ADAPTIVE BLIND CHANNEL IDENTIFICATION OF FIR CHANNELS FOR VITERBI DECODING

Robert W. Heath Jr., Steven D. Halford, and Georgios B. Giannakis
Dept. of Electrical Engr., Univ. of Virginia, Charlottesville, VA 22903-2442, USA
Tel: (804) 924-3659, Fax: (804) 924-8818, e-mail : georgios@virginia.edu

Abstract

Relying upon the diversity introduced by fractional-sampling or by multiple sensors, two different approaches for blind adaptive channel identification are proposed based on existing batch algorithms. We show that the adaptive approach allows tracking of time-variations in the channel and provides robustness to nearly common channel roots. Existing second-order adaptive procedures typically employ suboptimal linear equalizers whereas adaptive channel estimators we propose can be used in an optimal maximum likelihood sequence estimator. Simulations are provided that demonstrate the performance of the proposed channel estimators when used in the Viterbi algorithm for symbol detection.

1. Introduction

Modern communications systems employ equalization in order to remove the intersymbol interference (ISI) generated by multiple transmission paths or by bandwidth constraints. Recently, there is considerable interest in *blind equalization* techniques which exploit the diversity introduced by fractional-sampling or by multiple sensors [13] to equalize most channels without training sequences. In contrast, without such diversity, blind equalizers need higher- (than second-) order statistics (HOS) to completely capture the phase information. In comparison with HOS based methods, second-order based methods yield lower variance channel or equalizer estimates for short data records.

Most practical communications channels (e.g. mobile radio, cellular telephony) are not static; rather, the impulse response coefficients slowly drift or undergo occasional shifts in value. In order to track or follow these changes, practical equalizers need to be adaptive. While [6] and [5] provide adaptive linear equalizers, it is well known that the lowest probability of error is achieved via the nonlinear maximum likelihood sequence estimation (MLSE), typically implemented using the Viterbi algorithm [9, Ch. 10].

To remove an ISI channel that varies with time the optimal receiver incorporates an adaptive channel estimate into the MLSE. Most works on MLSE for ISI channels are based on the seminal paper by Forney in 1973 [4]. Extensions to diversity receivers have been proposed for spatial diversity [10] and for diversity through fractional sampling [16]. These systems require training sequences in order to compute their adaptive channel estimates. However, by using a blind adaptive channel estimator, the MLSE can be found without requiring training sequences. Presently, there are no blind second-order adpative channel estimators. Blind sequence estimation is also addressed by [12] although [12] does not explicitly incorporate a channel estimate and is not adaptive.

In this work, we show two adaptive channel estimation techniques and, unlike [12], can take advantage of existing results on MLSE for ISI channels. One is based on the deterministic channel identification method described in [15] (XLTK) while the other is based on the vector-correlations [6]. The advantage of adaptive channel estimation over the batch methods of [13], [8], [11], and [14] stem largely from the fact that these adaptive procedures do not require a computationally-intensive matrix inversion or an eigendecomposition. Other advantages are three-fold: (i) ability to follow time-variations in the channel; (ii) smaller computational and memory requirements for long data records; (iii) robustness to nearly common channel roots. These adaptive channel estimates are then employed in a MLSE algorithm for detection of the input sequence with the lowest probability of error. This paper is organized as follows. First we place the problem in an appropriate mathematical framework. Then we derive two adaptive channel estimators from their batch equivalents. Next we incorporate these estimates into a MLSE for ISI decoding. Finally we show simulations which demonstrate the performance of adaptive channel estimates and show the resulting bit-error rate when these estimates are incorporated into the MLSE Viterbi decoder.

*This work was supported by ONR AASERT N00014-95-1-0908

1058-6393/97 $10.00 © 1997 IEEE

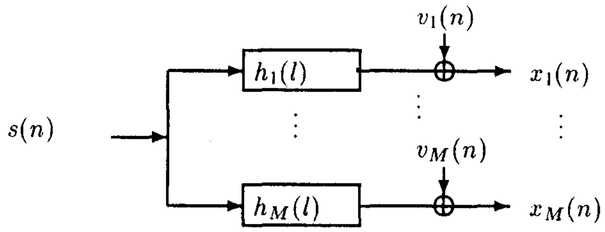

Figure 1. Single-input multiple-output system

2. Problem formulation

Consider the discrete-time single-input multiple-output (SIMO) system in Figure 1. Let $s(n)$ denote the sequence of input symbols from a finite alphabet, $\mathbf{h}(l) = [h_1(l) \ldots h_M(l)]', l \in [0, L]$ the vector of finite impulse-response (FIR) channels of order L, $\mathbf{v}(n) = [v_1(n) \ldots v_M(n)]'$ the vector of uncorrelated all white gaussian noise (AWGN) samples, and $\mathbf{x}(n) = [x_1(n) \ldots x_M(n)]'$ is the vector of channel outputs. We can then describe the output of Figure 1 mathematically by:

