Transactions Papers

Adaptive Soft-Input Soft-Output Algorithms for Iterative Detection with Parametric Uncertainty

Achilleas Anastasopoulos, *Member, IEEE*, and Keith M. Chugg, *Member, IEEE*

***Abstract*—The soft-input soft-output (SISO) module is the basic building block for established iterative detection (ID) algorithms for a system consisting of a network of finite state machines. The problem of performing ID for systems having parametric uncertainty has received relatively little attention in the open literature. Previously proposed adaptive SISO (A-SISO) algorithms are either based on an oversimplified channel model, or have complexity that grows exponentially with the observation length N (or the smoothing lag D). In this paper, the exact expressions for the soft metrics in the presence of parametric uncertainty modeled as a Gauss–Markov process are derived in a novel way that enables the decoupling of complexity and observation length. Starting from these expressions, a family of suboptimal (practical) algorithms is motivated, based on forward/backward adaptive processing with linear complexity in N. Recently proposed A-SISO algorithms, as well as existing adaptive hard-decision algorithms are interpreted as special cases within this framework. Using a representative application—joint iterative equalization-decoding for trellis-based codes over frequency-selective channels—several design options are compared and the impact of parametric uncertainty on previously established results for ID with perfect channel state information is assessed.**

***Index Terms*—Channel estimation, frequency-selective fading, intersymbol interference, iterative decoding, Kalman filters, per-survivor processing, serially concatenated codes, soft statistics, symbol-by-symbol MAP detection, turbo codes.**

I. Introduction

RECENTLY, there has been great interest in iterative detection (ID) schemes for systems consisting of multiple finite state machines (FSMs), which can be loosely defined as the set of rules to exchange, combine, and marginalize some sort of soft information related to the FSM input/output symbols, with the purpose of providing reliable decisions about the input sequence. Applications that utilize this scheme include turbo decoding of parallel and serial concatenated convolutional codes (PCCCs and SCCCs) [1], [2], decoding of trellis coded modulation (TCM) in interleaved frequency-selective fading channels [3], [4], as well as various multidimensional detection problems [5], [6]. The core building block in these iterative schemes is the soft-input soft-output (SISO) module [7]; an algorithm—similar to the Viterbi Algorithm (VA) [8]—that accepts *a priori* information on the input and output symbols of an FSM and outputs the corresponding *a posteriori* information, with complexity growing linearly with the record length N.

In most practical situations where perfect channel state information (CSI) is not available at the receiver (e.g., PCCCs and SCCCs with carrier phase tracking or TCM in fast frequency-selective fading channels), an adaptive[1] ID (AID) scheme is required to deal with the unknown, and possibly time-varying parameters. In this paper, a subclass of adaptive iterative receivers is investigated, in which the parameter estimates are not exchanged as part of the iterative procedure, rather, they are generated and are confined inside the adaptive SISO (A-SISO) modules, which are the natural extension of the SISO modules for the case of parametric uncertainty. Nevertheless, the exchange of soft information on the FSM symbols provides a implicit mechanism for the reestimation of the unknown parameters as well.

In the simplest case of the unknown parameter being modeled as a Markov chain with finite number of states, the optimal A-SISO is a modified SISO that runs on the augmented FSM [9]. Of more interest is the case of the parameter being continuous in nature (e.g., phase offset or channel taps). Early attempts to solve this more general problem were based on the Baum–Welch method (or equivalently the expectation maximization (EM) algorithm [10]). Since convergence to a locally optimal solution is possible, the optimality of the EM algorithm cannot always be guaranteed. In [11], a Gauss–Markov (GM) model is assumed for the unknown parameter and the optimal scheme is derived. Starting from a different viewpoint, structurally similar algorithms are derived in [12] and [13] for GM and deterministic parameter models, respectively. Finally, a suboptimal A-SISO with a single-parameter estimator was developed in [14]. The inherent limitation of all the above approaches is that they all operate in a fixed lag (FL) mode; thus, two major conflicting goals in designing a practical algorithm are coupled through a single parameter, the smoothing depth D. Indeed, in an FL algorithm, a large decision delay (smoothing depth) D is

Paper approved by T. Aulin, the Editor for Coding and Communications Theory of the IEEE Communications Society. Manuscript received December 23, 1998; revised March 31, 1999, March 28, 2000, and April 20, 2000. This work was supported in part by the National Science Foundation under Grant NCR-CCR-9726391. This paper was presented in part at the International Conference on Communications, Vancouver, BC, Canada, June 1999.

A. Anastasopoulos is with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: anastas@umich.edu).

K. M. Chugg is with the Electrical Engineering–Systems Department, University of Southern California, Los Angeles, CA 90089-2565 USA.

Publisher Item Identifier S 0090-6778(00)08773-0.

[1]The term "adaptive" means that the algorithm attempts to track the unknown time-varying parameters.

required to deliver reliable soft information. On the other hand, the same parameter determines the amount of pruning of the sequence tree and needs to be kept as small as possible, especially since it results in exponential complexity growth. Additional simplifications may then be summoned upon to decouple D and complexity (e.g., thresholding is used in [11], while reduced state sequence estimation and suboptimal filtering is used in [12] to further reduce the processing burden).

In this paper, meaningful[2] soft metrics for the GM parameter model are defined and exact expressions are derived in a novel way, that motivates a family of suboptimal practical algorithms, the unique characteristic of which is the decoupling of complexity and smoothing depth, leading directly to fixed interval (FI) schemes that have linear complexity with the record length N, as is the case for SISOs when no parametric uncertainty is present. In addition, all existing A-SISO algorithms for continuous valued parameter models can be viewed as *forward only* special cases within this framework. In the application examined in this paper—TCM in interleaved frequency-selective fading channels—the effectiveness of the various A-SISO options is assessed via extensive simulations, and the impact of parametric uncertainty on previously established conclusions (e.g., in [15]) for iterative detection in systems consisting of concatenated FSMs, is assessed.

In the next section, the system and channel models are presented. Soft metrics are defined in Section III and the exact A-SISOs are developed, while practical algorithms are derived in Section IV. An extensive discussion on the available options for receiver design and numerical results are offered in Section V, and the conclusions are summarized in Section VI.

II. System and Channel Model

Consider a typical time-division multiple-access (TDMA) cellular transmission system consisting of a memoryless source that feeds a convolutional code. The trellis-coded symbols are interleaved, mapped into a constellation, and pulse-shaped before transmission. The low-pass equivalent transmitted signal is of the form

$$s(t) = \sqrt{E_s} \sum_{k=-\infty}^{\infty} d_k p(t - kT) \tag{1}$$

where d_k is the coded symbol (normalized to unit energy), E_s is the symbol energy, $p(t)$ is the shaping pulse (normalized to unit energy), and T is the symbol duration. This signal is distorted by a time-varying frequency-selective fading channel with impulse response of the form

$$c(t, \tau) = \sum_{n=0}^{L_c} c_n(t) \delta(\tau - nT_r) \tag{2}$$

where $T_r = T/N_r$ with N_r integer, and the dynamics of the vector random process $\underline{c}(t) = [c_0(t), \ldots, c_{L_c}(t)]^T$ are assumed slow compared to the symbol duration T. The distorted signal is observed in additive white Gaussian noise (AWGN) with power spectral density level N_0.

$$z(t) = \sqrt{E_s} \sum_{k=-\infty}^{\infty} d_k \sum_{n=0}^{L_c} c_n(t) p(t - kT - nT_r) + n(t). \tag{3}$$

Several options are available at the receiver front-end (FE) for preprocessing the received signal: low-pass filtering or match filtering with $p^*(-t)$, followed by fractionally-spaced sampling every T_s (where $T_s = T/N_s$, with N_s integer), followed by noise whitening (if necessary), as is extensively discussed in [16]–[18].[3] Regardless of the specific FE structure, the FE output can be modeled as an equivalent symbol-spaced vector intersymbol interference (ISI) channel as follows:

$$\underline{z}_k = \sqrt{E_s} \sum_{n=0}^{L} d_{k-n} \underline{g}_k(n) + \underline{n}_k \tag{4a}$$

$$\begin{aligned} &= \sqrt{E_s}[d_k, \ldots, d_{k-L}] \diamond \underline{g}_k + \underline{n}_k \\ &= \underline{q}_k^T \diamond \underline{g}_k + \underline{n}_k = \underline{q}_k^T \diamond (V \underline{c}_k) + \underline{n}_k \end{aligned} \tag{4b}$$

where $\underline{z}_k$, $\underline{n}_k$, and $\underline{g}_k(n)$ are all N_s-dimensional vectors, and $\underline{n}_k$ is complex, circular AWGN with independently, identically distributed components and $E\{\|\underline{n}_k\|^2\} = N_0$. The $N_s(L_c + 1)$-dimensional vector $\underline{c}_k = [c_0(kN_sT_s), \ldots, c_0((kN_s + N_s - 1)T_s), \ldots, c_{L_c}((kN_s + N_s - 1)T_s)]^T$ contains all the information relevant to the channel process $\underline{c}(t)$, while all the details of the pulse shaping and the FE are included in the matrix V as shown in [18]. Finally, the $N_s(L+1)$-dimensional vector $\underline{g}_k = [\underline{g}_k^T(0), \ldots, \underline{g}_k^T(L)]^T$ is the equivalent channel (which includes the effect of pulse shaping, channel, and FE) at time k, and the shorthand *diamond* $(\diamond)$ notation is used in (4b) to denote the mixed inner product implied by (4a).

