MAP Sequence Estimation for Fading ISI Channels

H. Chen, R. Perry and K. Buckley[1]
Electrical & Computer Engineering Department
Villanova University
Villanova, PA 19085
perry@ece.vill.edu, (610)-519-4969

Abstract — **In this paper we first describe a Maximum A Posterior (MAP) based sequence estimation approach for unknown, fast fading, frequency selective digital communications channels. The approach incorporates prior probabilistic knowledge of the channel via a stochastic channel model. We then assume a first order Gauss-Markov channel model to derive a specific MAP estimator, and we describe a Per Survivor Processing (PSP) algorithm as a realistic approximation of the optimum estimator. Monte Carlo simulations results are then presented: first to illustrate the performance realized when incorporating prior channel knowledge and its dependence on channel fading rate; and second to study the sensitivity of this MAP estimator to inaccuracies in the assumed values of parameters of the Gauss-Markov model.**

I. Introduction

MAP sequence estimation is widely used in digital communication systems to estimate the transmitted data sequence observed over noisy, multipath channels. This objective becomes particularly challenging when the channel is unknown and fast-fading. If the channel is slow time-varying, effective algorithms [1, 2, 3, 4] have been developed to estimate the transmitted data sequence and/or identify the channel blindly. For fast time-varying unknown channels, adaptive MLSE approaches have been developed. These algorithms usually estimate the transmitted data sequence based on the current estimate of the channel coefficients, while the channel coefficients are estimated using an estimate of the transmitted data sequence. One method of channel estimation [5] is to make tentative decisions which are obtained by truncating the surviving path in the Viterbi algorithm [6] to a fixed delay. Another channel estimator is realized by an adaptive decision-feedback equalizer (DFE) embedded in the MLSE structure [7]. Due to the decision delay in the former method and the error propagation caused by the DFE in the latter, these two methods have difficulty in tracking fast time-varying intersymbol interference (ISI) channels.

In [8] we have developed a maximum a posterior (MAP) sequence estimator for unknown, fast fading ISI channels for which the channel is modeled as FIR with a first-order Gauss-Markov coefficient vector. We derived a closed-form MAP cost function by marginalizing over the channel coefficients. For a small number of symbols n this cost can be exhaustively searched to obtain the MAP solution. For large n we proposed a Per Survivor Processing (PSP) algorithm. PSP has been successfully applied for suboptimal sequence estimation and channel identification. For example, Raheli et al. proposed PSP-based MLSE [9] for fast time-varying channels. A PSP based algorithm was proposed in [10], in which a Viterbi algorithm is used to identify survivor sequences, and a data-assisted channel estimator is employed for each survivor. For each survivor Viterbi path, the corresponding sequence was used for LMS or RLS based channel estimation. Although this adaptive MLSE approach can result in good tracking performance for fast time-varying ISI channels, it is a sub-optimal sequence estimator approach. The PSP algorithm we proposed, being based on MAP estimator which marginalizes over the channel, avoids the limitations associated with sub-optimal LMS or RLS channel estimation. In [8] we illustrated that it provides close to optimal results for a practical level of computation.

In this paper we employ this estimator to study the advantage realized by incorporating knowledge of the channel into a channel model and marginalizing over the channel coefficients. We will explore the degree of this advantage as a function of channel fading rate. We also study the sensitivity of this MAP estimator to inaccuracies in the assumed values of parameters of the Gauss-Markov model.

II. Communication System Model

In the discrete-time FIR model of a fast-fading, frequency-selective noisy communications channel, for the received data sequence up to time n, the complex received data r_k at time k is given by

