A Unified Framework for Finite-Memory Detection

Gianluigi Ferrari, *Member, IEEE*, Giulio Colavolpe, *Member, IEEE*, and Riccardo Raheli, *Member, IEEE*

***Abstract*—In this paper, we present a general approach to finite-memory detection. From a semi-tutorial perspective, a number of previous results are rederived and new insights are gained within a unified framework. A probabilistic derivation of the well-known Viterbi algorithm, forward–backward, and sum-product algorithms, shows that a *basic metric* emerges naturally under very general *causality* and *finite-memory* conditions. This result implies that detection solutions based on one algorithm can be systematically extended to other algorithms. For stochastic channels described by a suitable parametric model, a conditional Markov property is shown to imply this finite-memory condition. This conditional Markov property, although seldom met *exactly* in practice, is shown to represent a reasonable and useful approximation in all considered cases. We consider, as examples, linear predictive and noncoherent detection schemes. While good performance for increasing complexity can often be achieved with a finite-memory detection strategy, key issues in the design of detection algorithms are the computational efficiency and the performance for limited complexity.**

***Index Terms*—Adaptive detection, finite-memory detection, forward–backward (FB) algorithm, graph-based detection, iterative detection, maximum *a posteriori* (MAP) sequence/symbol detection, sum-product (SP) algorithm, Viterbi algorithm (VA).**

I. Introduction

THE PROBLEM of decoding and detection over noisy channels has long been considered in the literature. In particular, the Viterbi algorithm (VA) [1], [2] provides an efficient way to implement the maximum *a posteriori* (MAP) sequence detection criterion, based on a suitable trellis diagram descriptive of the communication system. The key characteristic of the VA is the recursive computation of path metrics associated with a number of surviving paths which is kept constant and equal to the number of trellis states.

In [3], an efficient algorithm to compute the *a posteriori* probability (APP) of a particular symbol, with complexity on the same order of the VA, is proposed. This algorithm allows therefore to implement exactly the MAP symbol detection criterion. The trellis-based algorithm derived in [3] and usually termed, after the authors, BCJR algorithm, is based on a forward and a backward recursion, and it is thus also referred to as forward–backward (FB) algorithm. The discovery, in the early 1990s, of turbo codes [4]–[7] and of the concept of *iterative decoding*[1] has called for *soft-output* algorithms, i.e., algorithms computing "reliability values," rather than decisions, for the transmitted symbols. Hence, the FB algorithm, which allows the exact computation of the symbol APP, has received a new and ever-increasing attention since then.

In the last years, *low-density parity-check* (LDPC) codes, originally invented in [8], were also rediscovered [9]. One of the reasons behind the renewed interest for LDPC codes is the fact that a simple iterative message passing algorithm, operating on a graph descriptive of the linear block code [8], [10], can be devised for computing reliability values for the transmitted symbols. In particular, this algorithm is termed *sum-product* (SP) algorithm [11]. In [11], it is shown that if the *factor graph* corresponding to a linear block code does not contain cycles, then the SP algorithm leads to the exact computation of the APP of the transmitted symbols. For factor graphs with cycles, the soft-output value generated by the SP algorithm approximates well the APP in several important cases.

In this paper, we show that in order to perform detection for communications over channels with memory, possibly stochastic, the same *basic metric* is the key ingredient to implement any of the considered hard-output or soft-output detection algorithms (VA, FB algorithm, and SP algorithm). The algorithms are obtained through a probabilistic derivation based on minimal *causality* and *finite-memory* conditions. For instance, we show that a metric derived for a VA can be systematically extended to FB and SP algorithms. In particular, we point out that: 1) performing combined detection and decoding of *trellis codes* transmitted over channels with memory leads naturally to the introduction of an *augmented trellis* (with an increased number of states), whereas 2) in the case of a *Tanner* graph-based detection for linear block codes [10], taking into account the channel memory leads to the introduction of another *level* of nodes in the factor graph. In both cases, it is possible to conclude that there is an expansion of the original trellis or graph structures.

In stochastic channels with suitable parametric models, a *conditional Markov property* is shown to imply the finite-memory condition. Unfortunately, this property rarely holds in realistic channels. However, reasonable approximate detection algorithms can be derived. Unlike most of the existing works in the literature, the generality of the approach proposed in this paper resides in the fact that, *first*, a general conditional Markov property is assumed and, *then*, the channel statistical characterization is exploited. A few detection strategies (noncoherent and linear predictive) are considered to exemplify the general result in terms of special cases.

The rest of this paper is structured as follows. In Section II, causality and finite-memory conditions are introduced. In Section III, probabilistic derivations of VA, FB, and SP algorithms are proposed. In Section IV, detection for *stochastic* channels is considered and a conditional Markov property is introduced. In Section V, numerical results relative to a few

Manuscript received March 31, 2004; revised December 19, 2004 and February 28, 2005. This paper was presented in part at the Third International Symposium on Turbo Codes and Related Topics, Brest, France, September 2003.

The authors are with the Department of Information Engineering, University of Parma, 43100 Parma, Italy (e-mail: gianluigi.ferrari@unipr.it; giulio@unipr.it; raheli@unipr.it).

Digital Object Identifier 10.1109/JSAC.2005.853812

[1]This concept can be originally found in [8], but was crystallized in [7].

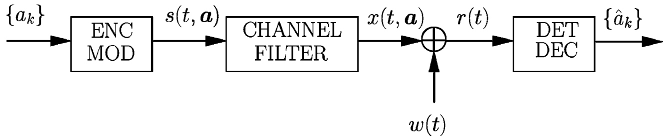

Fig. 1. Communication system.

significant examples are presented. Finally, concluding remarks are given in Section VI.

II. Causality and Finite Memory

We consider a generic time-continuous transmission system, depicted in Fig. 1. A sequence of independent and identically distributed M-ary information symbols $\{a_k\}$ are transmitted successively from epoch 0 to epoch $K-1$. A sequence of information symbols is denoted in vector notation as

$$\boldsymbol{a}_{k_1}^{k_2} = (a_{k_1}, a_{k_1+1}, \ldots, a_{k_2}) \quad k_2 \geq k_1.$$

For brevity, the entire sequence is denoted by $\boldsymbol{a}$. This sequence is input to the encoder and modulator. The coded and modulated signal is denoted as $s(t,\boldsymbol{a})$ to emphasize its dependence on the information sequence. The channel is viewed as a noiseless filter (possibly stochastic) with output signal $x(t,\boldsymbol{a})$, rendered noisy by the addition of white noise $w(t)$. The received signal $r(t)$ is observed by the demodulation and decoding block, which outputs a sequence of decisions $\{\hat{a}_k\}$.

The encoder/modulator block in Fig. 1 is a generic system which evolves, upon receiving at its input the information sequence $\boldsymbol{a}$, through a sequence of states $\{\mu_0, \mu_1, \ldots\}$. In many communication schemes, the encoder/modulator can be described as a time-invariant finite-state machine (FSM) (e.g., trellis coded modulation (TCM) [12] or continuous phase modulation (CPM) [13]). In this case, the state μ_k belongs to a set of finite cardinality and a *time-invariant* "next-state" function $\mathrm{ns}(\cdot,\cdot)$ describes the evolution of the system as

$$\mu_k = \mathrm{ns}(\mu_{k-1}, a_{k-1}). \tag{1}$$

Therefore, the evolution of the encoder/modulator can be described through a trellis diagram, in which there are M exiting branches (in correspondence with M different information symbols) from each state. A trellis branch corresponds to a transition, defined as $t_k \triangleq (\mu_k, a_k)$. In the rest of this paper, the initial state μ_0 is assumed to be known.

The received signal can be expressed as

$$r(t) = x(t,\boldsymbol{a}) + w(t). \tag{2}$$

By means of a *discretization process*, the received signal $r(t)$ can be converted into a time-discrete sequence $\boldsymbol{r}$ [14]. In particular, we assume that there is one observable r_k per information symbol a_k, or formally, $\boldsymbol{r} = \boldsymbol{r}_0^{K-1}$, with a notation similar to that used for the information sequence.