Equation (4) can be used under either a stochastic or an a-stochastic (i.e., deterministic) assumption for $\underline{c}(t)$. An often used model for $\underline{c}_k$, and thus $\underline{g}_k = V\underline{c}_k$ is that of a Gaussian autoregressive-moving average (ARMA) process, generated by the plant equations [19]

$$\underline{\phi}_{k+1} = \Phi \underline{\phi}_k + \underline{\nu}_k \quad \underline{c}_k = C \underline{\phi}_k \tag{5}$$

where $\underline{\nu}_k$ is a white noise sequence and the dimensionality of the state $\underline{\phi}_k$ is in general higher than $\underline{c}_k$. It has been shown (e.g., [20]) that such model can adequately approximate realistic fading channels with nonrational spectrum (e.g., [21]).

One of the conclusions in [16]–[18] is that, while the specific FE processing (i.e., the matrix V) is important for making quantitative claims and for claiming optimality, qualitative conclusions about different postprocessors remain the same for different FEs (e.g., per-survivor Processing (PSP) [22] outperforms the conventional, adaptive maximum-likelihood sequence detector (CA-MLSD) [23]). In this paper, we focus on the introduction of novel postprocessing approaches that

[2]The term "meaningful" is used to distinguish algorithms that are designed to compute quantities that are consistently defined and may be considered reasonable soft metrics, as opposed to algorithms constructed in an *ad hoc* manner, where the nature of the soft output produced is not understood.

[3]A fractionally-spaced $(T_s = T_r)$ matched filtering with $p^*(-t)$, followed by whitening, was shown to be optimal, i.e., it provides sufficient statistics when $\underline{c}(t)$ is slowly varying compared to T. Similarly, low-pass filtering and fractionally-spaced sampling is optimal as long as the filter and T_s are selected such that the signal part of $z(t)$ is not distorted.

are valid for any FE processing (i.e., for the general model in (4)–(5). However, in order to improve the readability of the development and to reduce the simulation effort, we focus on the following special case. In particular, regarding the channel model in (5), a first-order GM model is adopted for $\underline{c}_k$, and thus $\underline{g}_k$ (i.e., $\underline{\phi}_k = \underline{c}_k$ is assumed). Furthermore, we chose to illustrate the concepts using a simplified symbol-spaced scalar ISI model (i.e., $N_s = 1$). Once the concepts introduced are understood, it is straightforward to modify the algorithms to account for the more general case of $N_s > 1$ and arbitrary plant model. Finally, to simplify the notation, vector quantities will not be underlined.

In the following, a model for a generic FSM is presented. The output y_k of a generic FSM can be defined as a function of its input x_k and state s_k—together constituting the transition $t_k = (s_k, x_k)$—through the equations

$$y_k = \mathrm{out}(x_k, s_k) \quad s_{k+1} = \mathrm{ns}(x_k, s_k) \tag{6}$$

where each *integer* quantity u_k (i.e., x_k, y_k, s_k, or t_k) is assumed to take values in the set $A_u = \{0, 1, \ldots, N_u - 1\}$. The output y_k of the FSM is either used as an input to another FSM, or observed indirectly, through a function, which also involves the unknown parameter g_k. Under the simplifying assumption mentioned earlier, the $(L+1)$-dimensional vector process $\{g_k\}$ evolves in time according to the equations[4]

$$\begin{aligned} g_k &= G g_{k-1} + w_k \quad \text{(forward)} \\ g_k &= G^b g_{k+1} + v_k \quad \text{(backward)} \end{aligned} \tag{7}$$

where w_k, v_k are zero-mean Gaussian vectors with covariance $K_w(m) = Q\delta_K(m)$ and $K_v(m) = Q^b\delta_K(m)$, respectively.[5] Equation (4) can now be written as

$$z_k = f(y_k)^T g_k + n_k = q_k^T g_k + n_k = m_k + n_k \tag{8}$$

where $q_k = f(y_k)$ is a complex vector depending on the modulation format (e.g., in the TCM system, $f(\cdot)$ maps the output of the inner FSM, which is the entire transition $y_k = t_k$, to the $L+1$ constellation points, as shown in (4)).

III. Exact Evaluation of the Soft Metrics

The objective of a SISO algorithm is to provide soft information about the input and output symbols of the FSM based on the observation record. This reliability information can either be in the form of an *a posteriori* probability or any other related quantity. It would be advantageous at this point to generalize the notion of the state s_k and transition t_k to longer sequence portions (e.g., a super-state and super-transition can be defined as $s_k^s = (t_{k-d}, \ldots, t_{k-1}, s_k)$ and $t_k^s = (t_{k-d}, \ldots, t_k)$ for arbitrary d). This foreshadows the result that the optimal algorithms do not "fold" [24] onto a trellis as in the case of known channel and that the size of the trellis eventually used is a design parameter. For a generic quantity u_k (i.e., $x_k, y_k, s_k, t_k, s_k^s, t_k^s$, etc.), we define the *a posteriori* probability (APP) and minimum sequence metric (MSM) soft outputs as follows:

$$\begin{aligned} \mathrm{APP}_p(u_k) &\stackrel{\text{def}}{=} P(u_k \mid z_0^n) = c \sum_{x_0^n : u_k} P(z_0^n, x_0^n) \\ &= c \sum_{x_0^n : u_k} E_\Theta \{P(z_0^n, x_0^n \mid \Theta)\} \end{aligned} \tag{9a}$$

$$\begin{aligned} \mathrm{MSM}_p(u_k) &\stackrel{\text{def}}{=} -\log\left[\max_{x_0^n : u_k} P(x_0^n \mid z_0^n)\right] \\ &= c' - \log\left[\max_{x_0^n : u_k} P(z_0^n, x_0^n)\right] \\ &= c' - \log\left[\max_{x_0^n : u_k} E_\Theta \{P(z_0^n, x_0^n \mid \Theta)\}\right] \end{aligned} \tag{9b}$$

where $x_0^n : u_k$ denotes all input sequences consistent with u_k, and c, and c' are normalizing constants. These soft outputs are the direct generalizations of well-known soft outputs for perfect CSI [7] to the case of an unknown parameter Θ. When the SISO module is part of an iterative receiver, the soft output is usually normalized to the *a priori* information resulting in the so-called extrinsic information (e.g., $\mathrm{APP}_p(u_k)/P(u_k)$, or $\mathrm{MSM}_p(u_k) - (-\log P(u_k))$ is used in place of $\mathrm{APP}_p(\cdot)$ or $\mathrm{MSM}_p(\cdot)$, respectively). We observe that in both cases, the soft outputs can be derived from the quantity $E_\Theta\{P(z_0^n, x_0^n \mid \Theta)\}$ by either averaging or maximizing—for $\mathrm{APP}_p(\cdot)$ or $\mathrm{MSM}_p(\cdot)$, respectively—over the nuisance parameters $x_0^n : u_k$.

Equation (9) clearly suggests a way of manipulating $P(z_0^n, x_0^n \mid \Theta)$ to obtain the proposed soft metrics. Maintaining the conditioning over the entire input sequence, expectation can be performed on the unknown parameter. Combining of the resulting metrics over the nuisance parameters $x_0^n : u_k$ is performed as a final step, leading to the final two soft metrics for u_k. Since operators $\sum_{x_0^n : u_k}$ and E_Θ commute, an additional choice is available for the evaluation of the metric in (9a). Here, the sequence combining is done initially, followed by the parameter elimination. Different soft metrics can also be defined by interchanging the $\max_{x_0^n : u_k}$ operator with the E_Θ operator in (9b). This option will not be pursued in this work, mainly because it does not appear to lead to rigorously expressed optimal structures.

A. Parameter-First Combining

We begin by deriving optimal algorithms for the evaluation of the soft outputs defined in (9a) and (9b) and more precisely the quantity $P(z_0^n, x_0^n)$. It is noted once more that these algorithms are optimal for a given FE processing at the receiver. The obvious approach is a straightforward evaluation of this likelihood for each of the $(N_x)^{n+1}$ input sequences. The procedure is concluded with the appropriate combining of these quantities (summation or maximization for $\mathrm{APP}_p(u_k)$ or $\mathrm{MSM}_p(u_k)$, respectively). This type of processing is based on the fact that the likelihood $P(z_0^n, x_0^n)$ can be computed recursively as in [25]

$$\begin{aligned} P\left(z_0^k, x_0^k\right) &= P\left(z_k \mid z_0^{k-1}, x_0^k\right) P(x_k) P\left(z_0^{k-1}, x_0^{k-1}\right) \\ &= N\left(z_k; q_k^T \tilde{g}_{k\,|\,k-1}; N_0 + q_k^T \tilde{G}_{k\,|\,k-1} q_k^*\right) \\ &\quad \times P(x_k) P\left(z_0^{k-1}, x_0^{k-1}\right) \end{aligned} \tag{10}$$

[4]We assume a time-invariant model for notational and expositional simplicity. All results generalize to the time-variant case.

