On LDPC Codes over Channels with Memory

Giulio Colavolpe

***Abstract*— The problem of detection and decoding of low-density parity-check (LDPC) codes transmitted over channels with memory is addressed. A new general method to build a factor graph which takes into account both the code constraints and the channel behavior is proposed and the a posteriori probabilities of the information symbols, necessary to implement maximum a posteriori (MAP) *symbol* detection, are derived by using the sum-product algorithm. With respect to the case of a LDPC code transmitted on a memoryless channel, the derived factor graphs have additional factor nodes taking into account the channel behavior and not the code constraints. It is shown that the function associated to the generic factor node modeling the channel is related to the basic branch metric used in the Viterbi algorithm when MAP *sequence* detection is applied or in the BCJR algorithm implementing MAP *symbol* detection. This fact suggests that all the previously proposed solutions for those algorithms can be systematically extended to LDPC codes and graph-based detection.**

When the sum-product algorithm works on the derived factor graphs, the most demanding computation is in general that performed at factor nodes modeling the channel. In fact, the complexity of the computation at these factor nodes is in general *exponential* in a suitably defined channel memory parameter. In these cases, a technique for complexity reduction is illustrated. In some particular cases of practical relevance, the above mentioned complexity becomes *linear* in the channel memory. This does not happen in the same cases when detection is performed by using the Viterbi algorithm or the BCJR algorithm, suggesting that the use of factor graphs and the sum-product algorithm might be computationally more appealing. As an example of application of the described framework, the cases of noncoherent and flat fading channels are considered.

***Index Terms*— Factor graphs, Sum-product algorithm, Channels with memory, Phase-noise, Flat fading, Low-density parity-check codes, Iterative detection/decoding,**

I. INTRODUCTION

IN RECENT years, the extraordinary success of turbo codes has stimulated the rediscovery of another class of codes exhibiting similar performance and characteristics [1]. These codes, called *low-density parity-check* (LDPC) codes, were first introduced by Gallager [2] in their original *regular* version. The recently conceived *irregular* LDPC codes exhibit an impressive performance outperforming the best known turbo codes [3]–[5]. This paper proposes a new general framework, based on factor graphs (FGs) and the sum-product algorithm (SPA) [6], for designing decoding schemes for channel codes transmitted over channels with memory with particular emphasis to LDPC codes. Since the derived algorithms are iterative, they are particularly suited to turbo codes and LDPC codes whose decoding is iterative even when they are transmitted over a memoryless channel.

In the technical literature there are a lot of recent papers on detection and decoding of LDPC codes over channels with memory. In [7]–[12], the case of magnetic channels, where the memory is due to the presence of known intersymbol interference (ISI), is considered, and separate trellis-based detection, performed by means of the BCJR algorithm [13] as in turbo equalization [14]–[16], is proposed.

In the case of a noncoherent channel, i.e., a channel which introduces an unknown and possibly time-varying channel phase, the algorithms developed in [17]–[20] are designed for turbo codes and can be applied to LDPC codes only if trellis-based separate detection is performed. In [21] a channel model where the unknown carrier phase is constant over a block of N symbols and independent from block to block (the so-called block noncoherent channel) is considered, and detection algorithms for LDPC codes based on this model are developed. A non-Bayesian approach is adopted in [22], [23]. In [22] the channel phase is estimated by using the expectation-maximization (EM) algorithm, as originally proposed in [24]–[27] for turbo codes, and the estimation algorithm is embedded into the LDPC iterative decoding process. On the contrary, in [23] a class of problems is identified for which the optimal (in the sense of the generalized-likelihood ratio test) computation of the symbol a posteriori probabilities can be performed with polynomial complexity and the application to LDPC codes and the noncoherent channel is discussed.

For flat correlated Rayleigh fading channels, the algorithms in [28]–[30] can be applied to LDPC codes only performing trellis-based separate detection. On the contrary, for the block fading channel, an algorithm which can be directly applied to the case of LDPC codes is described in [31]. This algorithm is based on an estimation of the fading amplitude and a quantization of the channel phase for each fading block but similar approaches can be conceived for correlated fading channels, for example based on a channel Markov model (e.g., see [29]) and a Kalman smoothing approach [6], [32].