The considered discrete-time model is based on a sampling rate of one sample per symbol, which may be practically sufficient in many cases. In a more general setting, there may be two or more elements of $\boldsymbol{r}$ per information symbol a_k, e.g., when a convolutional code or a time-varying channel is considered.

A *causality condition* for the considered communication system can be formulated in terms of statistical dependence of the observation sequence $\boldsymbol{r}_0^k$, up to epoch k, on the information sequence. Accordingly, a system is causal if

$$p\left(\boldsymbol{r}_0^k|\boldsymbol{a}\right) = p\left(\boldsymbol{r}_0^k|\boldsymbol{a}_0^k\right). \tag{3}$$

Similarly, a *finite-memory condition* can be formulated, in statistical terms, as follows:

$$p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_0^k\right) = p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-C}^k, \mu_{k-C}\right) \tag{4}$$

where C is a suitable *finite-memory parameter* and μ_{k-C} represents the state, at epoch $k-C$, of the encoder/modulator. The finite-memory condition (4) is an extension of the *folding condition* introduced in [15], which accounts for the (possibly recursive) encoder/modulator state μ_k. The considered model includes any definition of state μ_k in terms of a suitable state variable, not necessarily defined in terms of input variables. It can easily be proved (see Appendixes I and II) that causality and finite-memory conditions imply the following equalities:

$$\begin{aligned} &p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-D}^k, \mu_{k-D}\right) \\ &\quad = p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-C}^k, \mu_{k-C}\right) \quad \forall D \geq C \end{aligned} \tag{5}$$

$$\begin{aligned} &p\left(\boldsymbol{r}_k^{K-1}|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_0^{k-1}\right) \\ &\quad = p\left(\boldsymbol{r}_k^{K-1}|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-C}^{k-1}, \mu_{k-C}\right). \end{aligned} \tag{6}$$

The first equality formalizes the intuition that considering past information symbols, before epoch $k-C$, adds no further information regarding the observation at epoch k. The second equality formalizes the idea that the finite-memory condition[2] extends to future observations beyond epoch k.

The encoder/modulator block in Fig. 1 can often be decomposed into the cascade of an encoder and a memoryless mapper. In this case, causality and finite-memory conditions imply analogous relations between the observation sequence $\boldsymbol{r}$ and *code* sequence $\boldsymbol{c} = \boldsymbol{c}_0^{K-1}$, where c_k is a generic code symbol. More precisely, a coded symbol c_k has to be interpreted as the discrete-time output of the encoder/modulator FSM according to a suitable "output" function $\mathrm{o}(\cdot,\cdot)$, such that $c_k = \mathrm{o}(\mu_k, a_k)$—this function was not clearly introduced at the beginning of this section because in the communication system model in Fig. 1 the output of the encoder/modulator is represented as a continuous-time signal. Accordingly, causality and finite-memory conditions can be formulated as follows:

$$p\left(\boldsymbol{r}_0^k|\boldsymbol{c}\right) = p\left(\boldsymbol{r}_0^k|\boldsymbol{c}_0^k\right) \tag{7}$$

$$p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{c}_0^k\right) = p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{c}_{k-C}^k\right). \tag{8}$$

We remark, however, that these conditions involve the transmission channel only and, in general, *do not* imply (3) and (4). A case of interest may be that of a linear block code followed by a memoryless modulator. In particular, a linear block code is not guaranteed to be causal and finite-memory[3] so that the channel causality (7) and finite memory (8) do not imply the system causality (3) and finite-memory condition (4).

[2]Note that there is a slight difference between the formal definition of the finite-memory condition (4) and (6), since in (6) the conditioning information sequence is $\boldsymbol{a}_0^{k-1}$ and does not include symbol a_k. This is, however, expedient for the derivation of the backward recursion of the FB algorithm in Section III-B.

[3]*Block-wise* causality and finite-memory must be indeed satisfied.

In the case of a linear block code, a trellis representation is possible, but the trellis is time-variant, both in terms of states and branches [3]. In this case, the evolution of the encoder/modulator could be described by a *time-variant* next-state function $\mathrm{ns}_k(\cdot,\cdot)$. A *Tanner graph* representation for a linear block code—where the parity checks determine the structure of the graph—may be more appealing, especially if the parity check equations involve a few code symbols as, for instance, in the case of LDPC codes [16].

III. Detection Strategies

Based on the statistical definition of causality and finite-memory conditions introduced in Section II, a probabilistic derivation of the principal detection/decoding algorithms is now presented. In particular, conditions (3) and (4) will be applied to *trellis-based* algorithms (VA and FB algorithm), whereas conditions (7) and (8) will be applied to *factor graph-based* algorithms (SP algorithm).

A. Viterbi Algorithm (VA)

The VA is an efficient method to perform MAP *sequence* detection. The causality and finite-memory conditions (3) and (4), the independence of the information symbols, and the chain factorization rule allow one to derive the following:

$$\begin{aligned} P\{\boldsymbol{a}|\boldsymbol{r}\} &\sim p(\boldsymbol{r}|\boldsymbol{a})P\{\boldsymbol{a}\} \\ &= \prod_{k=0}^{K-1} p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}\right)P\{a_k\} \\ &= \prod_{k=0}^{K-1} p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_0^{k}\right)P\{a_k\} \\ &= \prod_{k=0}^{K-1} p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)P\{a_k\} \end{aligned} \tag{9}$$

where the symbol $\sim$ indicates that two quantities are monotonically related with respect to the variable of interest (in this case, $\boldsymbol{a}$). Note that the last step in (9), where the finite-memory condition is applied, holds if $k \geq C$, i.e., in the algorithm "regime." In the initial transient period for $k < C$, (9) holds assuming that negative indeces are replaced by 0.

Defining augmented trellis state and branch (transition) as follows:

$$S_k \triangleq \left(\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right) \tag{10}$$

$$T_k \triangleq (S_k,a_k) = \left(\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right) \tag{11}$$

the kth factor in (9) can be expressed as

$$\gamma_k(T_k) \triangleq p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)P\{a_k\}. \tag{12}$$

As well known, the VA can now be formulated in the logarithmic domain, by defining the branch metric[4] $\lambda_k(T_k) \triangleq \log\gamma_k(T_k)$, and obtaining

$$\log P\{\boldsymbol{a}|\boldsymbol{r}\} \sim \sum_{k=0}^{K-1}\lambda_k(T_k).$$

[4]Following the definition of $\lambda_k(T_k)$ as metric, in the remainder of this paper, we will refer to $\gamma_k(T_k)$ as "exponential metric."

The MAP sequence, denoted as $\boldsymbol{a}^{\mathrm{MAP}}$, is such that

$$\boldsymbol{a}^{\mathrm{MAP}} = \underset{\boldsymbol{a}}{\operatorname{argmax}}\ \log P\{\boldsymbol{a}|\boldsymbol{r}\}.$$

The VA is an efficient trellis-based algorithm to determine $\boldsymbol{a}^{\mathrm{MAP}}$ [2]. Note that the state S_k is augmented *with respect to* the state μ_k of the encoder/modulator. This corresponds to considering *combined* detection and decoding. The next-state function for this augmented trellis diagram, denoted as $\mathrm{NS}(\cdot,\cdot)$, can be straightforwardly expressed as follows:

$$S_{k+1} = \mathrm{NS}(S_k,a_k) = \left(\boldsymbol{a}_{k-C+1}^{k},\mathrm{ns}(\mu_{k-C},a_{k-C})\right).$$