[5]A necessary and sufficient condition for stationarity of $\{g_k\}$ as well as the time-reversed process $\{g_{-k}\}$ is that the covariance of g_k satisfies the equation $K_g = GK_gG^+ + Q$, $G^b = K_gG^+K_g^{-1}$ and $Q^b = K_g - K_gG^+K_g^{-1}GK_g$, where $(\cdot)^+$ denotes complex conjugate and transpose.

where $N(z;m;\sigma^2)$ denotes the probability density function of a complex circular Gaussian random variable with mean m, and variance $\sigma^2/2$ for the real and imaginary part, while $\tilde{g}_{k\,|\,k-1}$ and $\tilde{G}_{k\,|\,k-1}$ are the channel one-step prediction and corresponding covariance matrix generated by a sequence-conditioned Kalman filter (KF). This technique, although efficient, results in suboptimal algorithms where complexity and smoothing depth are exponentially coupled, as mentioned in Section I.

An alternative optimal procedure for the likelihood calculation, based on which, several suboptimal useful algorithms will be developed in the next section, is now described. We observe that, due to the presence of the parameter process $\{g_k\}$, future observations depend on past observations conditioned on the state of the FSM. On the other hand, by conditioning on the parameter g_k as well, separation of the future and past observations occurs, yielding (11), shown at the bottom of the page (see the Appendix for the details of the derivation). The relation in (11) and subsequent analogous expressions are the basis for the practical algorithms proposed in Section IV and is a key contribution of this paper. It indicates that the likelihood can be split into three factors, of which the first two depend each on the past/present and future, respectively, while the third can be viewed as a weighting factor that binds them together. Indeed, the third factor quantifies the dependence of the future, present and past that is introduced due to the parameter process $\{g_k\}$ and in the absence of parametric uncertainty would be eliminated. An alternative interpretation can be offered by realizing that the expression in (11) is closely related to the total mean square error of a sequence-conditioned Kalman smoother. A closed-form expression can be found for the binding factor since it involves an integral of Gaussian densities (see the Appendix), and although the expression is fairly complicated (it involves inverse matrices and matrix determinants), we emphasize that it does not require any reprocessing of the observation record. The first factor in (11) is recursively evaluated using (10), while the second is calculated through a similar backward recursion.

$$\begin{aligned} P\left(z_{k+1}^n, x_{k+1}^n \,\middle|\, s_{k+1}\right) &= P\left(z_{k+1} \,\middle|\, z_{k+2}^n, s_{k+1}, x_{k+1}^n\right) P(x_{k+1}) \\ &\quad \times P\left(z_{k+2}^n, x_{k+2}^n \,\middle|\, s_{k+2}\right) \\ &= N\left(z_{k+1}; q_{k+1}^T \tilde{g}_{k+1\,|\,k+2}^b; N_0 + q_{k+1}^T \tilde{G}_{k+1\,|\,k+2}^b q_{k+1}^*\right) \\ &\quad \times P(x_{k+1}) P\left(z_{k+2}^n, x_{k+2}^n \,\middle|\, s_{k+2}\right). \end{aligned} \tag{12}$$

The scheme suggested by (10)–(12) is illustrated in Fig. 1 and can be described as follows. Starting at time 0, a forward N_x-ary tree is built, each node of which represents a sequence path. The likelihood $P(z_0^{k-1}, x_0^{k-1})$, together with $\tilde{g}_{k\,|\,k-1}$ and $\tilde{G}_{k\,|\,k-1}$ of that path are stored in each node. At each time k, the tree is expanded forward and the probabilities corresponding to the newly generated branches are calculated using (10). It is implied from this equation that a KF that depends on the entire path history is required to complete the recursion. Similarly, starting at time n, a backward tree is expanding according to the recursion (12). The relevant channel estimates are provided by a per-path backward running KF. After $k+1$ forward and $n-k$ backward steps, the two trees meet each other. The likelihood of each sequence x_0^n can now be evaluated as indicated by (11). The $(N_x)^{k+1}$ likelihoods corresponding to the nodes of the forward tree are combined with the $(N_x)^{n-k}$ likelihoods corresponding to the nodes of the backward tree (future) and weighted by the binding factor in (20). The final soft output for a generic quantity u_m is the summation (or maximization) over all factors with the same u_m. Note that the choice of k, the particular point in time when the past and future metrics are combined, is *completely arbitrary* (i.e., it is not related to m). In fact, the two extreme values $k=n$ and $k=0$ correspond to a single forward or a single backward tree. In a practical algorithm, however, the reference point k is chosen to be in the neighborhood of m, in order to maximize the number of relevant sequences combined to produce the soft information on u_m. Thus, while it may seem redundant to store and update both a forward and a backward tree (i.e., same result can be accomplished with a single forward tree), the fact

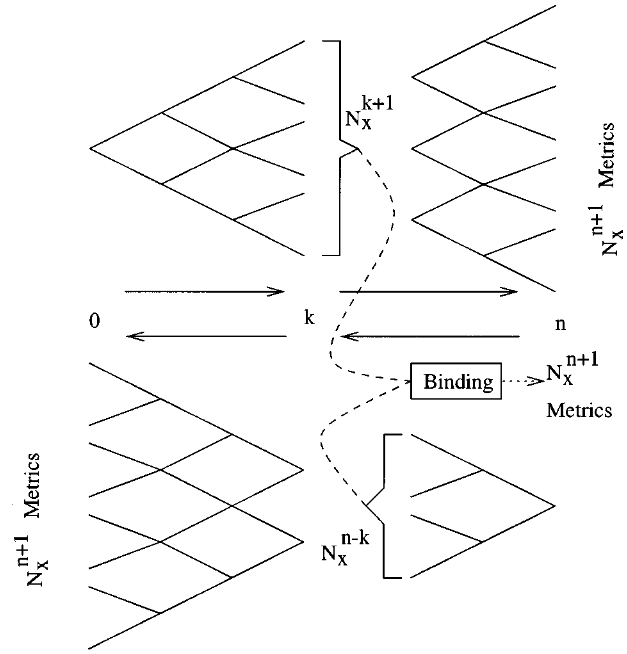


Fig. 1. Likelihood evaluation using forward/backward tree structures.

$$P(z_0^n, x_0^n) = \underbrace{P\left(z_0^k, x_0^k\right)}_{\text{past/present}} \underbrace{P\left(z_{k+1}^n, x_{k+1}^n \,\middle|\, s_{k+1}\right)}_{\text{future}} \underbrace{\int_{g_k} \frac{P\left(g_k \,\middle|\, x_0^k, z_0^k\right) P\left(g_k \,|\, s_{k+1}, x_{k+1}^n, z_{k+1}^n\right)}{P(g_k)} \, dg_k}_{\text{binding } b_p(\cdot)} \tag{11}$$

that the two trees can be pruned independently, decouples complexity and observation length, leading to practical algorithms, as will be discussed in Section IV.

B. Sequence-First Combining

The special form of $\mathrm{APP}_p(u_k)$ allows us to obtain alternative expressions for the optimal soft outputs by realizing that we can interchange the expectation operators in (9a), to obtain $\mathrm{APP}_p(u_k) = cP(z_0^n, u_k)$. In particular, a straightforward expression for $\mathrm{APP}_p(t_k)$ can be derived by utilizing the fact that the process $\{(t_k, g_k)\}$ is a mixed-state Markov chain. Unfortunately, the storage requirement for these recursions is infinite due to the fact that g_k takes values in a continuous space, making it of primarily conceptual value.[6] Although it is conceivable to quantize the channel values, we will follow another approach. A derivation similar to (11) leads to (13), shown at the bottom of the page. The forward and backward recursions for the first two quantities are as follows:

$$P\left(z_0^k, s_{k+1}\right) = \sum_{t_k : s_{k+1}} P\left(z_0^{k-1}, s_k\right) \times P\left(z_k \,|\, t_k, z_0^{k-1}\right) P(x_k) \tag{14a}$$

$$P\left(z_{k+1}^n \,\middle|\, s_{k+1}\right) = \sum_{t_{k+1} : s_{k+1}} P\left(z_{k+1} \,|\, t_{k+1}, z_{k+2}^n\right) \times P(x_{k+1}) P\left(z_{k+2}^n \,\middle|\, s_{k+2}\right). \tag{14b}$$

Aside from the evident similarity of (13) and (14) with (11), (10), and (12), there are two important differences as follows: 1) the recursions described here do not depend (at least explicitly) on the entire path history and 2) the evaluation of the third factor of (13) as well as the innovation factors in (14) is complicated due to the fact that they are mixed-Gaussian densities. Nevertheless, assuming that the latter difficulty can be overcome, the algorithm suggested by (13) and (14) is much simpler: only a forward and backward recursion is performed over a state trellis, followed by a combining (multiplication) of the updated quantities with an appropriate weight (third factor). This procedure is depicted in Fig. 2. Once more, we emphasize that the generalized states s_k^s and transitions t_k^s can be used with the corresponding updating equations unchanged.

C. Comments on the Deterministic Parameter Model

In the case when the unknown parameter is modeled as a deterministic constant, and expectation over the unknown Θ is not feasible, a reasonable soft output choice is

$$\mathrm{APP}_d(u_k) \stackrel{\mathrm{def}}{=} c \sum_{x_0^n : u_k} \max_{\Theta} P\left(z_0^n, x_0^n \,|\, \Theta\right) \tag{15}$$

[6]These recursions are basically the well-known BCJR [26] recursions for a mixed-state Markov process.