$$r_k = \mathbf{a}_k^T \mathbf{h}_k + n_k, \qquad k = 1, \ldots, n \ , \tag{1}$$

where $\mathbf{a}_k^T$ is a complex row vector containing transmitted data $\{a_{k-i+1}, i = 1, \ldots, M\}$, M is the FIR channel length, $\mathbf{h}_k$ is a complex column vector containing the channel impulse response coefficients at time k, and n_k is the white Gaussian complex noise at time k with variance σ^2. Let $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$ represent the matrix of channel coefficient vectors over time, arranged by columns. Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]^T$ represent the matrix of transmitted data arranged by rows. Also let $\mathbf{r} = [r_1, \ldots, r_n]^T$ represent the column vector of received data, and $\mathbf{n} = [n_1, \ldots, n_n]^T$ represent the column vector of noise over time. With this notation, the probability density function of the received data, given $\mathbf{H}$ and $\mathbf{A}$, is

$$f(\mathbf{r}|\mathbf{H}, \mathbf{A}) = \frac{1}{(\pi\sigma^2)^n} \prod_{k=1}^{n} e^{-\frac{|r_k - \mathbf{a}_k^T \mathbf{h}_k|^2}{\sigma^2}} . \tag{2}$$

The basic equations above describe the communication system under consideration, based only on the fact that the additive noise is white and Gaussian, without making any additional assumptions about the properties of the FIR channel itself.

To minimize BER, the MAP criterion should be used, i.e. we must find $\mathbf{A}$ to maximize $f(\mathbf{A}|\mathbf{r})$. Using Bayes rule,

$$f(\mathbf{A}|\mathbf{r}) = \frac{f(\mathbf{r}|\mathbf{A}) f(\mathbf{A})}{f(\mathbf{r})} \doteq f(\mathbf{r}|\mathbf{A}) f(\mathbf{A}) \ . \tag{3}$$

[1] This research was supported by ONR under Grant N00014-98-1-0892.

(The notation $\doteq$ means "equivalent within an additive and/or multiplicative constant" for equations in which such constants have no effect for optimization purposes.) In (3), the term $f(\mathbf{r})$ is ignored since it has no effect in the relative cost. MAP estimators minimize BER by maximizing $f(\mathbf{r}|\mathbf{A})f(\mathbf{A})$. If $f(\mathbf{A})$ is unknown or assumed to be uniform, then the ML estimator which maximizes the likelihood function $f(\mathbf{r}|\mathbf{A})$ can be used instead. Referring to (2), if the channel is known, the Viterbi algorithm can be used directly to estimate the data sequence. If the channel is unknown, the dependence of $f(\mathbf{r}|\mathbf{A})$ on the random channel coefficients is

$$f(\mathbf{r}|\mathbf{A}) = E[f(\mathbf{r}|\mathbf{H},\mathbf{A})] = \int_{\mathbf{H}} f(\mathbf{r}|\mathbf{H},\mathbf{A}) f(\mathbf{H}) d\mathbf{H} \ , \qquad (4)$$

where $f(\mathbf{r}|\mathbf{H},\mathbf{A})$ is given by (2).

Given prior information on the distribution of the channel coefficient vector sequence $\mathbf{h}_k$, in the form of a joint probability density function $f(\mathbf{H})$, (3) and (4) may be evaluated to derive the MAP sequence estimator.

III. First Order Gauss Markov Channel Model

Our objective is to incorporate prior knowledge of the fast-fading, FIR channel via a flexible, realistic stochastic linear model that can account for correlation across both channel taps and time. Here we choose a first order Gauss Markov model because it fulfills this objective – it can be used to represent a broad class of fast-varying FIR systems, and the resulting channel taps represent Rayleigh fading.

Let α represent a complex first-order Markov memory factor such that $|\alpha| = e^{-\omega T}$ where T is the sampling period and $\frac{\omega}{\pi}$ is the Doppler spread [11]. Let the time-varying channel coefficients follow a Gauss-Markov distribution such that, at time k, $\mathbf{h}_k = \alpha \mathbf{h}_{k-1} + \mathbf{v}_k$, where $\mathbf{v}_k$ is complex, white, and Gaussian with mean $\mathbf{d}$ and covariance $E\{\mathbf{v}_k \ \mathbf{v}_k^H\} = \mathbf{R}_\mathbf{v} = \mathbf{C}$. In [11] a scalar version of this model is used to represent a frequency-nonselective fast-fading channel. In [12] a slightly different first order Gauss-Markov channel model of a frequency-selective fast-fading channel is used for joint channel/sequence estimation.