B. Forward–Backward (FB) Algorithm

The FB algorithm allows to implement the MAP *symbol* detection criterion, since it explicitly computes the APP $P\{a_k|\boldsymbol{r}\}$. Based on the causality and finite-memory conditions (3) and (4), the following probabilistic derivation of an FB algorithm is obtained by marginalization:

$$\begin{aligned} P\{a_k|\boldsymbol{r}\} &= \sum_{\boldsymbol{a}_{k-C}^{k-1}}\sum_{\mu_{k-C}} P\left\{\boldsymbol{a}_{k-C}^{k},\mu_{k-C}|\boldsymbol{r}\right\} \\ &\sim \sum_{\boldsymbol{a}_{k-C}^{k-1}}\sum_{\mu_{k-C}} p\left(\boldsymbol{r}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)P\left\{\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right\} \\ &= \sum_{\boldsymbol{a}_{k-C}^{k-1}}\sum_{\mu_{k-C}} p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^{k},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right) \\ &\quad\cdot p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right) \\ &\quad\cdot p\left(\boldsymbol{r}_0^{k-1}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)P\left\{\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right\} \end{aligned} \tag{13}$$

where Bayes and chain rules have been used. Based on (6) and causality, it follows that:

$$p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^{k},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right) = p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^{k},\boldsymbol{a}_{k-C+1}^{k},\mu_{k-C+1}\right).$$

Based on causality, one can also write

$$p\left(\boldsymbol{r}_0^{k-1}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right) = p\left(\boldsymbol{r}_0^{k-1}|\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right).$$

Recalling the independence of the information symbols, (13) can be rewritten as follows:

$$\begin{aligned} P\{a_k|\boldsymbol{r}\} \sim \sum_{\boldsymbol{a}_{k-C}^{k-1}}\sum_{\mu_{k-C}} & p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^{k},\boldsymbol{a}_{k-C+1}^{k},\mu_{k-C+1}\right) \\ &\cdot p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)p\left(\boldsymbol{r}_0^{k-1}|\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right) \\ &\cdot P\left\{\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right\}P\{a_k\}. \end{aligned}$$

Considering augmented state S_k and transition T_k as in (10) and (11), and defining

$$\begin{aligned} \beta_{k+1}(S_{k+1}) &\triangleq p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^{k},\boldsymbol{a}_{k-C+1}^{k},\mu_{k-C+1}\right) \\ \alpha_k(S_k) &\triangleq p\left(\boldsymbol{r}_0^{k-1}|\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right)P\left\{\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right\} \end{aligned}$$

the symbol APP in (13) can be finally expressed as

$$P\{a_k|\boldsymbol{r}\} \sim \sum_{S_k}\beta_{k+1}\left(\mathrm{NS}(S_k,a_k)\right)\gamma_k(S_k,a_k)\alpha_k(S_k)$$

where $\gamma_k(T_k)$ is defined in (12).

Based on the causality and finite-memory conditions, the quantities $\alpha_k(S_k)$ and $\beta_{k+1}(S_{k+1})$ can be computed by means of forward and backward recursions, respectively. More precisely, in Appendix III, it is shown that

$$\alpha_k(S_k) = \sum_{T_{k-1}:S_k} \alpha_{k-1}(S_{k-1})\gamma_{k-1}(T_{k-1}) \tag{14}$$

$$\beta_k(S_k) = \sum_{T_k:S_k} \beta_{k+1}(S_{k+1})\gamma_k(T_k) \tag{15}$$

where the notation $T_k : S_k$ indicates all transitions T_k compatible with state S_k. As usual, proper boundary conditions $\{\alpha_0(S_0)\}$ and $\{\beta_{K-1}(S_{K-1})\}$ must be specified. The algorithm can be also formulated in the logarithmic domain based on the branch metric $\lambda_k(T_k) = \log\gamma_k(T_k)$. In particular, defining

$$\overline{\alpha}_k(S_k) \triangleq \log\alpha_k(S_k)$$
$$\overline{\beta}_k(S_k) \triangleq \log\beta_k(S_k)$$

and the operator $\max *$ as

$$\max_u * f(u) \triangleq \log\sum_u e^{f(u)}$$

where u belongs to a discrete set and $f(\cdot)$ is a given function of u, the FB algorithm can be equivalently described by the following formulas:

$$\begin{aligned}\log P\{a_k|\boldsymbol{r}\} &\sim \max_{S_k} * \{\overline{\alpha}_k(S_k) + \lambda_k(T_k) \\ &\qquad + \overline{\beta}_{k+1}(\mathrm{NS}(S_k, a_k))\} \\ \overline{\alpha}_k(S_k) &= \max_{T_{k-1}:S_k} * \{\overline{\alpha}_{k-1}(S_{k-1}) + \lambda_{k-1}(T_{k-1})\} \\ \overline{\beta}_k(S_k) &= \max_{T_k:S_k} * \{\overline{\beta}_{k+1}(S_{k+1}) + \lambda_k(T_k)\}.\end{aligned}$$

The approximation $\max_u * f(u) \approx \max_u f(u)$ leads to a widely used approximated version of the FB algorithm, referred to as *max-log* [17] or *min-sum* [18].

We remark that the derived formulation of an FB algorithm for detection over channels with memory is based on very general causality and finite-memory statistical conditions, and it does not make any assumption on the specific nature of the channel. Since any detection strategy, designed for implementation with a VA, is solely defined in terms of specific branch metric λ_k and trellis state S_k, it is immediate to conclude that such a detection strategy can be *systematically extended* to an FB algorithm, and *vice versa*.

C. Sum-Product Algorithm

The application of the SP algorithm [11] to a factor graph representing the APP of the transmitted code sequence $\boldsymbol{c}$, given an observation sequence $\boldsymbol{r}$, allows the exact or approximate computation of the symbol marginal APPs [11]. Therefore, this algorithm may be used to implement a MAP *symbol* detection algorithm.

Since an information sequence is in a one-to-one correspondence with a coded sequence, the information sequence APP

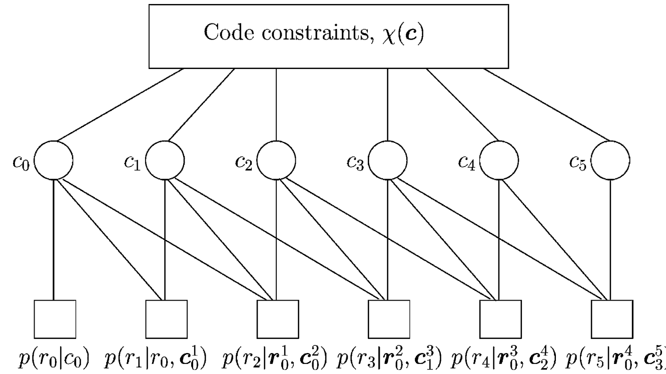

Fig. 2. Overall factor graph for $C = 2$.

may be conveniently expressed, in terms of the coded sequence, as

$$P\{\boldsymbol{a}|\boldsymbol{r}\} = P\{\boldsymbol{c}|\boldsymbol{r}\} \sim P\{\boldsymbol{c}\}p(\boldsymbol{r}|\boldsymbol{c}) = P\{\boldsymbol{c}\}\prod_{k=0}^{K-1} p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{c}_0^k\right)$$

where $\boldsymbol{c}$ is the *unique* coded sequence corresponding to $\boldsymbol{a}$ and the causality condition (7) has been used.

Assuming that the *a priori* distribution of the transmitted codewords is uniform and denoting by $\chi(\boldsymbol{c})$ the *code characteristic function*,[5] equal to 1 if $\boldsymbol{c}$ is a codeword and to zero, otherwise, under the finite-memory condition (8), we have

$$P\{\boldsymbol{c}|\boldsymbol{r}\} \sim \chi(\boldsymbol{c})\prod_{k=0}^{K-1} p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{c}_{k-C}^k\right).$$

The corresponding factor graph, representing both the code constraints [described by $\chi(\boldsymbol{c})$] and the channel behavior, is shown in Fig. 2 for $C = 2$. With respect to SP-based decoding schemes for linear block codes (e.g., LDPC codes) over a memoryless channel, *additional factor nodes* must be added at the bottom of the graph, as shown in Fig. 2. These additional factor nodes perform a marginalization, based on the channel model, without taking into account the code constraints—the application of the SP algorithm to this factor graph leads to a scheme for *separate* detection and decoding. This approach is different from that proposed in [19], where new variable nodes, representing the unknown channel parameters, are introduced.