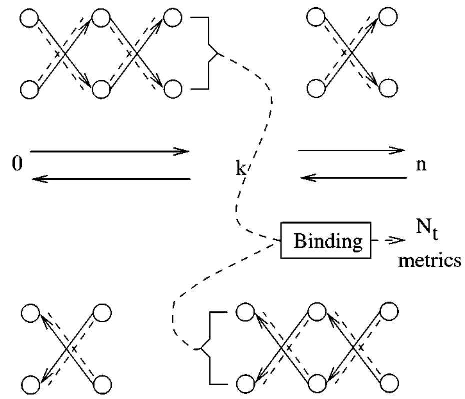


Fig. 2. Soft-metric evaluation in the case of sequence-first combining.

$$\mathrm{MSM}_d(u_k) \stackrel{\mathrm{def}}{=} c' - \log\left[\max_{x_0^n : u_k} \max_{\Theta} P\left(z_0^n, x_0^n \,|\, \Theta\right)\right]. \tag{16}$$

The development of the exact expressions for this modeling option is similar to that associated with the GM channel, and is not presented here for brevity (refer to [27] and [28] for a more detailed presentation). The resulting expressions are structurally similar with the main difference being the channel estimator, which is a recursive least-squares (RLS) estimator instead of the KF. Similarly to the GM case, by exchanging the order of maximization in (16), sequence-first expressions can be developed as well.

IV. Suboptimal (Fixed-Complexity) Algorithms

The exact evaluation of the soft metrics developed in the previous section involves likelihood updates on a forward and backward tree, assisted by per-path filters, followed by binding of the past and future metrics. In view of this fact, any suboptimal algorithm for the case of parameter-first combining can be interpreted as the result of applying one or more of the following simplifications: 1) nonexhaustive tree search; 2) non-Kalman channel estimators; and 3) suboptimal binding of the past and future metrics. Similarly, for the case of sequence-first combining, any suboptimal algorithm is the result of a simplifying assumption for the innovation factors, as well as a simpler form for the channel estimators and binding factor in (13). In the following, this design space is partially explored.

A. Parameter-First Combining

1) Tree-Search Techniques: Regarding the tree search, many options are available to prune the sequence tree (e.g., from the hard-decision literature [29]). Breadth-first schemes seem to be the most appropriate for soft-decisions, since they maintain a common front in the search process, which facilitates the combining task. One such algorithm is the VA, which

$$P\left(z_0^n, t_k\right) = P\left(z_0^{k-1}, s_k\right) P\left(z_{k+1}^n \,\middle|\, s_{k+1}\right) \underbrace{\int_{g_k} \frac{P\left(g_k \,|\, s_k, z_0^{k-1}\right) P(z_k \,|\, t_k, g_k) P(x_k) P\left(g_k \,|\, s_{k+1}, z_{k+1}^n\right)}{P(g_k)} \, dg_k}_{b'_p(\cdot)} \tag{13}$$

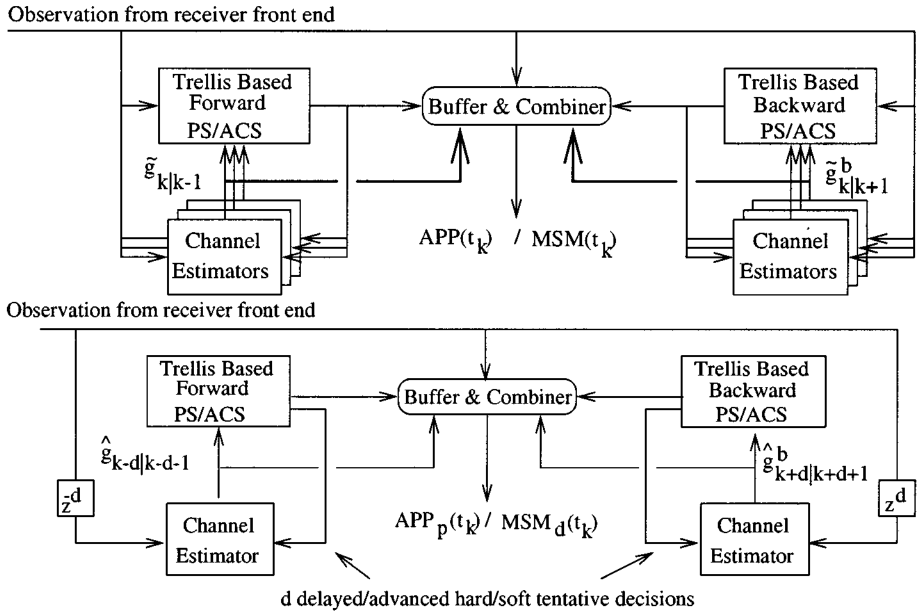


Fig. 3. Trellis-based practical A-SISO algorithms with a multiple or single estimator.

maintains and updates—through the familiar add compare select (ACS) operations—a fixed number of paths in such a way that they are forced to have different recent paths. Given that a set of paths—at the same depth—is available, an algorithm for evaluating $\mathrm{MSM}(\cdot)$ metrics, proceeds by extending and eliminating paths in the same way as in the hard-decision case [22], while the completion is performed by minimizing the corresponding transition metrics. The formulation of a practical algorithm for calculating $\mathrm{APP}(\cdot)$ metrics involves summation of the sequence metrics as well as tree pruning. An algorithm that combines these two tasks can be derived employing either the PSP principle [22], or equivalently, the decision-feedback (DF) assumption introduced in [30].

The resulting FI algorithms, shown in Fig. 3, consist of a single forward and backward recursion over the entire observation record, similar to the ones performed in the classical SISO. Product sum (PS) or ACS operations are performed for the metric updates, for APP or MSM soft metrics, respectively. A KF channel estimate is kept for every trellis state and updated in a PSP [22] fashion. The soft outputs for x_k and y_k are derived from the soft output of the transition t_k. The latter is computed as the product (sum) of the forward metric of the starting state s_k, the transition metric of t_k, the backward metric of the ending state s_{k+1}, and the binding factor corresponding to t_k. At this point, we emphasize once more that the trellis on which this algorithm operates is not tightly related to the FSM trellis. Its size is a design parameter that determines the amount of pruning in the forward and backward trees, and eventually, the complexity of the algorithm.

2) Channel Estimate and Binding Factor Simplification: Any near-optimal receiver has to search over as many paths as possible for a given amount of resources, so it is desirable to reduce the complexity associated with the metric updates and in particular the channel estimates. One such simplification is to substitute KF channel estimation with the least mean-squares (LMS) algorithm, so no matrix storage and update is required. In [28], this simplification is derived in a more rigorous manner for the case of deterministic parameter model, resulting in a simple and insightful expression for the binding factor shown in (17).

$$-\log b_p(\cdot) \sim \left\| \tilde{g}_{k\,|\,k-1} - \tilde{g}^b_{k+1\,|\,k+2} \right\|^2. \tag{17}$$

The above expression can be interpreted as follows. If the forward and backward channel estimates corresponding to a particular sequence are not consistent, a penalty is paid by means of increasing the sequence metric.

B. Sequence-First Combining

1) Metric Simplification: Starting from (14), suboptimal algorithms can be derived by employing a simplifying assumption for the innovation factors $P(z_k\,|\,t_k, z_0^{k-1})$, $P(z_{k+1}\,|\,t_{k+1}, z_{k+2}^n)$, which are in reality mixed-Gaussian density functions. The Gaussian approximation for the above innovation terms leads to an attractive algorithm since only the state-conditioned/sequence-averaged forward (i.e., $\tilde{g}_{k\,|\,k-1}(s_k) = E(g_k\,|\,s_k, z_0^{k-1})$) and backward channel one-step predictions together with the corresponding covariances need to be maintained and updated. Note that these estimates are only partially conditioned on the data sequence through the state s_k (or more generally the super-state s_k^s). Recursive update equations for these partially conditioned (PC) channel estimates, first derived in [12], are very similar to the KF recursions, thus we use the name PCKF. Furthermore, in the limiting case when the super-state represents the entire sequence, the innovation factors become precisely Gaussian and the PCKF becomes the sequence-conditioned KF; this is the exact scenario of the parameter-first combining in the GM case. Under the Gaussian assumption, a closed-form expression for the binding factor in (13) can be derived as well, resulting in a function similar to $b_p(\cdot)$ (see the Appendix for details).

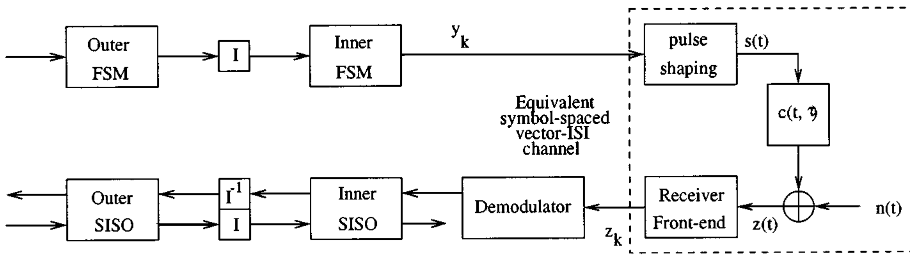


Fig. 4. Serial concatenation of FSMs and the associated iterative detection network for the case of perfect CSI.