For this model, the channel covariance matrix sequence is, in steady state,

$$\mathbf{R}_\mathbf{h}(m) \ = \ E\{\mathbf{h}_k \ \mathbf{h}_{k+m}^H\} \ = \ \frac{\alpha^{|m|}}{1-|\alpha|^2} \ \mathbf{R}_\mathbf{v} \ . \qquad (5)$$

For diagonal $\mathbf{R}_\mathbf{v}$, the taps weights of the channel are uncorrelated, which for example corresponds to an uncorrelated multipath scattering assumption ([13], p. 842). As indicated above, $|\alpha|$ is selected in accordance with the expected Doppler spread. $\angle\alpha$ controls model Doppler spectrum characteristics.

Figure 1 illustrates several $M = 3$ coefficient Gauss-Markov FIR channels. For each, the coefficients are zero-mean (i.e. $\mathbf{d} = \mathbf{0}$) and uncorrelated across delay with variances $\sigma_h^2 = 0.01$. Over a $n = 100$ symbol duration, example trajectories are plotted for the three coefficients. Figures 1a, 1b, 1c, 1d correspond respectively to $\alpha = 0.95, 0.975, 0.99, 0.995$ and $\omega T = 0.0513, 0.0253, 0.0101, 0.005$. Reliable MAP sequence estimation can not be accomplished for a Gauss-Markov channel with $\alpha = 0.95$. The channel varies too quickly. As expected, as α increases towards one, MAP sequence estimation is more reliable. This will be illustrated in the simulation section.

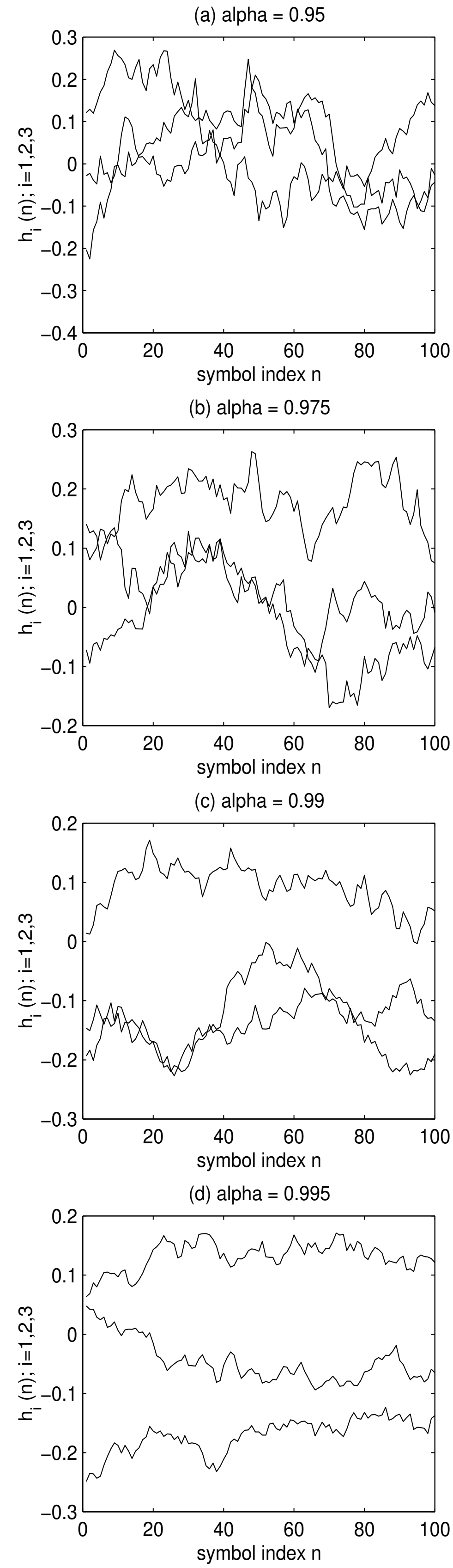

Figure 1: Examples Gauss-Markov channel coefficient trajectories, for channel memory length $M = 3$, for a duration of $n = 100$ symbols, and for $\alpha = 0.95, 0.975, 0.99, 0.995$.