$$\mathbf{x}(n) = \sum_{l=0}^{L} \mathbf{h}(l)s(n-l) + \mathbf{v}(n), \quad For \;\; n = 0, 1, \ldots, N-1. \tag{1}$$

In the sequel, we will assume that this system is identifiable from its second-order statistics; in other words, the subchannels $h_i(n)$ have no common roots [13]. This vector stationary system model applies to the reception of a signal by multiple sensors followed by symbol rate sampling [15], the reception of a communications signal by a single sensor followed by fractional sampling, or by a hybrid of these two techniques. We assume that $s(n)$ is an independent identically distributed (i.i.d.) sequence from a finite alphabet, each $h_i(l)$ is the cascade of the pulse-shaping filter, multipath channel impulse response, recieve filter, and perhaps noise whitening filter [9] for the i^{th} channel, and $v_i(n)$ are from a AWG noise distribution with no cochannel correlation.

3. Description of the adaptive algorithms

In this section we show the procedure for deriving blind adaptive channel estimation algorithms from their batch equivalents. To clarify the discussion we focus on the $M = 2$ case, though an extension to $M > 2$ is straightforward.

3.1. Channel estimation based on XLTK

From Figure 1 (see also [15]) we can relate the outputs by exploiting the commutivity of the convolution operator (which we denote by $\otimes$) by writing:

$$x_1(k) \otimes h_2(k) = x_2(k) \otimes h_1(k). \tag{2}$$

Collecting equations from (2) for $k = N-L, N-L-1, \ldots, 0$ (where N is the number of collected samples) yields the following matrix equation:

$$\begin{bmatrix} \mathbf{X}_1(L) & -\mathbf{X}_2(L) \end{bmatrix} \begin{bmatrix} \mathbf{h}_2 \\ \mathbf{h}_1 \end{bmatrix} = \mathbf{0} \;, \tag{3}$$

with $\mathbf{h}_m := [h_m(L), \cdots, h_m(0)]'$, $m \in [1,2]$ and

$$\mathbf{X}_m(L) = \begin{bmatrix} x_m(0) & \cdots & x_m(L) \\ \vdots & \vdots & \vdots \\ x_m(N-L) & \cdots & x_m(N) \end{bmatrix}. \tag{4}$$

All blind identification methods suffer from a possible scale ambiguity hence, without loss of generality, we assume that $h_1(0) = 1$. This is reasonable because in practice, the unknown scale factor is typically overcome by employing automatic gain control and/or differential encoding. With this assumption, we remove the last column of $\mathbf{X}_2(L)$, forming $\tilde{\mathbf{X}}_2(L)$, and place this column on the other side, thus (3) becomes

$$\underbrace{\begin{bmatrix} \mathbf{X}_1(L) & -\tilde{\mathbf{X}}_2(L) \end{bmatrix}}_{\mathcal{X}_{N-1}} \underbrace{\begin{bmatrix} \mathbf{h}_2 \\ \tilde{\mathbf{h}}_1 \end{bmatrix}}_{\hat{\mathbf{h}}} = \underbrace{\begin{bmatrix} x_2(L) \\ \vdots \\ x_2(N) \end{bmatrix}}_{\hat{\mathbf{x}}}, \tag{5}$$

where $\tilde{\mathbf{h}}_1 := [h_1(L), \cdots, h_1(1)]'$. The batch least-squares (LS) formulation in (5) leads naturally to a recursive formulation, since the channels are coprime and assuming $N - L + 1 \geq 2L + 1$, the matrix $(N - L + 1) \times 2L + 1$ matrix $\mathcal{X}_{N-1}$ is full rank. With this assumption, we can write the following least-squares batch method for finding $\hat{\mathbf{h}}$:

$$\hat{\mathbf{h}} = (\hat{\mathcal{X}}'\hat{\mathcal{X}})^{-1} \hat{\mathcal{X}}'\hat{\mathbf{x}} \,. \tag{6}$$