A more detailed discussion must be devoted to [33]. In that paper, a general framework to solve the problem of joint decoding and estimation, in the presence of unknown parameters, is described. The approach is Bayesian, i.e., the channel parameters are modeled as stochastic processes with known statistics and the use of FGs [6] that include both code constraints and channel statistics is advocated in a very general setting. The SPA [6] is then used to implement the maximum a posteriori (MAP) symbol detection strategy. Since the channel parameters, which are continuous random variables, are explicitly represented in the graph, the application of the SPA becomes impractical. To solve this problem, the

Paper submitted on April 20, 2004. Revised on January 28, 2005 and April 12, 2005.

G. Colavolpe is with Università di Parma, Dipartimento di Ingegneria dell'Informazione, Parco Area delle Scienze 181A, I-43100 Parma, Italy, e-mail: giulio@unipr.it.

This paper was presented in part at the IEEE International Conference on Communications (ICC'04), Paris, France, June 2004. This work is funded by the European Space Agency, ESA-ESTEC, Noordwijk, The Netherlands.

method of *canonical distributions* is adopted. By specializing the approach of [33] to particular channel phase statistics and canonical distributions, several algorithms for noncoherent detection of LDPC codes have been proposed [34]–[37]. For fading channels, it can be shown that the algorithm in [31] can be rederived using the framework in [33] and also the well-known Kalman smoother can be seen in this light [6], [32]. From the above mentioned papers, it is clear that the choice of the canonical distribution becomes crucial to determine the performance and the complexity of the derived algorithms.

In this paper, an alternative Bayesian framework is described. As in [33], our framework is based on FGs and the SPA [6] and, in fact, by means of a factorization of the joint a posteriori probability of the information symbols, we derive an FG representing both the code constraints and the channel model. In this FG, however, the channel parameters *are not* explicitly represented since they are a priori averaged out. The application of the SPA to this FG leads to an iterative scheme for joint detection and decoding. The above mentioned factorization is exact in the case of channels with *finite* memory, such as a channel with known ISI, and approximate for channels with *infinite* memory. This latter case includes a noncoherent channel and a flat correlated Rayleigh fading channel. For these channels, the factorization is approximate in the sense of a finite dependence assumption.

The proposed framework generates algorithms non-trellis-based. Although the logarithm of the function associated to the generic factor node is the same basic metric used in trellis-based algorithms (in the Viterbi algorithm when MAP *sequence* detection is applied or in the BCJR algorithm implementing MAP *symbol* detection), this must be considered a new result. In fact, the structure and the characteristics of the derived algorithms are very different. As an example, the algorithms developed by using our framework are well suited for a fully parallel implementation of the detector/decoder.

As shown in the numerical results, it is not possible to say, in general, which approach is better in terms of performance and/or complexity since for some channels the approach in [33] is favorable, whereas for other channels it is not. In the approach in [33], the choice of the used canonical distribution is a critical point and must be performed "ad-hoc" for the channel at hand [37], whereas the proposed framework is independent from subjective choices and in general gives good performance and affordable complexity, as it will be shown in the numerical results. Hence, it is interesting to have multiple frameworks since there is no a clear advantage of one framework on the other.

The case of a channel which introduces known ISI is not explicitly considered in this paper since analyzed in depth in [38]. On the contrary, the focus will be on LDPC codes transmitted on noncoherent or flat correlated Rayleigh fading channels. For these channels and equal energy signals, a modified version of the described FGs may be devised. The application of the SPA to these modified graphs has a complexity *linear* in a suitably defined channel memory parameter C, allowing a low-complexity receiver implementation for all practical values of C. Notice that this does not occur in the corresponding trellis-based algorithms in which the complexity is exponential in C. The possibility of an implementation of the receiver for any value of C is a key point since this parameter has a fundamental role in determining the receiver performance. In fact, the optimal value of C must be chosen as a function of the channel rate of change—the faster the channel, the lower the optimal value of C.

The remainder of this paper is organized as follows. Section II introduces the channel model. The basic concepts on FGs and the SPA are also reviewed. The proposed framework for approaching the problem of joint detection and decoding of LDPC codes is described in Section III and possible complexity reduction methods are described in Section IV. In Section V some examples of applications are detailed. Finally, in Section VI we present some numerical results and in Section VII we point out some concluding remarks.