IV. Sequence Estimation and Per Survivor Processing

In this Section we summarize the development of the MLSE algorithm for a fast fading Gauss-Markov channel model. Given prior information on the sequence distribution, the corresponding MAP estimator is a straightforward extension (e.g. [14]).

For the Gauss Markov model under consideration here, $f(\mathbf{h}_k|\mathbf{h}_{k-1})$ is Gaussian with mean $\mathbf{d}+\alpha\mathbf{h}_{k-1}$ and covariance $\mathbf{C}$. $f(\mathbf{H})$ can be written as

$$f(\mathbf{H}) = \prod_{k=1}^{n} f(\mathbf{h}_k|\mathbf{h}_{k-1}). \tag{6}$$

To simplify expressions, we assume below that $\mathbf{d}$ is zero. We also assume that the initial channel coefficient vector $\mathbf{h}_0$ is known (e.g. estimated from a previous data block). Substituting (2) and (6) into (4), we obtain

$$\begin{aligned} f(\mathbf{r}|\mathbf{A}) &= E[f(\mathbf{r}|\mathbf{H},\mathbf{A})] \\ &= \int_{\mathbf{H}} f(\mathbf{r}|\mathbf{H},\mathbf{A}) f(\mathbf{H}) d\mathbf{H} \\ &\doteq \int_{\mathbf{H}} \left(\prod_{k=1}^{n} e^{-\frac{|r_k - \mathbf{a}_k^T \mathbf{h}_k|^2}{\sigma^2}} \right) \prod_{k=1}^{n} f(\mathbf{h}_k|\mathbf{h}_{k-1}) d\mathbf{H}. \end{aligned} \tag{7}$$

We can marginalize over the channel coefficient distribution, integrating step by step as shown in [8]. After the integration over all of the channel coefficients, we have the likelihood function

$$f(\mathbf{r}|\mathbf{A}) \doteq \prod_{k=1}^{n} e^{\mathbf{q}_k{}^H \mathbf{G}_k \mathbf{q}_k} |\mathbf{G}_k| \ , \tag{8}$$

where the $\mathbf{q}_k$ and $\mathbf{G}_k$; $k = 1, 2, \cdots, n$ are given as

$$\begin{aligned} \mathbf{G}_1 &= \left[\frac{\mathbf{a}_1^* \mathbf{a}_1^T}{\sigma^2} + (1+|\alpha|^2)\mathbf{C}^{-1} \right]^{-1} \\ \mathbf{q}_1 &= \frac{r_1 \mathbf{a}_1^*}{\sigma^2} + \alpha \mathbf{C}^{-1} \mathbf{h}_0 \end{aligned} \tag{9}$$

$$\begin{aligned} \mathbf{G}_k &= \left[\frac{\mathbf{a}_k^* \mathbf{a}_k^T}{\sigma^2} + \mathbf{C}^{-1} + |\alpha|^2 \left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{G}_{k-1}\mathbf{C}^{-1} \right) \right]^{-1} \\ \mathbf{q}_k &= \frac{r_k \mathbf{a}_k^*}{\sigma^2} + \alpha \mathbf{C}^{-1} \mathbf{G}_{k-1} \mathbf{q}_{k-1} \qquad k = 2, ..., n-1 \ , \end{aligned} \tag{10}$$

and

$$\begin{aligned} \mathbf{G}_n &= \left[\frac{\mathbf{a}_n^* \mathbf{a}_n^T}{\sigma^2} + \mathbf{C}^{-1} - |\alpha|^2 \mathbf{C}^{-1}\mathbf{G}_{n-1}\mathbf{C}^{-1} \right]^{-1} \\ \mathbf{q}_n &= \frac{r_n \mathbf{a}_n^*}{\sigma^2} + \alpha \mathbf{C}^{-1} \mathbf{G}_{n-1} \mathbf{q}_{n-1} \ . \end{aligned} \tag{11}$$

