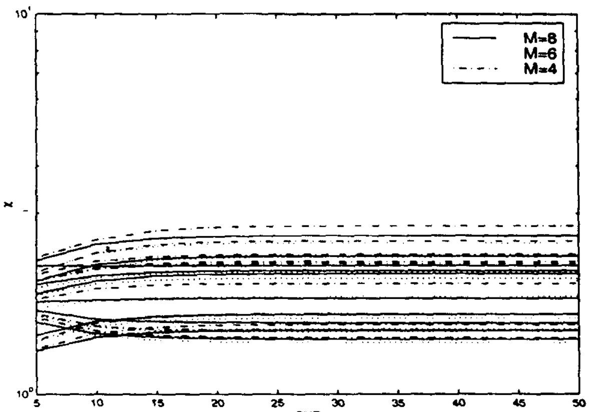


Figure 4: *Eigenvalue spread of the input process for non-fractional T DFE* $(M, 3)$ *with values* $M = 2, 4, 6$.

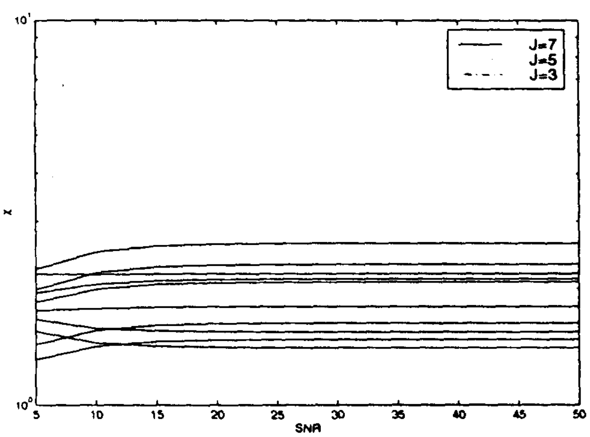


Figure 5: *Eigenvalue spread of the input process for non-fractional T DFE* $(2, J)$ *with values* $J = 3, 5, 7$.

An open question here is how the channel relates to the eigenvalue spread of the equalizer. A simple model can be assumed with two channel coefficients only: $c_0$ and $c_1 = r_2$. According to (2) $c_0 = \sqrt{K} + r_1 e^{\phi}$ includes the specular component K of the Rice channel. In other words, the channel has the form:

$$C(z) = (\sqrt{K} + r_1 e^{j\phi}) + r_2 z^{-1}. \qquad (10)$$

Assuming a T-spaced transversal equalizer, the acf matrix is of Toeplitz form with only two entries, $h_0$ and $h_1$. The eigenvalues are given by

$$\lambda_i = h_0 + 2|h_1| \cos\left(i \frac{\pi}{M+1}\right); i = 1..M. \qquad (11)$$

For (10) $h_0 = |\sqrt{K} + r_1 e^{j\phi}|^2 + |r_2|^2$ and $h_1 = (\sqrt{K} + r_1 e^{j\phi}) r_2$. Finally, the eigenvalue spread is given by:

$$\chi_o = \frac{1 + \frac{2|r_2||\sqrt{K}+r_1 e^{j\phi}|}{|\sqrt{K}+r_1 e^{j\phi}|^2+|r_2|^2} \cos(\frac{\pi}{M+1})}{1 - \frac{2|r_2||\sqrt{K}+r_1 e^{j\phi}|}{|\sqrt{K}+r_1 e^{j\phi}|^2+|r_2|^2} \cos(\frac{\pi}{M+1})}. \qquad (12)$$

Given the distributions of $r_1$, $r_2$, and $\phi$ the average eigenvalue spread can be computed over a whole set of channels for a predefined value $K$. Assuming large $M$ the cosine can be replaced by one. Assuming Rayleigh distribution for $r_1$ and $r_2$ and uniform phase distribution, the values in Table 1 are obtained and displayed in Figure 6.

Now the means to compute eigenvalue spreads for T-spaced equalizers are available. But how does the eigenvalue spread change for a decision feedback equalizer structure? To answer this question, the autocorrelation matrix needs to be written in the following form:

$$\mathbf{R}_{cc} = \begin{bmatrix} \mathbf{R}_{ff} & \mathbf{R}_{fb} \\ \mathbf{R}_{fb}^H & \mathbf{R}_{bb} \end{bmatrix} \qquad (13)$$

were the four acf blocks corresponding to the correlation in the feed-forward and the feedback section were written explicitly. For a statistically white symbol sequence $\mathbf{R}_{bb} = \mathbf{I}$, the identity matrix. The eigenvalue problem can now be written as:

$$\begin{bmatrix} \mathbf{R}_{ff} & \mathbf{R}_{fb} \\ \mathbf{R}_{fb}^H & \mathbf{R}_{bb} \end{bmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}. \qquad (14)$$