2) Further Channel Estimator Simplification: In addition to the Gaussian approximation, a further simplification occurs under the assumption that the conditional means and covariances of the channel are not functions of the states $E(g_k \,|\, s_k, z_0^{k-1}) \approx E(g_k \,|\, z_0^{k-1}) = \hat{g}_{k\,|\,k-1}$. This approximation—if valid—results in a desirable solution, since only a single forward and a single backward global estimator (averaged over the sequence) needs to be maintained and updated. Assuming that a probabilistic description $P'(t_k)$ is available for the transitions t_k, a recursion can be derived for $\hat{g}_{k\,|\,k-1}$. The application of this single-estimator idea is inhibited, since 1) the above approximation is not valid and 2) an accurate $P'(t_k)$ can only be derived from the observation z_0^k and is therefore tightly coupled with the estimation process. Both 1) and 2) are alleviated by introducing a delayed (advanced) by d channel estimate to evaluate the forward (backward) transition metric at time k, since by increasing the decision delay d, the accuracy of the approximation

$$E\left(g_{k-d} \,|\, s_k, z_0^{k-d-1}\right) \approx E\left(g_{k-d} \,\middle|\, z_0^{k-d-1}\right) = \hat{g}_{k-d\,|\,k-d-1} \tag{18}$$

is improved. The resulting recursion equations, summarized in the Appendix, closely resemble those of the KF. The intuitive justification of this algorithm is that since a probabilistic description of t_{k-d}—and consequently y_{k-d}—exists, an average $\hat{q}_{k-d\,|\,k-d-1} = \sum_{t_{k-d}} q_{k-d} P'(t_{k-d})$ can be used in place of q_{k-d} in the KF recursions, thus resulting in what we refer to as an average KF (AKF). The resulting A-SISO, that utilizes a d-lag (d-advanced) soft-decision-directed forward (backward) AKF, is depicted in Fig. 3, and proceeds as follows. The forward metrics at time k are updated as in (14a) using the d-delayed channel estimate $\hat{g}_{k-d\,|\,k-d-1}$. Starting at time k, a d-step nonadaptive backward recursion is performed, at the end of which, a smoothed soft metric $P'(t_{k-d}) = P(t_{k-d} \,|\, z_0^k)$ is obtained. The latter is now used in the AKF to update $\hat{g}_{k-d\,|\,k-d-1}$. A similar one-step adaptive backward/d-step nonadaptive forward recursion is required for the update of the backward quantities.

C. Interpretation of Existing Algorithms

By dropping the backward recursions in (11) or (13), the forward-only A-SISO algorithms proposed in the literature can be derived: The algorithm in [11] calculates $\mathrm{APP}_p(x_k)$ soft outputs in an FL configuration, using the T-algorithm [29] for path pruning and employing KF for channel estimation. To achieve the desired smoothing depth D, the forward algorithm is developed based on the super state $s_k^s = (t_{k-d}, \ldots, t_{k-1}, s_k)$, where d is selected such that x_{k-D} is included in s_{k+1}^s. Similarly, in [13], a forward-only recursion is considered to produce $\mathrm{APP}_d(x_k)$ and $\mathrm{MSM}_d(x_k)$ soft outputs for the special FL case of the delay being equal to the channel length, with the VA used to prune the tree, and RLS channel estimation. The algorithm described in [12] is a forward-only special case of the A-SISO employing per-state PCKF. Although this algorithm was not intended to provide soft decisions, the metric updates and channel recursions (in the form of the PCKF) are precisely those developed therein. The A-SISO of [14] is an FL, forward only version, of the single-estimator (AKF) A-SISO, operating on the super-trellis $s_k^s = t_{k-1}$ with $d = 0$. Although the zero tentative decision delay eliminates the need for additional backward recursions, it seriously compromises the accuracy of the approximation in (18), motivating the nonzero delay d proposed herein.

V. TCM in Interleaved Frequency-Selective Fading Channels

A. Receiver Structures

As mentioned earlier, the TCM system can be modeled as a serial concatenation of two FSMs—the outer TCM encoder and the inner ISI channel—through the interleaver.

In [4], three receiver types were identified for the case of perfect CSI. They included the traditional hard-decision Viterbi equalizer[7] (VE) followed by a Viterbi decoder (VD), as well as the more sophisticated iterative structure shown in Fig. 4. An adaptive receiver can be derived in a straightforward way from the nonadaptive version, by replacing the inner detector (i.e., the equalizer) with its adaptive equivalent, while leaving the outer detector (i.e., the decoder) intact. In the more traditional hard-decision scheme, the VE is replaced by either a CA-MLSD VE [23], or a PSP-based VE [22], while in a soft-decision iterative receiver, one of the A-SISOs proposed herein is used in place of the inner SISO. An additional distinction of the adaptive iterative receiver from the nonadaptive version proposed in [7] and depicted in Fig. 4, is that the demodulator needs to be incorporated within the inner A-SISO.

Although there are many possible A-SISOs arising from the framework in Section III, we only utilize trellis-based algorithms. Several notes on the details of the implementation follow.

- APP algorithms operating in the log domain, result in a small complexity increase compared to MSM as reported

[7]The term "equalizer" is only used to signify that the particular VA is associated with the inner FSM, i.e., the ISI channel. We emphasize that this does not imply that linear or DF equalization is taking place.

in [7]. Indeed, all APP algorithms can be constructed from their MSM counterparts by replacing the $\min(x, y)$ function in the ACS operation by $\min^* = \min(x, y) - \log(1 + \exp(-|x - y|))$.

- Trellis-based multiple-estimator structures store and update one estimator per state with zero delay, while single-estimator schemes require d backward steps—for every forward step—to provide reliable tentative soft or hard data estimates to update their single estimator.
- Regarding the particular channel estimator used, the complexity increases in the order LMS, RLS, KF, AKF, PCKF, with the KF and the AKF having almost equal complexity.
- Optimal binding is, in general, a costly operation as shown in (20) and (21), while the suboptimal binding proposed in (17) results in a small increase in the adaptive SISO complexity.
- Forward-only algorithms have significantly lower requirements in computation and memory than forward/backward algorithms with the same number of states, since they do not require the additional backward recursion and binding. As was discussed in Section I, however, the exponential dependence of complexity and smoothing depth D is expected to give rise to much higher overall requirements for forward-only algorithms, if the performance of forward/backward algorithms is to be obtained.

B. Numerical Results and Discussion

Simulations were run for a transmission scheme comparable to GSM [31]. The convolutionally encoded sequence is interleaved using a 57 × 30 block interleaver. Each interleaver column is formatted into a TDMA burst together with a training sequence, equally split in 13 leading and 13 trailing symbols. Each burst is modulated and sent over a three-tap equal power Rayleigh fading channel (each tap is assumed independent from the others) with normalized Doppler spread $\nu_d = 0.005$. Referring back to the generic model in (4b), the above-described scenario corresponds to a system with root-raised cosine pulses, symbol-spaced independent fading taps (i.e., $T_r = T$) and a whitened-matched-filter symbol-spaced (i.e., $T_s = T$) receiver FE. Although the decorrelation time of such a channel is much larger than 57 symbols, for the purpose of simulation efficiency, a smaller interleaver depth is used in conjunction with the assumption of burst-to-burst independent channel. Three systems are considered as follows: i) a rate $1/2$, 16-state coded QPSK system (S1); ii) a rate $2/3$, 32-state coded 8PSK system (S2); iii) and an uncoded QPSK system (S3). Regarding the naming of the presented algorithms, each algorithm is identified by a four-part label, each part of which denoting: 1) the type of the soft decision (i.e., APP or MSM); 2) the multiplicity of the channel estimators (i.e., SING or MULT); 3) the particular channel estimator used (i.e., KF, RLS, LMS, AKF); and 4) the binding method [i.e., optimal binding (OB), suboptimal binding (SB), or no binding (NB)]. The trellis size of all algorithms considered here is chosen to be the same as the size of the underlying FSM trellis.

Fig. 5 presents performance curves for system S1, employing the iterative receiver described in the previous section with

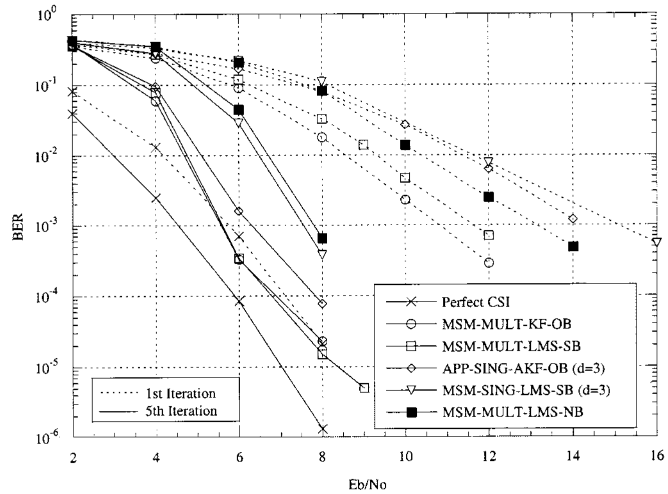


Fig. 5. BER versus E_b/N_0 for system S1 and various configurations for the inner A-SISO. Performance is compared between (i) MSM-MULT-LMS-SB and MSM-MULT-LMS-NB, (ii) MSM-MULT-LMS-SB and MSM-MULT-KF-OB, and (iii) MULT and SING.