If we take the negative natural logarithm of (8), we will obtain the time-recursive form of the cost function:

$$-\log f(\mathbf{r}|\mathbf{A}) \doteq -\sum_{k=1}^{n} \left[\mathbf{q}_k{}^H \mathbf{G}_k \mathbf{q}_k + \log |\mathbf{G}_k| \right]. \tag{12}$$

From the expressions for $\mathbf{G}_k$ and $\mathbf{q}_k$, we see that the transition costs of the Viterbi algorithm depend on all of the previous states. So the standard Viterbi algorithm can not be optimally applied here. Optimum results can be obtained by exhaustive search. Alternatively, computationally feasible suboptimum algorithms can be considered. Here we describe a suboptimum algorithm which employs the generalized Viterbi algorithm (GVA) [10] to prune possible data sequences. Considering a trellis formulation, at each state of each stage we keep $L \geq 1$ survivor paths (i.e. surviving hypothesized sequences). For each survivor, the $\mathbf{q}_k$ and $\mathbf{G}_k$ required for the incremental cost are computed. (Since for each survivor path the $\mathbf{G}_k$ and $\mathbf{q}_k$ are computed from the sequence associated with that path, we can interpret this as data-aided estimation of the incremental cost corresponding to the particular path, so this algorithm is effectively a form of PSP [9].) At time n the sequence estimate corresponds to the best surviving path. As L increases, the PSP algorithm results will approach the optimum. Simulations in [8] show that this approach results in effective algorithms for ML (or MAP) sequence estimation over unknown, fast time-varying channels for reasonably small values of L. We have found that $L \leq 3$ is typically adequate.

V. Simulation Results

In [8] simulation results are presented that illustrate that the MAP sequence estimator described above, and its suboptimal PSP implementation, performs better than the RLS and LMS based PSP algorithms proposed in [10]. It is also shown that near optimal performance is realized using the list Viterbi PSP algorithm mentioned above, with as low as $L = 3$ paths kept per trellis node. Here we concentrate on two issues: the accuracy to which symbols can be estimate as influenced by the channel fading rate; and the sensitivity of the Gauss-Markov channel model based MAP estimator on inaccuracies in assumed model parameters.

Here we model the channel as described in Section III and illustrated in Figure 1. A $M = 3$ tap channel is considered, with zero-mean coefficients (i.e. $\mathbf{d} = \mathbf{0}$) which are uncorrelated across delay and have variances $\sigma_h^2 = 0.01$. For each trial, a $n = 100$ symbol transmission is simulated, with $\mathbf{h}_0$ generated using the assumed $\mathbf{h}$ statistics noted above. Figure 2 shows achieved bit-error-rates (BER's) for three values of α: 0.98, 0.99, and 0.995. Performance is probably unexceptable for $\alpha < 0.97$, and as expected performance improves for increasing α (i.e. reduced fading rate). Eventually, as SNR increases, performance improvement becomes negligible, corresponding to a channel-uncertainty limited mode of operation. Even for very fast fading channels, very good performance can be achieved if the channel is known. However, when the channel is unknown, temporal correlation is an important performance factor. The larger the temporal correlation (i.e. the closer $|\alpha|$ is to one for a Gauss-Markov channel), the better the sequence of symbols can be estimated.

Next we consider the sensitivity of the Gauss-Markov channel model based MAP sequence estimator to assumed model parameters. For Figures 3-5, $n = 10$ symbols were used, and the optimum Gauss-Markov model results were obtained by exhaustive search. This allows us to accurately measure the model parameter sensitivity without affects which may be caused by the sub-optimum PSP approach.

Figures 3 and 4 illustrate the sensitivity to the assumed value of α for two different values of true α. For model $\alpha = 0$, the Gauss-Markov channel model degenerates to a temporally-independent Gaussian model (see [8]). Although there is significant performance variation over the range $0 \leq \alpha \leq 1$, over the range of useful α (say $0.975 \leq |\alpha| < 1$), performance does not vary significantly.