The quality of the convergence of the SP algorithm to the exact marginal probabilities is in general determined by the *girth* of the graph.[6] As an example, in designing LDPC codes, cycles of length 4 must be avoided to ensure good decoding convergence. The graph derived from the proposed factorization has, in general, girth 4, involving the factor nodes which model the channel behavior. However, we verified by computer simulations that these length-4 cycles often do not affect the convergence of the algorithm (see [20] for more details). An important scenario where these length-4 cycles do affect the performance of the SP algorithm is given by the case of transmission over *intersymbol interference* (ISI) channels. In this case, however,

[5]In the hypothesis of a uniformly distributed codebook, $P\{\boldsymbol{c}\} = \chi(\boldsymbol{c})/M^K$, where M^K is the number of codewords.

[6]A *cycle* is a closed path in the graph and its *length* is defined as the corresponding number of path edges. The length of the shortest cycle is the *girth* of the graph.

particular factor graph transformations (e.g., *stretching*) can be used to overcome these limitations [21].

D. Exact Applications

Significant examples where the causality and finite-memory conditions strictly hold are involved in transmission over channels with finite ISI, possibly encompassing a *nonlinearity with finite memory*. The exponential metric $\gamma_k(T_k)$ in (12) simplifies, by dropping the conditioning observations to

$$\gamma_k(T_k) = p\left(r_k|\boldsymbol{a}_{k-L}^k, \mu_{k-L}\right) P\{a_k\} \tag{16}$$

where L accounts for the channel dispersion and μ_{k-L} for the encoder/modulator memory. In this particular case, the finite-memory parameter C is equal to L. We remark that $\gamma_k(T_k)$ in (16) can be directly used both in a VA and an FB algorithm. A similar property holds for (8), which is of interest, for example, in the case of transmission of linear block codes over ISI channels. Hence, the SP algorithm can also be applied [21].

IV. Stochastic Channels

In the case of a channel characterized by parameters affected by stochastic uncertainty, the observations $\{r_k\}$ are *dependent*, so that the channel memory may not be finite. A very general parametric model for the observation r_k is the following:

$$r_k = g\left(\boldsymbol{a}_{k-L}^k, \mu_{k-L}, \boldsymbol{\xi}_0^k\right) + w_k \tag{17}$$

where L is an integer, $\boldsymbol{\xi}_0^k$ is a sequence of stochastic parameters independent from $\boldsymbol{a}$, and w_k is an additive noise sample.[7] Under this channel model, the following *conditional Markov property*

$$p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_0^k\right) = p\left(r_k|\boldsymbol{r}_{k-N}^{k-1}, \boldsymbol{a}_0^k\right) \tag{18}$$

where N is the order of Markovianity, is sufficient to guarantee a finite-memory condition. In fact, as shown in Appendix IV, (18) implies the following:

$$p\left(r_k|\boldsymbol{r}_0^{k-1}, \boldsymbol{a}_0^k\right) = p\left(r_k|\boldsymbol{r}_{k-N}^{k-1}, \boldsymbol{a}_{k-C}^k, \mu_{k-C}\right) \tag{19}$$

where the finite-memory parameter is $C = N + L$. It is immediate to recognize that (19) represents a special case of (4). As a consequence, all the derivations in the previous section hold by using the exponential metric $\gamma_k(T_k) = p(r_k|\boldsymbol{r}_{k-N}^{k-1}, T_k)P\{a_k\}$. In other words, (19) is the key relation which "links" the algorithms derived in Section III with the detection problem over channels with memory. A statistical description of the stochastic channel parameter allows one to compute this exponential metric as

$$\begin{aligned}\gamma_k(T_k) &= \frac{p\left(\boldsymbol{r}_{k-N}^k|T_k\right)}{p\left(\boldsymbol{r}_{k-N}^{k-1}|S_k\right)} P\{a_k\}\\ &= \frac{\mathrm{E}_{\boldsymbol{\xi}_0^k}\left\{p\left(\boldsymbol{r}_{k-N}^k|T_k, \boldsymbol{\xi}_0^k\right)\right\}}{\mathrm{E}_{\boldsymbol{\xi}_0^{k-1}}\left\{p\left(\boldsymbol{r}_{k-N}^{k-1}|S_k, \boldsymbol{\xi}_0^{k-1}\right)\right\}} P\{a_k\}. \end{aligned} \tag{20}$$

[7]Gaussianity of the additive noise is not required for the validity of the following derivation in this section.

The above exact result, although theoretically limited by the fact that *in realistic scenarios* the conditional Markov property (18) is seldom met *exactly*, suggests a reasonable approach to devise effective approximate detection algorithms whenever the conditional observations are asymptotically independent for increasing index difference [15], [22].

V. Examples of Applications

A. Noncoherent Channel

As a first example, we assume that the channel introduces an unknown phase rotation, modeled as a time-invariant random variable θ with uniform distribution in $[0, 2\pi)$. We consider coded linear modulations at the transmitter side. In this case, the samples at the output of a matched filter have the following expression:

$$r_k = c_k e^{j\theta} + w_k \tag{21}$$

where w_k is an additive white Gaussian noise (AWGN) sample of variance σ_w^2. The channel model (21) is a special case of (17) with $L = 0$ ($C = N$) and a dependence from a single time-invariant stochastic parameter. It is immediate to conclude that, being θ a random variable, the channel memory is infinite. Hence, the conditional Markov property can be claimed in an approximate sense only. On the basis of the considered phase model, γ_k can be expressed as

$$\begin{aligned}\gamma_k(T_k) &= \frac{\mathrm{E}_\theta\left\{p\left(\boldsymbol{r}_{k-C}^k|\theta, T_k\right)\right\}}{\mathrm{E}_\theta\left\{p\left(\boldsymbol{r}_{k-C}^{k-1}|\theta, S_k\right)\right\}} P\{a_k\}\\ &\sim \exp\left\{-\frac{|c_k|^2}{2\sigma_w^2}\right\} \frac{I_0\left(\frac{1}{\sigma_w^2}\left|\sum_{i=0}^{C} r_{k-i}c_{k-i}^*\right|\right)}{I_0\left(\frac{1}{\sigma_w^2}\left|\sum_{i=1}^{C} r_{k-i}c_{k-i}^*\right|\right)} P\{a_k\}\end{aligned}$$

where $I_0(x)$ is the zeroth order modified Bessel function of the first kind [23]. This result, obtained here as a special case of (20), is equivalent to previous solutions devised for noncoherent detection [24]–[27].