II. System Model and Preliminaries

A. System model

In the considered transmission system, a sequence of M-ary code symbols $\mathbf{c} = \{c_k\}$ is transmitted from epoch 0 to epoch $K-1$. These code symbols are obtained from the encoding, by means of a code $\mathcal{C}$, of a sequence of information symbols $\mathbf{a} = \{a_k\}$. The encoding function mapping information sequences $\mathbf{a}$ into the codewords $\mathbf{c}$ will be denoted by $\eta_{\mathcal{C}}$. This function will also include pilot symbols inserted in the sequence $\mathbf{c}$ to avoid phase ambiguity problems. A sub-sequence of code symbols is denoted in vector notation as

$$\mathbf{c}_{k_1}^{k_2} = (c_{k_1}, c_{k_1+1}, \ldots, c_{k_2}),\ k_2 > k_1. \tag{1}$$

The sequence of code symbols is then modulated and transmitted over a channel which is modeled as a noiseless filter (possibly stochastic) rendered noisy by additive white Gaussian noise (AWGN) with one-sided power spectral density N_0.

At the receiver side, by means of a discretization process, the received signal $r(t)$ is converted into a discrete-time sequence $\mathbf{r} = \{r_k\}$ [39]. We assume that a single sample r_k is used for each code symbol, which is practically sufficient in many cases. In the case of oversampling, the extension is straightforward—the observation r_k must be considered as a vector whose dimensionality is given by the number of samples per code symbol. With a notation similar to that used for code symbols, the sub-sequence of observations $\{r_k\}_{k=k_1}^{k_2}$ is denoted by $\mathbf{r}_{k_1}^{k_2}$.

We also assume that the channel is *causal*, that is the observation sequence $\mathbf{r}_0^k$ up to epoch k depends on the code sequence up to epoch k only. This condition may be formulated in terms of the following statistical dependence

$$p(\mathbf{r}_0^k|\mathbf{c}) = p(\mathbf{r}_0^k|\mathbf{c}_0^k) \tag{2}$$

and is verified in all commonly used channel models. As an example, for the cases of a noncoherent channel, a flat or a frequency-selective fading channel, and a channel with known and time-invariant ISI, the condition is verified, by considering both linear or continuous phase modulations (CPM).

Two channel models will be considered in the following. For a noncoherent channel, which is characterized by an unknown stochastic and possibly time-varying phase θ_k, considering a

linear modulation at the transmitter side and assuming that one sample per code symbol is adequate (as in the absence of strong phase variations), if transmit and receive filters are such that there is absence of ISI, we have the following observation model:

$$r_k = c_k e^{j\theta_k} + w_k \tag{3}$$

where w_k is a discrete-time complex AWGN noise sample with each component of variance $\sigma^2 = N_0$. In Section V-A, for the phase noise process $\{\theta_k\}$, different statistics will be considered leading to different detection algorithms.

For a flat correlated Rayleigh fading channel, assuming for simplicity that a sampling rate of one sample per information symbol is adequate, the observation can be expressed as

$$r_k = f_k c_k + w_k \tag{4}$$

where $\{f_k\}$ is a sequence of zero-mean complex Gaussian random variables with autocorrelation sequence modeled according to isotropic scattering [40], i.e., given by $\mathrm{E}\{f_k f^*_{k-n}\} = J_0(2\pi f_D T n)$, where $J_0(\cdot)$ is the zero-th order Bessel function and $f_D T$ is the normalized Doppler rate.

B. Factor graphs and the sum-product algorithm.

We now review the basic concepts of FGs and the SPA by using the notation of the excellent tutorial paper [6]. Let $X = \{x_1, \ldots, x_N\}$ denote a set of variables and $f(X)$ a multivariate function. Let $\mathcal{B}_1, \ldots, \mathcal{B}_m$ denote subsets of X. We say that $f(X)$ admits a factorization with supports $\mathcal{B}_1, \ldots, \mathcal{B}_m$, if $f(X)$ can be written as the product of the functions $\{f_j : j = 1, \ldots, m\}$, where f_j has the variables in $\mathcal{B}_j$ as arguments. The FG representing the factorization $f = \prod_j f_j$ is a *bipartite* graph $\mathcal{G} = \{\mathcal{V}, \mathcal{F}, \mathcal{E}\}$, where nodes in $\mathcal{V}$ (variable nodes) are associated with the variables $x_i \in X$, nodes in $\mathcal{F}$ (factor nodes) are associated with the functions f_j, and there exists an edge $e \in \mathcal{E}$ joining x_i and f_j if and only if $x_i \in \mathcal{B}_j$ (i.e., if x_i is an argument of f_j).