different A-SISOs for the inner equalizer. Bit-error rate (BER) curves for the first and fifth iteration are shown; no significant improvement was observed for more than five iterations. For the A-SISOs employing KF or AKF, the channel estimators were obtained by approximating the Clarke spectrum [21] with a first order model having 10-dB bandwidth equal to ν_d. Comparing the two curves corresponding to MSM-MULT-LMS, a loss of 2 dB (1 dB) is observed for the fifth (first) iteration when no binding is performed. This outcome clearly indicates the significant practical—aside from the conceptual—value of the binding factor. The comparison between MSM-MULT-LMS-SB and MSM-MULT-KF-OB shows that LMS channel estimation with suboptimal binding is nearly as good as the KF with optimal—and computationally expensive—binding. In the first iteration, the latter performs slightly better (by 0.7 dB at $\mathrm{BER} = 10^{-3}$), while in the fifth iteration, no notable difference is observed. Multiple-estimator schemes are shown to be 2–4 dB better than single-estimator counterparts in the first iteration, while this gain is decreased to 0.5–2 dB after the fifth iteration as can be observed from the comparison of MSM-MULT-LMS-SB and MSM-MULT-KF-OB with MSM-SING-LMS-SB or APP-SING-AKF-OB. Note that the optimal value for the tentative delay was found to be $d = 3$ for both SING estimators. The best A-SISO achieves performance that is just 1 dB away from that of perfect CSI. Regarding the iteration gain, as much as 6–7 dB can be gained using five iterations for both single- or multiple-estimator SISOs. This result is the direct antithesis with the perfect CSI case, where an iteration gain of only 1 dB does not even justify the need for ID. Finally, simulation results that are not shown here confirm the negligible difference between APP and MSM algorithms for these operational SNRs, a fact which was noted in [4] and [32] for the case of CSI as well.

In Fig. 6, the performance of MSM-MULT-LMS-SB of Fig. 5 is compared with that of the corresponding receiver employing a forward-only A-SISO (as the one in [11]) with decision delays $D = 3, 4$, and 5 symbols. Other than the different inner

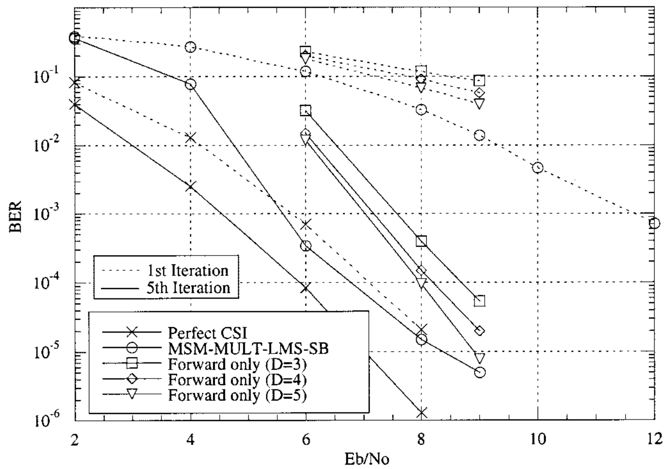


Fig. 6. Comparison between forward/backward and forward-only inner A-SISOs for system S1, for various values of the decision lag D.

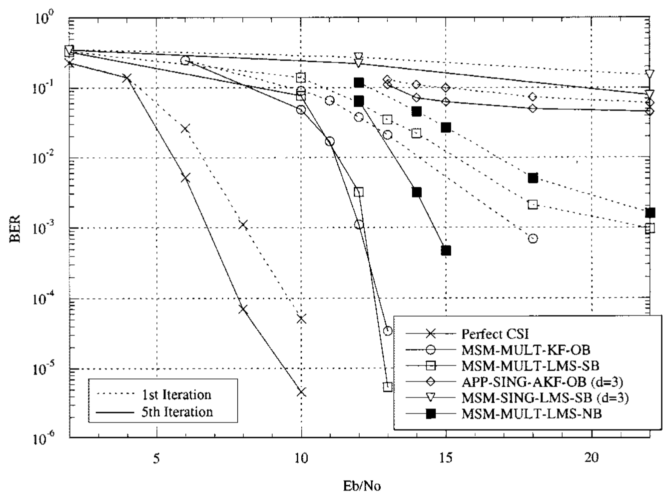


Fig. 7. BER versus E_b/N_0 for system S2 and various configurations for the inner A-SISO. Performance is compared between (i) MSM-MULT-LMS-SB and MSM-MULT-LMS-NB, (ii) MSM-MULT-LMS-SB and MSM-MULT-KF-OB, and (iii) MULT and SING.

A-SISOs, all other components of the compared receivers are identical. As expected, performance is improved by increasing the smoothing depth D, but gives rise to exponential complexity growth. The comparison with the proposed A-SISO shows that even with a high complexity forward-only algorithm ($D = 5$ corresponds to a 1024-state trellis) a performance gain of 1–1.5 dB can be achieved with the FI A-SISO with only a fraction of the complexity (a forward and a backward recursion on a 16-state trellis is required).

Similar performance curves are reproduced in Fig. 7 for system S2 over the same channel as in the previous simulation. The presence of the denser 8-PSK constellation produces quantitatively different performance curves. Single-estimator schemes reach an error floor at BER values greater than 10^{-2}, regardless of the channel estimator used (i.e., LMS or AKF). Multiple estimator algorithms using either KF and OB or LMS and SB perform almost identically at BERs smaller than 10^{-2}. Both of these adaptive algorithms yield much worse performance compared to perfect CSI (the loss is on the order of 5 dB for the fifth iteration for the best A-SISO at BER of 10^{-3}, while is reduced to approximately 3 dB for a BER of 10^{-5}).

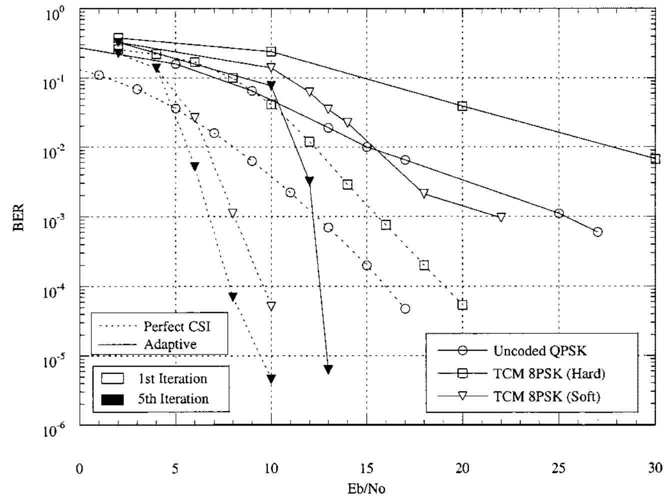


Fig. 8. BER versus E_b/N_0 for systems S2 and S3 employing hard-decision and soft-decision decoding BER versus E_b/N_0 for the receiver employing adaptive and nonadaptive (using interpolated channel estimates) inner SISOs for different payload sizes.

Coded modulation techniques have been considered as a method to provide improved performance (i.e., coding gain) with the only cost being increased receiver complexity (i.e., no bandwidth expansion). The design tradeoffs for this frequency-selective channel are more complex than those for an ideal AWGN channel. In [32], those tradeoffs were studied under the perfect CSI assumption. Fig. 8 presents a comparison between systems S3 (uncoded QPSK) and S2 (8PSK-TCM), both having the same throughput and occupying the same bandwidth. In the AWGN channel, S2 provides a 4.6-dB gain over the uncoded system. Similar to [4] and [32], conclusions are obtained for the case of perfect CSI: coding gain without bandwidth expansion is not possible using hard-decision receivers. The utilization of soft-decision receivers results in 4-dB coding gain at a BER of 10^{-3} for the first iteration. Additional iterations slightly improve the performance, resulting in 5.5-dB gain at the fifth iteration. When perfect CSI is not available, and adaptive processing is performed, the hard-decision PSP receiver still cannot provide any performance improvement over the uncoded system. Furthermore, the adaptive soft-decision algorithms, provides a poor coding gain when only a single iteration is performed (i.e., 3.5 dB). On the other hand, the use of iterative soft-decision adaptive processing results in a gain of approximately 13 dB.

C. Factors Impacting Performance

The conclusions drawn in the previous section are tightly coupled with the particular channel conditions and system configuration, and can be significantly altered when different operating conditions are considered. One channel characteristic, that has a significant effect on receiver design, is the level of dynamics (measured by the normalized Doppler spread ν_d). While high dynamics were considered here, in the case of low dynamics, the need for adaptive processing is questionable; an initial channel estimate may suffice for use in conjunction with a nonadaptive iterative detector. Similar conclusions have been drawn for adaptive hard-decision algorithms [33].