B. Flat Rayleigh-Fading Channel

As a second example, we consider transmission over a flat Rayleigh-fading channel ($L = 0$ and $C = N$). Assuming, for simplicity, that a sampling rate of one sample per information symbol is adequate, the observation sequence can be expressed as

$$r_k = f_k c_k + w_k$$

where $\{f_k\}$ is a sequence of realizations of zero mean Gaussian random variables with autocovariance sequence modeled according to isotropic scattering [28], i.e., given by $\mathrm{E}\{f_k f_{k-n}^*\} = J_0(2\pi Bn)$, where $J_0(\cdot)$ is the zeroth-order Bessel function [23] and B is the normalized Doppler rate. In this case, the conditional Markov property is an approximation as well, and the ex-

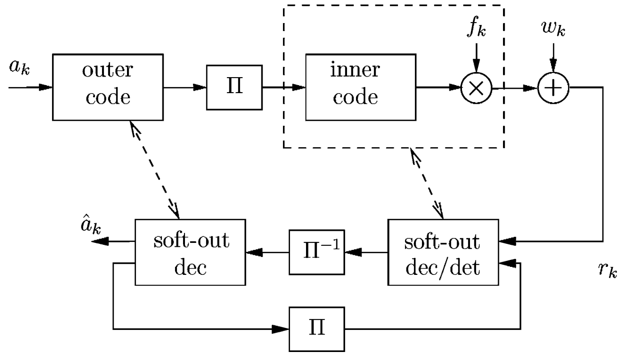

Fig. 3. System block diagram in the case of transmission of a SCCC over a fading channel with iterative detection at the receiver.

ponential metric γ_k can be computed, considering linear prediction, according to

$$\gamma_k(T_k) = \frac{1}{2\pi\sigma_k^2} \exp\left\{ -\frac{1}{\sigma_k^2} \left| r_k - \sum_{i=1}^{C} r_{k-i} \frac{c_k}{c_{k-i}} p_i \right|^2 \right\} P\{a_k\} \tag{22}$$

where the order of Markovianity C can be interpreted as the *prediction order*, $\{p_i\}_{i=1}^{C}$ are the prediction coefficients (which depend on state S_k, but not on symbol a_k), and σ_k^2 represents the mean square prediction error at epoch k. The result in (22), which can be derived from (19) owing to the Gaussianity of the observable, was obtained in [29]–[33] as a solution for maximum-likelihood sequence detection over fading channels. Related work for detection over Rayleigh/Ricean flat fading channels can be found in [34] and [35].

C. Numerical Results

We now evaluate the performance, by means of computer simulations, of a few iterative detection schemes using the finite-memory detection algorithms previously derived. Our results will show that the performance of the proposed schemes tends to that of the equivalent coherent schemes for increasing complexity. In particular, the performance is assessed in terms of bit-error rate (BER) versus the bit signal-to-noise ratio (SNR) E_b/N_0, E_b being the received energy per information bit and N_0 the one-sided noise power spectral density.

We first consider transmission of a serially concatenated convolutional code (SCCC) on a Rayleigh flat fading channel with normalized Doppler rate $B = 0.01$. The system block diagram is shown in Fig. 3. The code consists of an outer four-state, rate-1/2 convolutional code connected through a length-1024 pseudorandom interleaver to an inner four-state, rate-1/2 convolutional code. The generator polynomial matrices of outer and inner codes are given by, respectively

$$G_{\rm o}(D) = [1 + D + D^2 \quad 1 + D^2]$$
$$G_{\rm i}(D) = \left[1 \quad \frac{1 + D^2}{1 + D + D^2}\right].$$

The output symbols are mapped to a quaternary phase shift-keying (QPSK) constellation with Gray mapping. In Fig. 4, the performance of a receiver based on linear predictive detection at the inner decoder is shown, for various values of the prediction order C. For comparison, the ideal coherent performance,[8] i.e., assuming that the fading coefficients are perfectly known, is also shown. In all cases, five decoding iterations are considered. Increasing further the prediction order calls for the use of complexity reduction techniques at the inner detector/decoder [36].

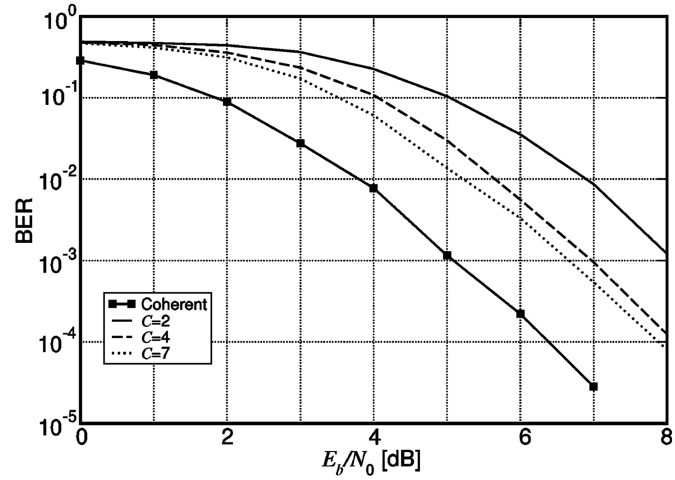

Fig. 4. Linear prediction-based iterative detection of a SCCC with QPSK in a fading channel.

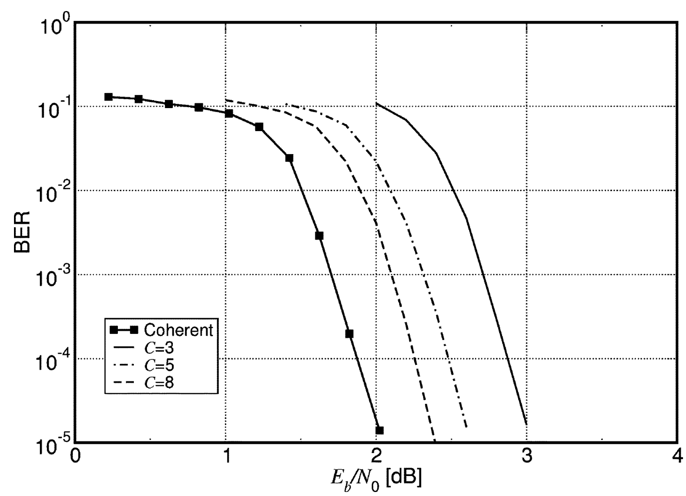

Fig. 5. Noncoherent iterative detection of a (3,6) regular LDPC code with BPSK.

In Fig. 5, the performance of the SP algorithm on the factor graph described in Fig. 2, for different values of C, is shown for a channel introducing an unknown phase rotation modeled as a time-invariant random variable uniformly distributed in $[0, 2\pi)$. The code is a (3,6) regular LDPC code with codewords of length 4000. A binary phase shift keying (BPSK) modulation format is used and a maximum of 200 iterations of the SP algorithm on the overall graph is allowed, using the *flooding* schedule [11]. A pilot symbol every 19 coded bits is added for ambiguity problems, and accounted for in the computation of the SNR—this makes the effective spectral efficiency equal to 0.487 bits/channel use. For increasing values of C, the performance approaches that of the corresponding coherent receiver.

[8]We remark that the ideal coherent receiver provides a lower bound to the performance of any practical receiver. Being the channel time-varying, this bound cannot be achieved with vanishing SNR loss by increasing C.

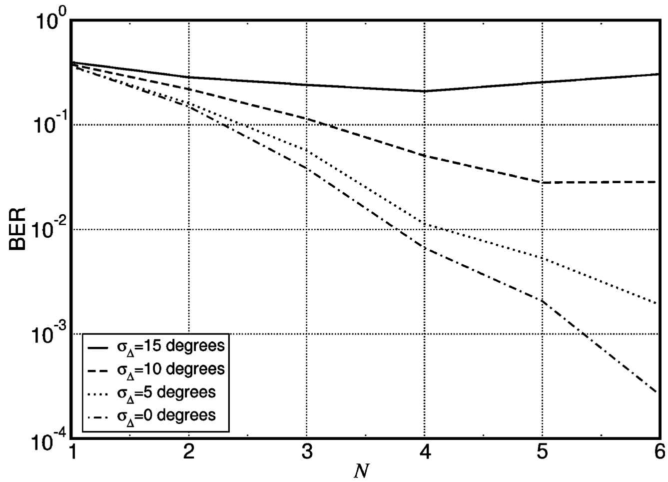

Fig. 6. Performance of noncoherent iterative detection of a PCCC with BPSK, as a function of finite-memory parameter $N = C$. Various channel dynamics (in terms of standard deviation σ_Δ of the Wiener phase model) are considered. In all cases, $E_b/N_0 = 2.5$ dB and five decoding iterations are considered.

In this case as well, the complexity can be reduced by applying techniques similar to reduced-state sequence detection [20], [21].