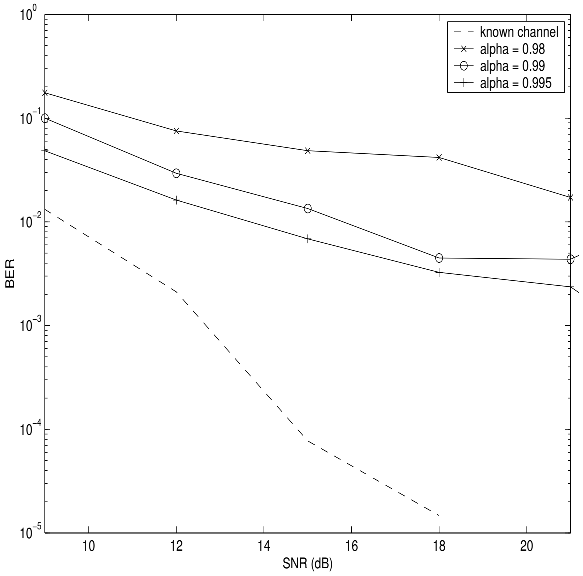

Figure 2: Bit-Error-Rate (BER) vs. SNR for several values of the Gauss-Markov memory factor α.

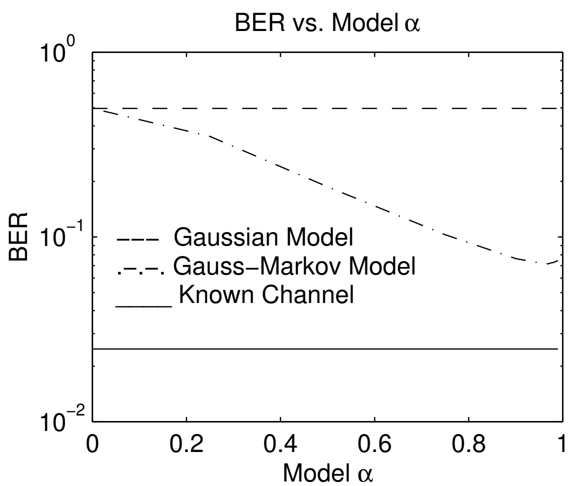

Figure 3: BER vs. model α. True $\alpha = 0.975$

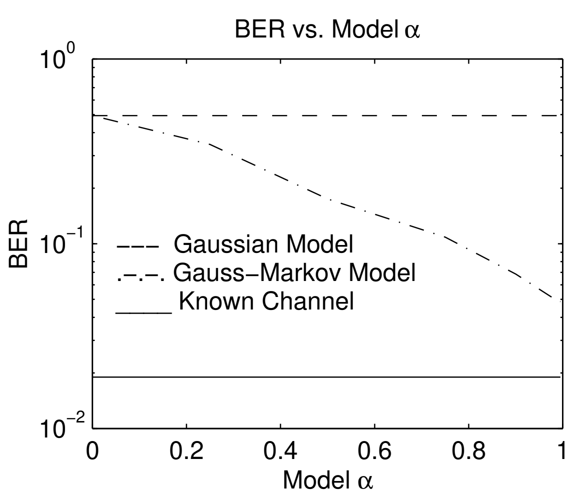

Figure 4: BER vs. model α. True $\alpha = 0.99$

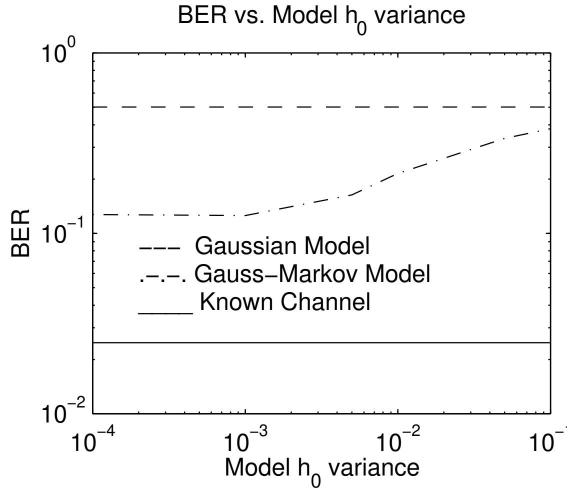

Figure 5: BER vs. model h_0 variance. $\alpha = 0.975$

For Figure 5, the true value of α was used in the Gauss-Markov model, but the assumed value of $\mathbf{h}_0$ was randomly perturbed by adding a zero-mean Gaussian component. For variance of the perturbation less than .001, performance is unaffected. As the variance increases, performance degrades and approaches the Gaussian model. This illustrates that the algorithm is not overly sensitive to errors in the assumed value of $\mathbf{h}_0$. Thus, using an $\mathbf{h}_0$ estimated from a previous data block is possible.

VI. Summary

In this paper we describe a MAP sequence estimator for frequency-selective fast-fading channels. The estimator incorporates prior channel knowledge using a Gauss-Markov model of the FIR channel by marginalizing over the resulting channel coefficient joint probability density function. A computationally efficient PSP algorithm is also described, and its performance illustrated. With simulations, we consider the accuracy to which symbols can be estimate as influenced by the channel fading rate, and we study the sensitivity of the Gauss-Markov channel model based MAP estimator on inaccuracies in assumed model parameters.

References

[1] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," *Optical Engineering*, vol. 31, pp. 1224–1228, June 1992.

[2] K. Chang and C. N. Georghiades, "Iterative joint sequence and channel estimation for fast time-varying intersymbol interference channels," *ICC'95*, pp. 357–361, June 1995.

[3] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. on Communications*, vol. 42, pp. 1000–1011, April 1994.