At each time k, let $\eta_k = [x_1(k-L), x_1(k-L-1), \ldots, x_1(k), x_2(k-L), x_2(k-L-1), \ldots, x_1(k+1)]'$ be the input data vector and let $x_2(k)$ be the addition to $\hat{\mathbf{x}}$. Following similar derivations for the least-mean squares algorithm in [7, Ch. 9], (6) can be solved in an adaptive manner using a stochastic gradient descent algorithm. At each time $k+1$ a row η_k containing new data is added to $\hat{\mathcal{X}}_{N-1}$ in (5). An error can be formed from the difference between the predicted and the actual value of the $(k+1)^{th}$ data point $x_2(k+1)$ from the other channel as in:

$$\epsilon_{k+1} = x_2(k+1) - \eta_{k+1}' \hat{\mathbf{h}}_k. \tag{7}$$

This weighted error is then used to update the channel estimate:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + 2\mu\eta_{k+1}\epsilon_{k+1}, \qquad (8)$$

where μ is an appropriately chosen stemp size (see e.g., [7, Ch. 9]). Convergence speed will depend on the step size and the initialization. The least mean-square (LMS) algorithm described by Equations 7 and 8 provides a computationally efficient adaptive method for finding the channel $\hat{\mathbf{h}}$. In addition, because it does not require explicit matrix inversion, the LMS algorithm is robust to nearly comonzeros which would lead to numerical instability if (5) were used to directly find $\hat{\mathbf{h}}$. [2]

While more computationally intensive, (5) also lends itself to adpative implementation with a recursive least-squares (RLS) algorihtm [7, Ch. 13]. Let $\mathcal{X}_k$ correspond to that in (5) constructed from $x_1(0)\cdots x_1(k)$ and $x_2(0)\cdots x_2(k-1)$ received data. Then define the inverse of the k^{th} weighted correlation matrix $\mathbf{P}_k = [\mathcal{X}_k'\mathbf{W}_k\mathcal{X}_k]^{-1}$ where the weighting matrix $\mathbf{W}_k = \text{diag}(\ldots, \lambda^3, \lambda^2, \lambda)$, and λ is the forgetting factor. Then if we define the gain vector $\mathbf{K}_k = \mathbf{P}_k\eta_k\left[\lambda + \eta_k'\mathbf{P}_k\eta_k\right]^{-1}$, the RLS algorithm can be stated as follow. At each time k compute:

$$\begin{aligned} \epsilon_{k+1} &= x_2(k+1) - \eta_{k+1}'\hat{\mathbf{h}}_k \\ \mathbf{K}_k &= \mathbf{P}_k\eta_{k+1}\left[\lambda + \eta_{k+1}'\mathbf{P}_k\eta_{k+1}\right]^{-1} \\ \hat{\mathbf{h}}_{k+1} &= \hat{\mathbf{h}}_k + \mathbf{K}_k\epsilon_{k+1} \\ \mathbf{P}_{k+1} &= \lambda^{-1}\left[1 - \mathbf{K}_k\eta_{k+1}'\right]\mathbf{P}_k. \end{aligned} \qquad (9)$$

Tracking speed and convergence depend on the choise of λ. We note however, when initialized with the batch method, the RLS solution will converge to the LS solution provided $\lambda = 1$. For $\lambda < 1$, convergence analysis is beyond the scope of this paper.

3.2. Channel estimation from output correlations

For comparison purposes we present an adaptive channel estimation algorithm that relies on the i.i.d. properties of the input. Here we present a brief summary of the batch algorithm because the adaptive approach follows from the batch approach in a similar manner [6].

Considering the input sequence $s(n)$ as an i.i.d. random sequence with variance, $\sigma_s^2 := \text{E}\{s(n)s^*(n)\}$, the output correlation matrix can be defined as:

$$\mathbf{C}_{2x}(d) := \text{E}\{\ \mathbf{x}_L(n)\ \mathbf{x}_L^{*'}(n-d)\}\ , \qquad (10)$$

where $*$ indicates conjugation and $'$ indicates transpose (Hermetian) and $\mathbf{x}_L(n) := [x_1(n), x_1(n-1), \ldots x_1(n-L), x_2(n), x_2(n-1)]', \ldots, x_2(n-L)]'$. Arguing as in [6, Section 3], we have that the channel vector $\mathbf{h}$ can be found (to within an unknown scale) by solving:

$$\mathbf{h} = \mathbf{C}_{2x}^*(L+1)\left[\mathbf{C}_{2x}^*(0)\right]^{-1}\mathbf{k}\ , \qquad (11)$$

where $\mathbf{k} := [\sigma_s^2, 0, 0, \ldots 0]'$. As before LMS and RLS solutions follow in a straightforward manner from the batch solution.