In order to investigate the impact of the finite-memory condition for communication over a channel introducing a phase rotation on the transmitted signal, we consider the case where the channel phase rotation is no longer modeled as a time-invariant random variable. In this case, the observation can be written as

$$r_k = c_k e^{j\theta_k} + w_k$$

where we assume that $\{\theta_k\}$ is modeled as a discrete-time Wiener process, with incremental variance over an observation interval equal to σ_Δ^2. In particular, we consider a parallel concatenated convolutional code (PCCC) of rate 1/2 with 16-state binary recursive systematic convolutional (RSC) component codes with generators (in octal notation) $G_1 = (37)_8$ and $G_2 = (21)_8$. The inner pseudorandom bit interleaver is 32 $\times$ 32. The output modulation is BPSK. The overall code is noncoherently noncatastrophic [37]. In Fig. 6, the BER is shown as a function of the finite-memory parameter C, for various values of the phase standard deviation σ_Δ. In all cases, the SNR is equal to 2.5 dB, and five decoding iterations are considered. As one can see, while for low values of the finite-memory parameter the phase dynamics have little influence on the performance of the detection algorithm (which depends, basically, on the SNR value), for larger values of the finite-memory parameter the phase dynamics significantly influence the performance. In particular, for each value of standard deviation σ_Δ there is an optimal value of the finite-memory parameter $C = N$—this is clearly visible in the cases with $\sigma_\Delta = 15^\circ$ and $\sigma_\Delta = 10^\circ$, whereas it is not visible in the other cases, since the optimal value is larger than six and the use of computer simulations becomes impractical. These results underline that the proposed detection strategy is "inherently" limited. In particular, the fact that the basic metric is obtained by averaging over the statistics of a random time-invariant phase rotation prevents the algorithm from tracking efficiently the channel dynamics.

VI. Concluding Remarks

In this paper, a general framework on finite-memory detection for channels with memory has been presented. We have shown that the same *basic metric* can be used in all considered hard-output and soft-output algorithms. Accounting for the channel memory, one can consider (combined) trellis-based detection and decoding for trellis codes or (separate) graph-based detection and decoding for linear block codes. A finite-memory detection algorithm designed in one case (either trellis-based or graph-based) can be *systematically extended* to the other case. We have applied the proposed approach to the case of transmission over *stochastic* channels. While the conditional Markov property is seldom met exactly in this scenario, it is possible to specialize the basic metric depending on the specific channel, and, consequently, to design trellis-based or graph-based detection algorithms.

Appendix I
Proof of (5)

We now show that (4) implies (5). By marginalization, we obtain

$$p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-D}^k, \mu_{k-D}\right) = \sum_{\boldsymbol{a}_0^{k-D-1}} p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_0^k, \mu_{k-D}\right) \cdot P\left\{\boldsymbol{a}_0^{k-D-1} | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-D}^k, \mu_{k-D}\right\}. \quad (23)$$

If a sequence $\boldsymbol{a}_0^{k-D-1}$, given μ_0, is incompatible with μ_{k-D}, then $P\{\boldsymbol{a}_0^{k-D-1} | \mu_{k-D}\} = 0$. Hence, for any sequence $\boldsymbol{a}_0^{k-D-1}$ compatible with μ_{k-D}

$$\begin{aligned} p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_0^k, \mu_{k-D}\right) &= p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_0^k\right) \\ &= p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-C}^k, \mu_{k-C}\right) \end{aligned}$$

where the last equality is based on the finite-memory condition (4). Hence, (23) becomes

$$\begin{aligned} &p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-D}^k, \mu_{k-D}\right) \\ &\quad= \sum_{\boldsymbol{a}_0^{k-D-1}} p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-C}^k, \mu_{k-C}\right) \\ &\qquad \cdot P\left\{\boldsymbol{a}_0^{k-D-1} | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-D}^k, \mu_{k-D}\right\} \\ &\quad= p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-C}^k, \mu_{k-C}\right) \\ &\qquad \cdot \underbrace{\sum_{\boldsymbol{a}_0^{k-D-1}} P\left\{\boldsymbol{a}_0^{k-D-1} | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-D}^k, \mu_{k-D}\right\}}_{1} \\ &\quad= p\left(r_k | \boldsymbol{r}_0^{k-1}, \boldsymbol{a}_{k-C}^k, \mu_{k-C}\right). \end{aligned}$$

Appendix II
Proof of (6)

We now show that (3) and (4) imply (6). By independence and causality assumptions, we obtain

$$p\left(\boldsymbol{r}_{k+1}^{K-1} | \boldsymbol{r}_0^k, \boldsymbol{a}_0^k\right) = \sum_{\boldsymbol{a}_{k+1}^{K-1}} p\left(\boldsymbol{r}_{k+1}^{K-1} | \boldsymbol{r}_0^k, \boldsymbol{a}_0^{K-1}\right) \cdot \underbrace{P\left\{\boldsymbol{a}_{k+1}^{K-1} | \boldsymbol{r}_0^k, \boldsymbol{a}_0^k\right\}}_{P\left\{\boldsymbol{a}_{k+1}^{K-1}\right\}}. \quad (24)$$

Applying the chain rule, it is immediate to write

$$p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_0^{K-1}\right)=\prod_{i=k+1}^{K-1}p\left(r_i|\boldsymbol{r}_0^{i-1},\boldsymbol{a}_0^{K-1}\right) \tag{25}$$

where based on the causality and finite-memory conditions, each term in (25) can be expressed as[9]

$$p\left(r_i|\boldsymbol{r}_0^{i-1},\boldsymbol{a}_0^{K-1}\right)=p\left(r_i|\boldsymbol{r}_0^{i-1},\boldsymbol{a}_{k-C+1}^{K-1},\mu_{k-C+1}\right). \tag{26}$$

Hence, substituting (26) in (25), one obtains

$$\begin{aligned}p&\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_0^{K-1}\right)\\&=\prod_{i=k+1}^{K-1}p\left(r_i|\boldsymbol{r}_0^{i-1},\boldsymbol{a}_{k-C+1}^{K-1},\mu_{k-C+1}\right)\\&=p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_{k-C+1}^{K-1},\mu_{k-C+1}\right).\end{aligned} \tag{27}$$

Finally, (24) becomes

$$\begin{aligned}p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_0^k\right)&=\sum_{\boldsymbol{a}_{k+1}^{K-1}}p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_{k-C+1}^{K-1},\mu_{k-C+1}\right)\\&\quad\cdot P\left\{\boldsymbol{a}_{k+1}^{K-1}\right\}\\&=p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_{k-C+1}^{k},\mu_{k-C+1}\right).\end{aligned}$$

Appendix III
Proof of (14) and (15)

Applying the Bayes and marginalization rules, it is possible to write

$$\begin{aligned}\alpha_k(S_k)&=p\left(\boldsymbol{r}_0^{k-1}|\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right)P\left\{\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right\}\\&=P\left\{\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}|\boldsymbol{r}_0^{k-1}\right\}p\left(\boldsymbol{r}_0^{k-1}\right)\\&=\sum_{\substack{\mu_{k-C-1},a_{k-C-1}:\\ \mathrm{ns}(\mu_{k-C-1},a_{k-C-1})=\mu_{k-C}}}P\left\{\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}|\boldsymbol{r}_0^{k-1}\right\}\\&\quad\cdot p\left(\boldsymbol{r}_0^{k-1}\right).\end{aligned} \tag{28}$$

Indicating concisely by $T_{k-1}:S_k$ the summation set in (28) and applying Bayes and chain factorization rules, (28) can be expressed as follows:

$$\begin{aligned}\alpha_k(S_k)&=\sum_{T_{k-1}:S_k}p\left(\boldsymbol{r}_0^{k-1}|\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right)\\&\quad\cdot P\left\{\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right\}\\&=\sum_{T_{k-1}:S_k}p\left(r_{k-1}|\boldsymbol{r}_0^{k-2},\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right)\\&\quad\cdot p\left(\boldsymbol{r}_0^{k-2}|\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right)\\&\quad\cdot P\left\{\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right\}.\end{aligned} \tag{29}$$

Owing to causality and independence of the information symbols, respectively, the following identities hold:

$$\begin{aligned}p\left(\boldsymbol{r}_0^{k-2}|\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right)&=p\left(\boldsymbol{r}_0^{k-2}|\boldsymbol{a}_{k-C-1}^{k-2},\mu_{k-C-1}\right)\\P\left\{\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right\}&=P\left\{\boldsymbol{a}_{k-C-1}^{k-2},\mu_{k-C-1}\right\}\\&\quad\cdot P\{a_{k-1}\}.\end{aligned}$$

[9]Note that, based on the causality condition, the conditional probability density function at the right-hand side of (26) can be written as $p(r_i|\boldsymbol{r}_0^{i-1},\boldsymbol{a}_{k-C+1}^{i},\mu_{k-C+1})$. However, leaving it as indicated is expedient for the application of chain factorization rule in (27).