[4] E. Zervas, J. G. Proakis, and V. Eyuboglu, "A quantized channel approach to blind equalization," *Proc. ICC'91, Chicago, Il*, June 1991.

[5] F. R. Magee, Jr. and J. G. Proakis, "Adaptive maximum-likelihood sequence estimation for digital signaling in the presence of intersymbol interference," *IEEE Trans. on Information Theory*, pp. 120–124, Jan 1973.

[6] G. D. Forney, "The Viterbi algorithm," *Proc. of the IEEE*, vol. 61, pp. 268–278, Mar 1973.

[7] S. U. H. Qureshi, "An adaptive decision-feedback receiver using maximum-likelihood sequence estimation," *Proc. 1973 IEEE Int. Conf. Commun.*, pp. 14.10–14.16, 1973.

[8] H. Chen, K. Buckley, and R. Perry, "Time-recursive maximum likelihood based sequence estimation for unknown ISI channels," *Proc. 33-rd Asilomar Conference on Circuits, Systems and Computers*, 2000.

[9] R. Raheli, A. Polydoros, and C. K. Tzou, "The principle of per-survivor processing: a general approach to approximate and adaptive MLSE," *IEEE Trans. on Communications*, vol. 43, pp. 354–364, Feb/Mar/Apr 1995.

[10] K. M. Hiroshi Kubo and T. Fujino, "An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels," *IEEE Transactions on Communications*, pp. 1872 – 1880, February 1994.

[11] M. Stojanovic and Z. Zvonar, "Performance of multiuser diversity reception in Rayleigh fading CDMA channels," *IEEE Trans. on Communications*, vol. 47, pp. 356–359, Mar 1999.

[12] E. Baccarelli and R. Cusani, "Combined channel estimation and data detection using soft statistics for frequency-selective fast-fading digital links," *IEEE Trans. on Communications*, vol. 46, pp. 424–427, Apr 1998.

[13] J. G. Proakis, *Digital Communications, 4-rd ed.* McGraw-Hill, 2001.

[14] A. Logothetis and V. Krishnamurthy, "Expectation maximization algorithms for MAP estimation of jump Markov linear systems," *IEEE Transactions on Signal Processing*, pp. 2139–2156, August 1999.