4. MLSE with adaptive channel estimation

The finite memory of the channel and the finite number of elements in the transmission alphabet lead us to frame the problem of input data estimation into a sequence detection framework. Thus given N samples of the vector output $\mathbf{x}(N-1)\cdots\mathbf{x}(0)$, where $N >> L$ we wish to find the sequence $s = s(N-1)\cdots s(0)$ which maximizes the conditional probability $P(\mathbf{x}(N-1)\cdots\mathbf{x}(0)|s(N-1)\cdots s(0))$. From (1), relying on the fact that $\mathbf{v}(n)$ is AWGN with variance σ^2 and is uncorelated with the input, leads us to write this conditional probability P_s as:

$$P(\mathbf{x}(N-1)\cdots\mathbf{x}(0)|s(N-1)\cdots s(0)) = \frac{1}{(\sqrt{2\pi}\sigma)^{MN}}\exp\left(\frac{-\sum_{k=0}^{N-1}|\mathbf{x}(k)-\sum_{l=0}^{L}\mathbf{h}_k(l)s(k-l)|^2}{2\sigma^2}\right). \qquad (12)$$

Note that (12) explicitly employs the channels estimate at a time k which we denote as $\mathbf{h}_k(l)$. Straightforward computations yield that maximizing P_s in (12) is equivalent to finding the sequence $s(N-1)\cdots s(0)$ that satisfies arg min P_s satisfies the following:

$$\arg\min \sum_{k=0}^{N-1} |\mathbf{x}(k) - \sum_{l=0}^{L}\mathbf{h}_k(l)s(k-l)|^2. \qquad (13)$$

To develop a procedure based on the Viterbi algorithm we note that arg min P_s satisfies:

$$\begin{aligned} &\arg\min P(\mathbf{x}(N-1)\cdots\mathbf{x}(0)|s(N-1)\cdots s(0)) = \\ &\arg\min P(\mathbf{x}(N-1)|s(N-1)\cdots s(N-L-1)) = \\ &\arg\min P(\mathbf{x}(N-2)\cdots\mathbf{x}(0)|s(N-2)\cdots s(0)) \end{aligned} \qquad (14)$$

Thus, each time k we have M^{L+1} states corresponding to each possible $s(k)\cdots s(k-L-1)$. Each state has M possible transitions. The branch metric computed is $|\mathbf{x}(k) - \sum_{l=0}^{L}\mathbf{h}_k(l)s(k-l)|^2$. Figure **??** shows an example state transition diagram for the demodulation of BPSK symbols transmitted through a vector channel of order 1.

The complexity of the Viterbi algorithm grows linearly with the number of data received, exponentially with the order of the channel, and logarithmically with the size of the alphabet. Ideally we would like to collect all the data and then make a decision on the best sequence. Unfortunately because of latency constraints and memory limitations we force a decision after a certain number of symbols have been received. It has been reported the error

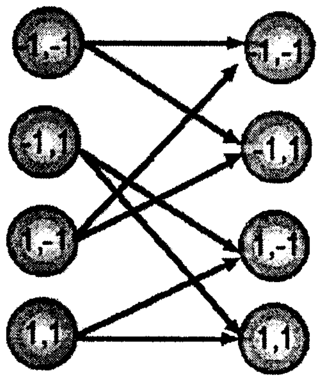

Figure 2. State transition diagram for demodulation of BPSK symbols transmitted through a vector channel of order 1

is still negligable if we force a decision after more than than $5L$ symbols have been received then error is negligable [9, Ch. 10]. As shown in [1] for some mobile communcations channels we can decrease complexity and still obtain good results by truncating the impulse response which effectively reducing the number of states necessary for decoding.