Finally, (29) can be expressed as follows:

$$\begin{aligned}\alpha_k(S_k)&=\sum_{T_{k-1}:S_k}\underbrace{p\left(r_{k-1}|\boldsymbol{r}_0^{k-2},\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right)P\{a_{k-1}\}}_{\gamma_{k-1}(T_{k-1})}\\&\quad\cdot\underbrace{p\left(\boldsymbol{r}_0^{k-2}|\boldsymbol{a}_{k-C-1}^{k-1},\mu_{k-C-1}\right)P\left\{\boldsymbol{a}_{k-C-1}^{k-2},\mu_{k-C-1}\right\}}_{\alpha_{k-1}(S_{k-1})}\\&=\sum_{T_{k-1}:S_k}\gamma_{k-1}(T_{k-1})\alpha_{k-1}(S_{k-1})\end{aligned}$$

which corresponds to (14).

The backward recursion (15) can be similarly obtained. In fact, applying Bayes, marginalization, and chain factorization rules, it is possible to write

$$\begin{aligned}\beta_k(S_k)&=p\left(\boldsymbol{r}_k^{K-1}|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right)\\&=\sum_{a_k}p\left(\boldsymbol{r}_k^{K-1}|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)\\&\quad\cdot P\left\{a_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right\}\\&=\sum_{a_k}p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^{k},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)\\&\quad\cdot p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)\\&\quad\cdot P\left\{a_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right\}.\end{aligned} \tag{30}$$

Indicating concisely by $T_k:S_k$ the summation set in (30) and owing to (6) and independence of the information symbols, respectively, the following identities hold:

$$\begin{aligned}p&\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_{k-C}^k,\mu_{k-C}\right)\\&=p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_{k-C+1}^k,\mu_{k-C+1}=\mathrm{ns}(\mu_{k-C},a_{k-C})\right)\\P&\left\{a_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right\}\\&=P\{a_k\}.\end{aligned}$$

Finally, (30) can be written as

$$\begin{aligned}\beta_k(S_k)&=\sum_{a_k}\underbrace{p\left(\boldsymbol{r}_{k+1}^{K-1}|\boldsymbol{r}_0^k,\boldsymbol{a}_{k-C+1}^k,\mu_{k-C+1}\right)}_{\beta_{k+1}(S_{k+1})}\\&\quad\cdot\underbrace{p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_{k-C}^k,\mu_{k-C}\right)P\{a_k\}}_{\gamma_k(T_k)}\\&=\sum_{a_k}\beta_{k+1}(S_{k+1})\gamma_k(T_k)\end{aligned}$$

which corresponds to (15).

Appendix IV
Proof of (19)

Based on (18), one can write

$$p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_0^k\right)=p\left(r_k|\boldsymbol{r}_{k-N}^{k-1},\boldsymbol{a}_0^k\right)=\frac{p\left(\boldsymbol{r}_{k-N}^k|\boldsymbol{a}_0^k\right)}{p\left(\boldsymbol{r}_{k-N}^{k-1}|\boldsymbol{a}_0^k\right)}. \tag{31}$$

The conditional probability density function at the numerator of (31) can be expressed, by applying the total probability theorem, as follows:

$$p\left(\boldsymbol{r}_{k-N}^k|\boldsymbol{a}_0^k\right)=\underbrace{\int\cdots\int}_{\boldsymbol{\xi}_0^k}p\left(\boldsymbol{r}_{k-N}^k|\boldsymbol{a}_0^k,\boldsymbol{\xi}_0^k\right)p\left(\boldsymbol{\xi}_0^k|\boldsymbol{a}_0^k\right)\mathrm{d}\boldsymbol{\xi}_0^k. \tag{32}$$

Owing to the considered observation model (17), it is immediate to conclude that

$$p\left(\boldsymbol{r}_{k-N}^{k}|\boldsymbol{a}_{0}^{k},\boldsymbol{\xi}_{0}^{k}\right)=p\left(\boldsymbol{r}_{k-N}^{k}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C},\boldsymbol{\xi}_{0}^{k}\right)$$

where $C=N+L$. Being the stochastic parameters independent from the information symbols, the second probability density function inside the integral in (32) can be equivalently expressed as $p(\boldsymbol{\xi}_0^k|\boldsymbol{a}_{k-L-C}^k,\mu_{k-L-C})$. Finally, the integral (32) becomes

$$\begin{aligned}p\left(\boldsymbol{r}_{k-N}^{k}|\boldsymbol{a}_{0}^{k}\right)=&\underbrace{\int\cdots\int}_{\boldsymbol{\xi}_0^k}p\left(\boldsymbol{r}_{k-N}^{k}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C},\boldsymbol{\xi}_{0}^{k}\right)\\&\cdot p\left(\boldsymbol{\xi}_{0}^{k}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)\mathrm{d}\boldsymbol{\xi}_{0}^{k}\\=&\,p\left(\boldsymbol{r}_{k-N}^{k}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right).\end{aligned}$$

Applying the same line of reasoning to the denominator of (31) (taking also into account the causality of the system), one can conclude that

$$\begin{aligned}p\left(r_k|\boldsymbol{r}_0^{k-1},\boldsymbol{a}_0^k\right)=&\frac{p\left(\boldsymbol{r}_{k-N}^{k}|\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)}{p\left(\boldsymbol{r}_{k-N}^{k-1}|\boldsymbol{a}_{k-C}^{k-1},\mu_{k-C}\right)}\\=&\,p\left(r_k|\boldsymbol{r}_{k-L-C}^{k-1},\boldsymbol{a}_{k-C}^{k},\mu_{k-C}\right)\end{aligned}$$

which corresponds to (19).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, in particular Reviewer C, for their very detailed comments. They are also grateful to M. Franceschini, University of Parma, for useful discussions.

REFERENCES

[1] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT–13, no. 2, pp. 259–260, Apr. 1967.

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

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT–20, no. 2, pp. 284–287, Mar. 1974.

[4] J. H. Lodge, P. Hoeher, and J. Hagenauer, "The decoding of multidimensional codes using separable MAP "filters"," in *Proc. Queen's Univ. 16th Biennial Symp. Commun.*, Kingston, ON, Canada, May 1992, pp. 343–346.

[5] J. Lodge, R. Young, and J. H. P. Hoeher, "Separable MAP "filters" for the decoding of product and concatenated codes," in *Proc. IEEE Int, Conf. Commun. (ICC)*, Geneva, Switzerland, Jun. 1993, pp. 102–106.

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

[7] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[8] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[9] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEE Electron. Lett.*, vol. 33, pp. 457–458, Mar. 1997.

[10] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inf. Theory*, vol. IT–27, no. 5, pp. 533–547, Sep. 1981.

[11] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[12] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[13] J. Anderson, T. Aulin, and C.-E. Sundberg, *Digital Phase Modulation*. New York: Plenum, 1986.

[14] H. Meyr, M. Oerder, and A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3208–3214, Dec. 1994.

[15] K. M. Chugg, "The condition for the applicability of the Viterbi algorithm with implications for fading channel MLSD," *IEEE Trans. Commun.*, vol. 46, no. 9, pp. 1112–1116, Sep. 1998.