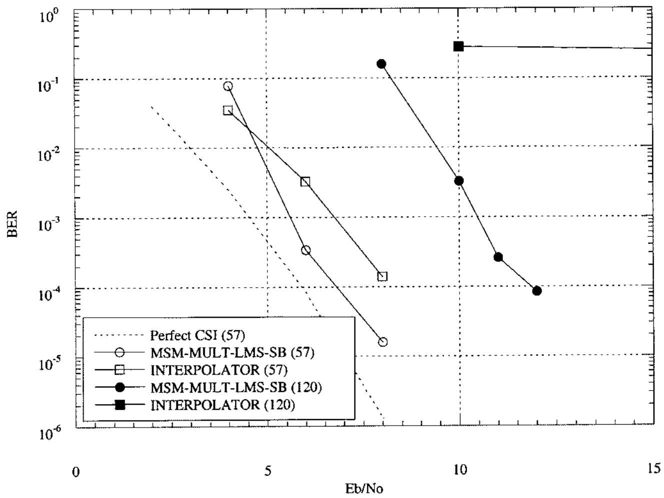


Fig. 9. BER versus E_b/N_0 for the receiver employing adaptive and nonadaptive (using interpolated channel estimates) inner SISOs for different payload sizes.

The signaling format and in particular the configuration of the training sequence is another system characteristic that has a great impact on receiver design. When only a leading training sequence is available, a more reasonable choice is to use an FL A-SISO. The design of an FL A-SISO is not a trivial extension of the FI scenario presented here, and is a topic of current research [34].

Regarding tracking versus acquisition operating mode, a relevant measure is the product of the payload (i.e., burst) size J with the normalized Doppler spread of the channel ν_d ($P = J \times \nu_d$); the smaller the value of P, the lower the probability of loosing lock. For systems operating with small P values, and utilizing leading/trailing training, a low complexity nonadaptive SISO algorithm derives channel estimates by linear interpolation between the initial and final channel estimates. In Fig. 9, the performance of this scheme is compared with that of MSM-MULT-LMS-SB for system S1. It is shown that the interpolator based nonadaptive SISO operates with 1-dB degradation compared to MSM-MULT-LMS-SB for a BER of 10^{-3} and payload size $J = 57$. Unfortunately, such high-performance/low-complexity A-SISO is not feasible when either a trailing training sequence is unavailable or when the value of P is increased. The latter is demonstrated in Fig. 9, where the doubling of payload size ($J = 120$) results in catastrophic performance for the interpolator based SISO.

VI. CONCLUSION

ID can be viewed as the exchange of soft information between "soft inverses" of each subsystem in the concatenated network, which combine and marginalize this information. In this paper, the soft inverse of a system with parametric uncertainty present was developed. The adaptive soft inverse of an FSM (i.e., the A-SISO) was the particular focus. It was demonstrated how an A-SISO can be used to perform AID with numerical results given for the TCM-ISI serially concatenated system. It was found that qualitative conclusions regarding performance (e.g., the iteration gain) are substantially different for the case with parametric uncertainty. In particular, iteration gains for the time-varying fading channel were considerably larger than the perfect CSI case.

By deriving the algorithms starting from a reasonable definition for soft outputs, we obtained several classes of practical adaptive forward–backward algorithms. A more detailed development of this general framework is contained in [27]. The resulting set of practical algorithms are, in hindsight, intuitive combinations of forward/backward SISOs for perfect CSI and adaptive hard-decision algorithms. Thus, one may now suggest several similar approaches based on existing hard-decision algorithms and/or SISO architectures. For example, the class of algorithms for the linear Gaussian fading channel that utilize steady-state finite-memory estimators [35], [36] can readily be adapted to a forward–backward SISO using the framework developed (i.e., the binding term). Similarly, generalization of AID to other activation schedules and architectures (i.e., FL SISOs, parallel message passing architectures, etc.) is an interesting area of current and future research.

APPENDIX

A. Proof of (11)

To prove (11), we first condition on the channel g_k and then use the fact that conditioned on the channel and the transition at time k, past and future observations are independent.

$$\begin{aligned}
&P\left(z_0^n, x_0^n\right)\\
&\quad= \int_{g_k} P\left(z_0^n, g_k, x_0^n\right)\, dg_k\\
&\quad= P\left(z_0^k, x_0^k\right) \int_{g_k} P\left(z_{k+1}^n, x_{k+1}^n \,\middle|\, z_0^k, x_0^k, g_k\right)\\
&\qquad\times P\left(g_k \,\middle|\, z_0^k, x_0^k\right)\, dg_k\\
&\quad= P\left(z_0^k, x_0^k\right) \int_{g_k} \frac{P\left(z_{k+1}^n, x_{k+1}^n, g_k \,|\, s_{k+1}\right)}{P(g_k \,|\, s_{k+1})}\\
&\qquad\times P\left(g_k \,\middle|\, z_0^k, x_0^k\right)\, dg_k\\
&\quad= P\left(z_0^k, x_0^k\right) P\left(z_{k+1}^n, x_{k+1}^n \,\middle|\, s_{k+1}\right)\\
&\qquad\times \int_{g_k} \frac{P\left(g_k \,\middle|\, x_0^k, z_0^k\right) P\left(g_k \,|\, s_{k+1}, x_{k+1}^n, z_{k+1}^n\right)}{P(g_k)}\, dg_k.
\end{aligned} \tag{19}$$

The closed-form expression for the binding factor is given by

$$\begin{aligned}
&b_P\left(\tilde{g}_{k\,|\,k}, \tilde{G}_{k\,|\,k}, \tilde{g}^b_{k\,|\,k+1}, \tilde{G}^b_{k\,|\,k+1}\right)\\
&\qquad= \frac{|K_g||P|}{|\tilde{G}_{k\,|\,k}|\left|\tilde{G}^b_{k\,|\,k+1}\right|} \exp(\beta^+ P\beta - \gamma)
\end{aligned} \tag{20a}$$

with

$$P^{-1} = \tilde{G}^{-1}_{k\,|\,k} + \left(\tilde{G}^b_{k\,|\,k+1}\right)^{-1} - K_g^{-1} \tag{20b}$$

$$\beta = \tilde{G}^{-1}_{k\,|\,k}\tilde{g}_{k\,|\,k} + \left(\tilde{G}^b_{k\,|\,k+1}\right)^{-1} \tilde{g}^b_{k\,|\,k+1} \tag{20c}$$

$$\begin{aligned}
\gamma &= \tilde{g}^+_{k\,|\,k}\tilde{G}^{-1}_{k\,|\,k}\tilde{g}_{k\,|\,k} + \left(\tilde{g}^b_{k\,|\,k+1}\right)^+\\
&\quad\times \left(\tilde{G}^b_{k\,|\,k+1}\right)^{-1} \tilde{g}^b_{k\,|\,k+1}
\end{aligned} \tag{20d}$$

where $\tilde{g}_{k\,|\,k}$, $\tilde{g}^b_{k\,|\,k+1}$ are the sequence-conditioned forward channel estimate, the one-step sequence-conditioned backward channel predictor and $\tilde{G}_{k\,|\,k}$, $\tilde{G}^b_{k\,|\,k+1}$ are the corresponding covariances.

B. Binding Factor Under the Gaussian Assumption for (13)

The binding factor in (13) under the Gaussian assumption is given by the expression

$$b'_p\left(\tilde{g}_{k\,|\,k-1}, \tilde{G}_{k\,|\,k-1}, \tilde{g}^b_{k\,|\,k+1}, \tilde{G}^b_{k\,|\,k+1}\right) = P(x_k)\frac{|K_g||P|}{|\tilde{G}_{k\,|\,k}|\left|\tilde{G}^b_{k\,|\,k+1}\right| N_0}\exp(\beta^+ P\beta - \gamma) \quad (21a)$$

with

$$P^{-1} = G^{-1}_{k\,|\,k} + {G^b_{k\,|\,k+1}}^{-1} - K_g^{-1} + \frac{q_k^* q_k^T}{N_0} \quad (21b)$$

$$\beta = G^{-1}_{k\,|\,k} g_{k\,|\,k} + {G^b_{k\,|\,k+1}}^{-1} g^b_{k\,|\,k+1} + \frac{q_k^* z_k}{N_0} \quad (21c)$$

$$\gamma = g^+_{k\,|\,k} G^{-1}_{k\,|\,k} g_{k\,|\,k} + {g^b_{k\,|\,k+1}}^+ {G^b_{k\,|\,k+1}}^{-1} g^b_{k\,|\,k+1} + \frac{|z_k|^2}{N_0} \quad (21d)$$

where $\tilde{g}_{k\,|\,k-1}$, $\tilde{g}^b_{k\,|\,k+1}$ are the PC one-step forward and backward channel predictors and $\tilde{G}_{k\,|\,k-1}$, $\tilde{G}^b_{k\,|\,k+1}$ are the corresponding covariances.