5. Simulations

Experiment 1: Here we demonstrate the performance of the RLS adaptive channel identification algorithm based on the XLTK algorithm for an *empirically measured digital radio channel* (see [2] for details). We have used an i.i.d. input sequence drawn from a 16-QAM constellation and have used an oversample ratio of 2 (i.e., $M = 2$). The noise is drawn from a white Gaussian distribution. Figure 3 shows the evolution of the root mean square channel error (RMSCE) as data are received at different signal-to-noise ratios (SNR). The RMSCE is defined as the average of $\| \mathbf{h}_k - \hat{\mathbf{h}}_k \| / \| \mathbf{h}_k \|$ over 100 Monte-Carlo simulations. From the Figure the fast convergence characteristic of the RLS algorithm is evident. In Figure 4 we show the evolution of the mean-squared prediction error $|\epsilon_k|^2$ over time, averaged over 100 Monte-Carlo trials. The prediction error indicates how well the algorithm is adapting to the true channels. These performance curves show similar results as the RMSCE. In Figure 5 we show the RMSCE performance of the Giannakis/Halford adaptive channel identification algorithm averaged over 100 monte-carlo simulations for the same channels as before. Here only the 0 dB and 40 dB noise levels are shown for clarity because the other curves were approximately the same as that of the $40dB$ curve. In Figure 6 we compare the RMSCE of the final channel estimate for each of the two RLS algorithms. As we would expect the performance of the deterministic XLTK approach steadily increases as the SNR increases, while in the stochastic Giannakis/Halford the performance levels off.

Experiment 2: It is of interest to examine the probability of error achieved using various channel estimation techniques. In Figure 7 we show the estimated bit error rate using $5 \cdot 10^5$ bits with forced decisions every 500 bits. We used BPSK input data with a channel of memory 2. First, a random channel estimate is employed into the Viterbi ISI decoder with that predictable result that decoding is not much better than guessing the input. Second, the batch XLTK algorithm was applied to every 500 symbols to provide the Viterbi ISI decoder with a channel estimate every 500 symbols. It seems that the channel estimates for below $10dB$ are worse than random estimates. Third, we estimate the zero-forcing equalizer from the batch identified channel then use it to equalize the data and make a decision. Fourth, we employed the LMS version of the XLTK algorithm, providing the Viterbi ISI decoder with a channel update at every symbol. The LMS algorithm was restarted every 500 symbols. Surprisingly despite the slow convergence of the channel estimates, the LMS approach yields good performance even at noise levels below 10 dB. Finally we showed the estimate bit error rate in the case where the channel is known.

References

[1] B. Desplanches, S. Buljore, and J. Diouris, "Complexity reduction for a multisensor Viterbi equaliser," *Electronics Letters*, vol. 32, no. 2, January 18, 1996.

[2] T. J. Endres, S. D. Halford, C. R. Johnson, Jr., and G. B. Giannakis, "Blind Adaptive Channel Equalization using Fractionally-Spaced Receivers: A Comparison Study," *Proc. of 30th Conf. on Information Sciences & Systems*, Princeton Univ., NJ, March 1996 (to appear).

[3] M. Feder and J. Catipovic, "Algorithms for joint channel estimation and data recovery–application to equalization in underwater communications," *IEEE Journal of Oceanic Engineering*, vol. 16, pp. 42-55, January 1991.

[4] G. D. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Inform. Theory*, vol. IT-18, pp. 363-378, May 1972.

[5] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based Adaptive Algorithms for the Blind Equalization of Multichannel FIR Filters," *Proc. of EUSIPCO*, 1994.

[6] G. B. Giannakis, and S. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: direct and adaptive solutions", *IEEE Trans. on Signal Proc.*, 1996 (accepted).

[7] S. Haykin, *Adaptive Filter Theory*, 2nd Edition, Prentice Hall, Englewood Cliffs, NJ, 1991.

[8] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, " Subspace Methods for the Blind Identification of Multichannel FIR Filters," *IEEE Trans. on Signal Processing*, vol. 43, pp. 516-525, February 1995.

[9] J. Proakis, *Digital Communications*, 2nd edition, New York, NY: McGraw-Hill, 1989.

[10] W. Sheen and G. Stüber, "MLSE Equalization and Decoding for Multipath-Fading Channels," *IEEE Trans. on Communications*, vol. 39, no. 10, October 1991.

[11] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction,"*Proc. of the Intl. Conf. on ASSP*, vol. IV, pp. 585-588, Adelaide, Australia 1994.

[12] L. Tong, "Blind Sequence Estimation," *IEEE Trans. on Communications*, vol. 43, no. 12, December 1995.

[13] L. Tong, G. Xu and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach", *IEEE Trans. on Information Theory*, vol. 40, no. 2, pp. 340-349, March 1994.