The lower part can be solved and leads to the relation

$$\mathbf{x}_2 = \frac{1}{\lambda - 1} \mathbf{R}_{fb}^H \mathbf{x}_1 \qquad (15)$$

and substituting $\mathbf{x}_2$ in the upper part of the equation finally leads to

$$\left[\mathbf{R}_{ff} + \frac{1}{\lambda - 1} \mathbf{R}_{fb}^H \mathbf{R}_{fb}\right] \mathbf{x}_1 = \lambda \mathbf{x}_1 . \qquad (16)$$

For very small eigenvalues the second term on the LHS will smaller $\mathbf{R}_{ff}$. For very large eigenvalues, $\lambda \approx M+J$ and the term in $1/(\lambda - 1)$ is of little consequence. If the cursor position is moved towards the last filter taps in the feed-forward part, the cross-correlation matrix $\mathbf{R}_{fb}$ will be small compared to $\mathbf{R}_{ff}$. Thus, as a rough approximation, it is assumed that the eigenvalue spread is basically given by the feed-forward section. Note that this condition can easily be violated and the assumption serves only to simplify matters.

Simulations on a DFE (4,3) structure and a set of 500 channels for K=8dB showed $\chi_o = 3.3$ thus, not too far away from the 11 obtained by using all these approximations. Note that this technique cannot be applied to the fractionally spaced equalizer since the feed-forward section is not of Toeplitz form and thus the derivation does not hold.

| K in dB | $\chi_o$ small $M$ | $\chi_o$ large $M$ |
|---|---|---|
| 3 | 15.8 | 1320 |
| 5 | 14.7 | 1192 |
| 8 | 11.0 | 718 |
| 10 | 8.4 | 396 |
| 30 | 1.17 | 1.2 |

Table 1: *Eigenvalue spread for various values of K.*

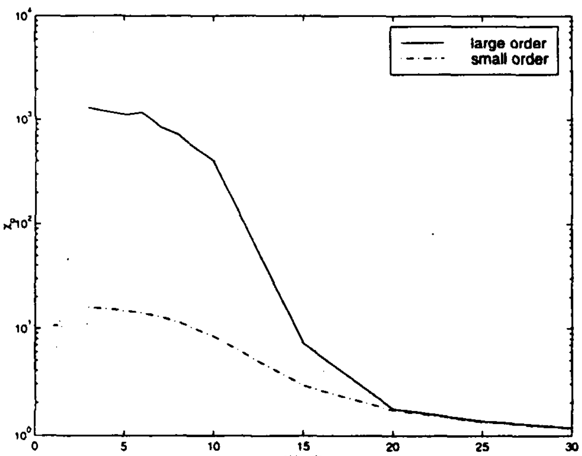


Figure 6: *Eigenvalue spread of the input process for non-fractional T DFE*

**Remark:** Note in (16) that scaling will change the eigenvalue spread considerably. Assume, for example, that the received sequence is scaled by a factor $\alpha$. Eq. (16) then reads:

$$\left[\mathbf{R}_{ff} + \frac{1}{\lambda - 1}\mathbf{R}_{fb}^H\mathbf{R}_{fb}\right]\mathbf{x}_1 = \frac{\lambda}{\alpha^2}\mathbf{x}_1 \,. \qquad (17)$$

Note that since the eigenvalue also appears on the LHS, scaling the input sequence is not equivalent to simply scaling the eigenvalue spread. Also the set of eigenvalues corresponding to $\mathbf{x}_2$ will change according to (15). Having some knowledge about the channel and noise it may be possible to find the optimal scaling for the design by prior optimizations.

## 5 Conclusion

This paper gives some insight in the tremendous performance differences that can occur when adaptive filters are used with oversampled input data compared to symbol-spaced sampling. Due to the strong correlation of the input process, simple gradient algorithms, like LMS, will not perform very well any longer. Decorrelation techniques might be useful to consider here. In particular, fractionally spaced equalizers tend to learn slower when the SNR increases while symbols spaced equalizers improve their learning with increasing SNR.

**References**

[1] J.Salz, "Optimum, Mean-Square Decision Feedback Equalizion," Bell Syst. Techn. Journal, vol. 52, pp. 1341-1373, Oct. 73.

[2] J.G.Proakis, *Digital Communications*, McGraw Hill 1995.

[3] C.Papadias, M.Rupp, "Performance analysis of finite-length DFE receivers based on a polyphase representation," 32nd. Asilomar Conference, Monterey, California, Nov. 1998.

[4] V.Erceg, et.al., "A Model for the multipath delay profile of fixed wireless channels," Journal on sel. areas in communications, vol. 17, no.3, pp. 399-410, March 1999.

[5] W.K.Jenkins, A.W.Hull, J.C.Strait, B.A.Schnaufer, X.Li, "Advanced Concepts in Adaptive Signal Processing," Kluwer Academic Publishers, 1996.

[6] S.Haykin, Adaptive Filter Theory. Englewood Cliffs, NJ: Prentice–Hall, Information and System Sciences Series, 1996.

[7] M.Rupp, *Über die Analyse von Gradientenverfahren zur Echokompensation*, VDI, No. 242, 1993.

[8] H.Meyr, M.Moeneclaey, S.A.Fechtel, "Digital Communication Receivers," Wiley & Sons, 98.