C. Channel Update Equations under the Gaussian Assumption and the Approximation in (18)

The forward recursions for the d-delayed AKF are given below. Backward recursions are similar. Assume that $P'(t_{k-d})$ is available.

$$\hat{q}_{k-d\,|\,k-d-1} = \sum_{t_{k-d}} q_{k-d} P'(t_{k-d}) \quad (22a)$$

$$\hat{Q}_{k-d\,|\,k-d-1} = \sum_{t_{k-d}} (q_{k-d} - \hat{q}_{k-d\,|\,k-d-1}) \times (q_{k-d} - \hat{q}_{k-d\,|\,k-d-1})^+ P'(t_{k-d}) \quad (22b)$$

$$K_{k-d} = \hat{G}_{k-d\,|\,k-d-1}\hat{q}^*_{k-d\,|\,k-d-1} \times \Big[N_0 + \mathrm{trace}(\hat{Q}^*_{k-d\,|\,k-d-1}\hat{G}_{k-d\,|\,k-d-1}) + \hat{q}^T_{k-d\,|\,k-d-1}\hat{G}_{k-d\,|\,k-d-1}\hat{q}^*_{k-d\,|\,k-d-1} + g^+_{k-d\,|\,k-d-1}\hat{Q}^*_{k-d}\hat{g}_{k-d\,|\,k-d-1}\Big]^{-1} \quad (22c)$$

$$\hat{g}_{k-d\,|\,k-d} = \hat{g}_{k-d\,|\,k-d-1} + K_{k-d} \times \left(z_{k-d} - \hat{q}^T_{k-d\,|\,k-d-1}\hat{g}_{k-d\,|\,k-d-1}\right) \quad (22d)$$

$$\hat{G}_{k-d\,|\,k-d} = \left(I - K_{k-d}\hat{q}^T_{k-d\,|\,k-d-1}\right)\hat{G}_{k-d\,|\,k-d-1} \quad (22e)$$

$$\hat{g}_{k-d+1\,|\,k-d} = G\hat{g}_{k-d\,|\,k-d} \quad (22f)$$

$$\hat{G}_{k-d+1\,|\,k-d} = G\hat{G}_{k-d\,|\,k-d}G^+ + Q. \quad (22g)$$

In the practical algorithm described in Subsection IV-B, $P'(t_{k-d}) = P(t_{k-d}\,|\,z_0^{k-1})$ is used as the tentative soft-decision.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their comments that helped improve the presentation of this work. In particular, they would also like to thank the anonymous reviewer who suggested that the issue of front-end processing be included in the development.

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Thitmajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. Int. Conf. Communications*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[2] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. Inform. Theory*, pp. 909–926, May 1998.

[3] A. Picart, P. Didier, and A. Glavieux, "Turbo-detection: A new approach to combat channel frequency selectivity," in *Proc. Int. Conf. Communications*, Montreal, PQ, Canada, 1997.

[4] A. Anastasopoulos and K. M. Chugg, "Iterative equalization/decoding of TCM for frequency-selective fading channels," in *Proc. Asilomar Conf. Signals, Systems, Comp.*, Nov. 1997, pp. 177–181.

[5] X. Chen and K. M. Chugg, "Near-optimal page detection for two-dimensional isi/awgn channels using concatenated modeling and iterative detection," in *Proc. Int. Conf. Communications*, Atlanta, GA, 1998.

[6] M. Moher, "An iterative multiuser decoder for near-capacity communications," *IEEE Trans. Commun.*, vol. 46, pp. 870–880, July 1998.

[7] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Soft-input soft-output modules for the construction and distributed iterative decoding of code networks," *Eur. Trans. Commun.*, vol. 9, pp. 155–172, Mar./Apr. 1998.

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

[9] J. G. Frias and J. D. Villasenor, "Turbo codes for binary markov channels," in *Proc. Int. Conf. Communications*, Atlanta, GA, 1998.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, pp. 1–38, 1977.

[11] Y. Zhang, M. P. Fitz, and S. B. Gelfand, "Soft output demodulation on frequency-selective Rayleigh fading channels using AR channel models," in *Proc. GLOBECOM Conf.*, Phoenix, AZ, Nov. 1997, pp. 327–331.

[12] R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filtering," *IEEE Trans. Commun.*, vol. 42, pp. 1017–1032, Feb./Mar./Apr. 1994.

[13] A. Anastasopoulos and A. Polydoros, "Adaptive soft-decision algorithms for mobile fading channels," *Eur. Trans. Commun.*, vol. 9, pp. 183–190, Mar./Apr. 1998.

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

[15] R. Mehlan, J. Wittkopp, and H. Meyr, "Soft output equalization and trellis coded modulation for severe frequency-selective fading channels," in *Proc. Int. Conf. Communications*, 1992, pp. 331.4.1–331.4.5.

[16] K. M. Chugg, "Sequence Estimation in the Presence of Parametric Uncertainty," Ph.D. dissertation, Univ. of Southern California, Los Angeles, CA, Aug. 1995.

[17] K. M. Chugg and A. Polydoros, "MLSE for an unknown channel—Part I: Optimality considerations," *IEEE Trans. Commun.*, vol. 44, pp. 836–846, July 1996.

[18] ——, "MLSE for an unknown channel—Part II: Tracking performance," *IEEE Trans. Commun.*, pp. 949–958, Aug. 1996.

[19] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[20] K. M. Chugg and K. Lerdsuwanakij, "Fading channel sequence detection based on rational approximations to the clarke doppler spectrum," in *Proc. Int. Conf. Telecommunications*, Chalkidiki, Greece, 1998.

[21] R. Clark, "A statistical theory of mobile radio reception," *Bell Sys. Tech. J.*, vol. 47, pp. 97–1000, 1968.

[22] R. Raheli, A. Polydoros, and C. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb./Mar./Apr. 1995.

[23] H. Kobayashi, "Simultaneous adaptive estimation and decision algorithm for carrier modulated data transmission systems," *IEEE Trans. Commun.*, vol. COM-19, pp. 268–280, June 1971.

[24] K. M. Chugg, "The condition for applicability of the Viterbi algorithm with implications for fading channel MLSD," *IEEE Trans. Commun.*, vol. 46, pp. 1112–1116, Sept. 1998.

[25] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for *a priori* unknown channels and symbol timing," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 579–588, Apr. 1992.

[26] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[27] A. Anastasopoulos, "Adaptive Soft-Input Soft-Output Algorithms for Iterative Detection," Ph.D. dissertation, Univ. of Southern California, Los Angeles, CA, Aug. 1999.

[28] A. Anastasopoulos and K. M. Chugg, "Adaptive SISO algorithms for iterative detection with parametric uncertainty," in *Proc. Int. Conf. Communications*, Vancouver, BC, Canada, June 1999.

[29] J. B. Anderson and S. Mohan, "Sequential coding algorithms: A survey cost analysis," *IEEE Trans. Commun.*, vol. COM-32, pp. 169–176, Feb. 1984.

[30] J. P. Seymour and M. P. Fitz, "Near-optimal symbol-by-symbol detection schemes for flat rayleigh fading," *IEEE Trans. Commun.*, vol. 43, pp. 1525–1533, Feb./Mar./Apr. 1995.

[31] R. Steele, *Mobile Radio Communications*. London, U.K.: Pentech, 1992.

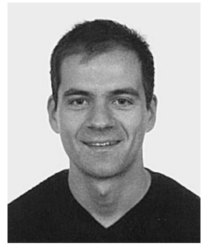
[32] A. Anastasopoulos and K. M. Chugg, "TCM for frequency-selective, interleaved fading channels using joint diversity combining," in *Proc. Int. Conf. Communications*, Atlanta, GA, June 1998.

[33] C.-K. Tzou, "Per-Survivor Processing: A General Approach to MLSE in Uncertain Environments," Ph.D. dissertation, Univ. Southern California, Los Angeles, Dec. 1993.

[34] J. Heo, K. M. Chugg, and A. Anastasopoulos, "A comparison of forward-only and bi-directional fixed-lag adaptive sisos," in *Proc. Int. Conf Communications (ICC)*, New Orleans, LA, June 2000, pp. 1660–1664.

[35] J. Lodge and M. Moher, "Maximum likelihood estimation of CPM signals transmitted over Rayleigh flat fading channels," *IEEE Trans. Commun.*, vol. 38, pp. 787–794, June 1990.

[36] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 43, pp. 1534–1544, Feb./Mar./Apr. 1995.



Achilleas Anastasopoulos (S'97–M'99) was born in Athens, Greece, in 1971. He received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1993, and the M.S. and Ph.D. degrees in electrical engineering from University of Southern California, Los Angeles, in 1995 and 1999, respectively.

He is currently an Assistant Professor at the University of Michigan, Ann Arbor, Department of Electrical Engineering and Computer Science. His scientific interests lie in the general area of communication theory, with emphasis in joint parameter estimation and data detection. His current research focuses on adaptive iterative detection, coding for fading channels, and receiver algorithms for multicarrier systems. He is co-author of the book *Iterative Detection: Adaptivity, Complexity Reduction, and Applications* (Norwell, MA: Kluwer Academic Press, scheduled for publication in late 2000).

Dr. Anastasopoulos is the recipient of the 1996 Myronis Fellowship from the University of Southern California.

Keith M. Chugg (S'88–M'95) was born in Los Angeles, CA, on March 16, 1967. He received the B.S. degree (with high distinction) in engineering from Harvey Mudd College, Claremont, CA, in 1989, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), Los Angeles, in 1990 and 1995, respectively.

During the 1995–1996 academic year, he was Assistant Professor with the Electrical and Computer Engineering Department, University of Arizona, Tucson. He is currently with the EE-Systems Department at USC, where he has been an Assistant Professor since 1996. He is a co-founder of Trellisware Technologies, Inc., a company dedicated to providing complete likelihood-based digital baseband solutions, where he serves as a consultant and member of the Technical Advisory Board. Along with his former Ph.D. students, A. Anastasopoulos and X. Chen, he is co-author of the book *Iterative Detection: Adaptivity, Complexity Reduction, and Applications* (Norwell, MA: Kluwer Academic Press, scheduled for publication in late 2000). His research interests include the general areas of signaling, detection, and estimation for digital communication and data storage systems. He is also interested in architectures for efficient implementation of the resulting algorithms.