[14] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind Channel Identification Based on Second-Order Statistics: A Frequency-Domain Approach," *IEEE Trans. on Information Theory*, vol. 41, pp. 329 - 334, January 1995.

[15] G. Xu, H. Liu, L. Tong, and T. Kailath, "A Least-Squares Approach to Blind Channel Identification", *IEEE Trans. on Signal Processing*, vol. 43, pp. 2983-2993, December 1995.

[16] Y. Wan, Q. Liu, and A. Sendyk, "A fractionally-spaced maximum-likelihood sequence estimation receiver in a multipath fading environment," *Proc. of the Intl. Conf. on ASSP*, vol. IV, pp. 689-692, San Francisco, USA 1992.

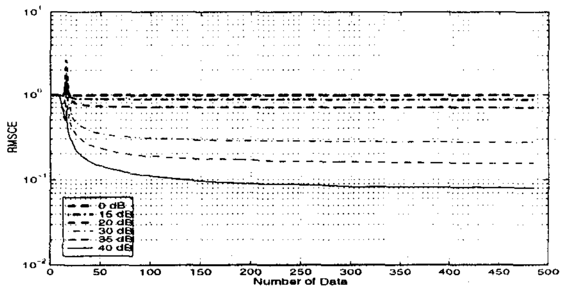

Figure 3. RMSCE performance of the RLS version of the XLTK algorithm

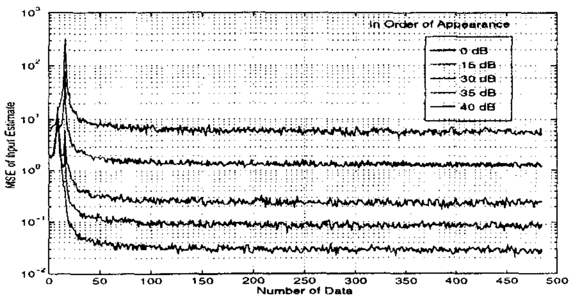

Figure 4. Prediciton error performance of the RLS version of the XLTK algorithm

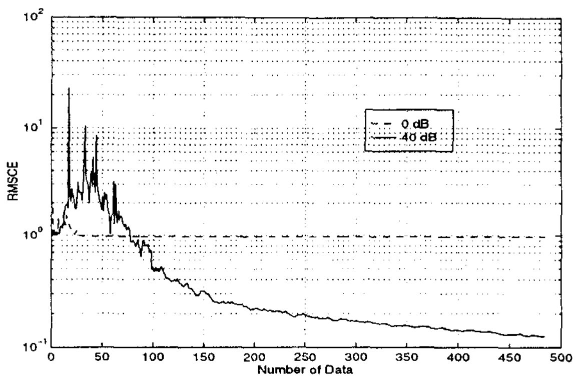

Figure 5. RMSCE Performance of the RLS version of the Giannakis/Halford algorithm

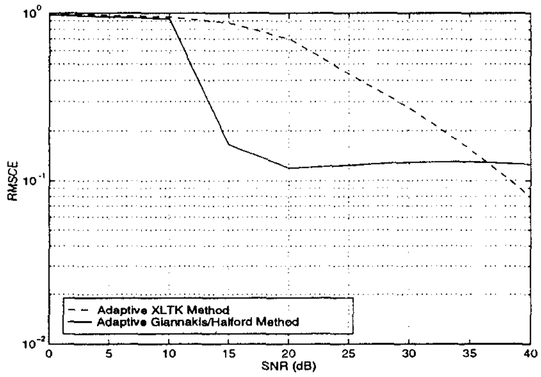

Figure 6. RMSCE Performance of comparison of both adaptive algorithms at different SNRs

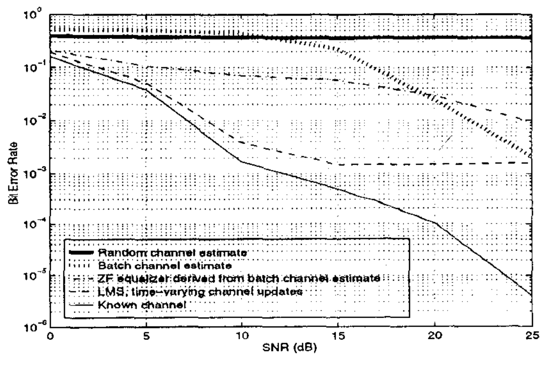

Figure 7. Probability of errror comparison