[16] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Feb. 1999.

[17] P. Robertson, E. Villebrun, and P. Hoeher, "Optimal and suboptimal maximum *a posteriori* algorithms suitable for turbo decoding," *Eur. Trans. Telecommun.*, vol. 8, no. 2, pp. 119–125, Mar./Apr. 1997.

[18] K. M. Chugg, A. Anastasopoulos, and X. Chen, *Iterative Detection: Adaptivity, Complexity Reduction, and Applications*. Norwell, MA: Kluwer, 2001.

[19] A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 843–849, Feb. 2001.

[20] G. Colavolpe, "LDPC codes over channels with memory," in *Proc. IEEE Int. Conf. Commun*, Paris, France, Jun. 2004, pp. 757–761.

[21] G. Colavolpe and G. Germi, "On the application of factor graphs and the sum-product algorithm to ISI channels," *IEEE Trans. Commun.*, vol. 53, no. 5, pp. 818–825, May 2005.

[22] G. Colavolpe and R. Raheli, "Noncoherent sequence detection," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1376–1385, Sep. 1999.

[23] I. S. Gradshtein and I. M. Ryzhik, *Table of Integrals, Series and Products*. New York: Academic, 1980.

[24] G. Colavolpe, G. Ferrari, and R. Raheli, "Noncoherent iterative (turbo) detection," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1488–1498, Sep. 2000.

[25] I. D. Marsland and P. T. Mathiopoulos, "On the performance of iterative noncoherent detection of coded M-PSK signals," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 588–596, Apr. 2000.

[26] G. Ferrari, G. Colavolpe, and R. Raheli, "Noncoherent iterative decoding of spectrally efficient coded modulations," *Ann. Telecommun.*, vol. 56, pp. 409–421, Jul./Aug. 2001.

[27] ——, *Turbo Codes: Error-Correcting Codes of Widening Application*, M. Jézéquel and R. Pyndiah, Eds. London, U.K.: Hermes Penton Science, 2002, pp. 59–81. ch. "Noncoherent iterative decoding of spectrally efficient coded modulationslt.

[28] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 957–1000, Aug. 1968.

[29] J. Lodge and M. Moher, "Maximum-likelihood estimation of CPM signals transmitted over Rayleigh flat fading channels," *IEEE Trans. Commun.*, vol. 38, no. 6, pp. 787–794, Jun. 1990.

[30] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1534–1544, Feb./Apr. 1995.

[31] G. M. Vitetta and D. P. Taylor, "Maximum-likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat fading channels," *IEEE Trans. Commun.*, vol. 43, no. 11, pp. 2750–2758, Nov. 1995.

[32] I. D. Marsland and P. T. Mathiopoulos, "Multiple differential detection of parallel concatenated convolutional (turbo) codes in correlated fast Rayleigh fading," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 265–275, Feb. 1998.

[33] P. Hoeher and J. Lodge, "'Turbo DPSK': Iterative differential PSK demodulation and channel decoding," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 837–843, Jun. 1999.

[34] D. Makrakis, P. T. Mathiopoulos, and D. Bouras, "Optimal detection of coded differentially encoded QAM and PSK signals with diversity reception in correlated fast fading channels," *IEEE Trans. Veh. Technol.*, vol. 42, no. 3, pp. 245–258, Aug. 1993.

[35] ——, "Optimal decoding of coded PSK and QAM signals in correlated fast fading channels and AWGN: A combined envelope, multiple differential and coherent detection approach," *IEEE Trans. Commun.*, vol. 42, no. 1, pp. 63–75, Jan. 1994.

[36] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR-type algorithms," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 848–859, May 2001.

[37] D. Raphaeli, "Noncoherent coded modulation," *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 172–183, Feb. 1996.

Gianluigi Ferrari (S'97–M'03) was born in Parma, Italy, in November 1974. He received the "Laurea" degree (five-year program) (*summa cum laude*) and the Ph.D. degree in electrical engineering from the University of Parma, Parma, Italy, in October 1998 and January 2002, respectively.

From July 2000 to December 2001, he was a Visiting Scholar at the Communication Sciences Institute, University of Southern California, Los Angeles. Since 2002, he has been a Research Professor at the Department of Information Engineering, University of Parma. Between November 2002 and December 2004, he visited, as a Research Associate, the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA. He has published more than 50 papers in leading international conferences and journals. He is coauthor of *Detection Algorithms for Wireless Communications, With Applications to Wired and Storage Systems* (New York: Wiley, 2004), *Ad Hoc Wireless Networks: A Communication-Theoretic Perspective* (New York: Wiley, 2006), and *Teoria Della Probabilità e Variabili Aleatorie Con Applicazioni* (New York: McGraw-Hill, 2005). He serves as a frequent reviewer for many international journals and conferences. His research interests include digital communication systems design, adaptive signal processing (with particular emphasis on iterative detection techniques for channels with memory), information theory, and ad hoc wireless networking.

Dr. Ferrari is listed in Marquis *Who's Who in the World*, 2005, and in *2000 Outstanding Intellectuals of the 21st Century*, International Biographical Centre, U.K., 2005. He serves as a Technical Program Member for several international conferences, among them the IEEE International Conference on Communications (ICC'05), Seoul, Korea, 2005, and the IEEE International Conference on Wireless Networks, Communications and Mobile Computing (WirelessCom 2005), Maui, HI, 2005.

Giulio Colavolpe (S'96–M'00) was born in Cosenza, Italy, in 1969. He received the Dr. Ing. degree in telecommunications engineering (*cum laude*) from the University of Pisa, Pisa, Italy, in 1994 and the Ph.D. degree in information technology from the University of Parma, Parma, Italy, in 1998.

Since 1997, he has been at the University of Parma, where he is now an Associate Professor of Telecommunications. In 2000, he was a Visiting Scientist at the Institut Eurécom, Valbonne, France. He is also coauthor of the book *Detection Algorithms for Wireless Communications, With Applications to Wired and Storage Systems* (New York: Wiley, 2004). His research activity has led to more than 60 scientific publications in leading international journals and conference proceedings and a few industrial patents. His main research interests include digital transmission theory, channel coding, and signal processing.

Riccardo Raheli (M'87) received the Dr. Ing. degree (Laurea) in electrical engineering (*summa cum laude*) from the University of Pisa, Italy, in 1983, the M.S. degree in electrical and computer engineering from the University of Massachusetts, Amherst, in 1986, and the Ph.D. degree (Perfezionamento) in electrical engineering (*summa cum laude*) from the Scuola Superiore di Studi Universitari e di Perfezionamento (now "S. Anna"), Pisa, Italy, in 1987.

From 1986 to 1988, he was with Siemens Telecomunicazioni, Cassina de' Pecchi, Milan, Italy. From 1988 to 1991, he was a Research Professor at the Scuola Superiore di Studi Universitari e di Perfezionamento S. Anna, Pisa, Italy. In 1990, he was a Visiting Assistant Professor at the University of Southern California, Los Angeles. Since 1991, he has been with the University of Parma, Italy, first as a Research Professor, then as an Associate Professor, and currently as a Professor of Communications Engineering. His scientific interests are in the general area of statistical communication theory, with application to wireless, wired and storage systems, and special attention to data detection in uncertain environments, iterative information processing and adaptive algorithms for communications. His research work has lead to numerous scientific publications in leading international journals and conference proceedings, as well as a few industrial patents. In 1990, he conceived (with A. Polydoros) the principle of *Per-Survivor Processing*. He is coauthor of the book *Detection Algorithms for Wireless Communications, with Applications to Wired and Storage systems* (New York: Wiley, 2004).

Dr. Raheli served on the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS as an Editor for Detection, Equalization, and Coding from 1999 to 2003. He is currently serving as a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special Issue on Differential and Noncoherent Wireless Communications), 2005. Since 2003, he has been on the Editorial Board of the *European Transactions on Telecommunications* as an Editor for Communication Theory.