Let $f(X)$ be a probability mass function (pmf). Then, if the FG corresponding to the factorization of f has no cycles,[1] the marginal pmfs can be computed exactly in a finite number of steps by the SPA [6]. The SPA is defined by the computation rules at variable and at factor nodes, and by a suitable node activation schedule. Denoting by $\mu_{x_i \to f_j}(x_i)$ a message sent from the variable node x_i to the factor node f_j, by $\mu_{f_j \to x_i}(x_i)$ a message in the opposite direction, and by $\mathcal{A}_i$ the set of functions f_j having x_i as argument, the message computations performed at variable and factor nodes are, respectively [6]

$$\mu_{x_i \to f_j}(x_i) = \prod_{h \in \mathcal{A}_i \setminus \{f_j\}} \mu_{h \to x_i}(x_i) \tag{5}$$

$$\mu_{f_j \to x_i}(x_i) = \sum_{\sim\{x_i\}} \left[f_j(\{y \in \mathcal{B}_j\}) \prod_{y \in \mathcal{B}_j \setminus \{x_i\}} \mu_{y \to f_j}(y) \right] \tag{6}$$

[1]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 smallest cycle is the *girth* of the graph.

where, following the notation of [6], we indicate by $\sum_{\sim\{x_i\}}$ the *summary* operator, i.e., a sum over all variables excluding x_i. If the FG contains cycles, convergence to the exact marginal pmfs is not guaranteed. Moreover, the SPA is inherently iterative. Nevertheless, for many relevant problems characterized by FGs with cycles, the SPA was found to provide very good results and therefore it represents a viable solution to the approximated marginalization of multivariate pmfs when exact calculation is not feasible because of complexity. Remarkably, the ubiquitous *belief propagation* iterative decoding algorithm used to decode LDPC codes [41], the turbo-decoding algorithm [6],[42],[43], turbo multiuser detection/decoding [44] and turbo equalization algorithms [14],[16] are instances of the SPA or approximations thereof.

A *message-passing schedule* in the SPA is the specification of the order in which messages are updated. In general, especially for graphs with cycles, the so-called *flooding schedule* is adopted [45]: in each iteration, all variable nodes and subsequently all factor nodes pass new messages to their neighbors. This schedule is well suited for a fully parallel implementation of the iterative detectors/decoders presented in the next section. Other schedules may be adopted, serial or mixed serial-parallel, according to the specific implementation requirements.

III. Detection Algorithm

The application of the SPA [6] to an FG representing the joint a posteriori pmf of the information sequence $\mathbf{a}$ conditioned on a given observation sequence $\mathbf{r}$, allows the exact or approximate computation of the marginal a posteriori pmfs $P(a_k|\mathbf{r})$ [6]. Therefore, this algorithm may be used to implement the MAP *symbol* detection algorithm. In fact, the optimal MAP *symbol* detection rule minimizing the average symbol-error probability is given by

$$\widehat{a}_k = \underset{a_k}{\operatorname{argmax}}\, P(a_k|\mathbf{r}). \tag{7}$$

From the definition of the encoding function $\eta_{\mathcal{C}}$ we obtain the factorization

$$\begin{aligned} P(\mathbf{a}|\mathbf{r}) &\sim P(\mathbf{a})p(\mathbf{r}|\mathbf{a}) \sim P(\mathbf{a})\chi[\mathbf{c} = \eta_{\mathcal{C}}(\mathbf{a})]p(\mathbf{r}|\mathbf{c}) \\ &\sim \chi[\mathbf{c} = \eta_{\mathcal{C}}(\mathbf{a})] \prod_{k=0}^{K-1} p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_0^k) \end{aligned} \tag{8}$$

where we have used the causality condition (2) and the fact that the output signal pdf $p(\mathbf{r})$ does not depend on $\mathbf{a}$, assumed that the information symbols are independent, uniformly and identically distributed (hence $P(\mathbf{a}) = \text{const.}$), and denoted by $\chi[\mathbf{c} = \eta_{\mathcal{C}}(\mathbf{a})]$ the code indicator function, equal to 1 if $\mathbf{c}$ is the codeword corresponding to $\mathbf{a}$ and to zero otherwise. In (8), the symbol $\sim$ indicates that two quantities are monotonically related with respect to the variable of interest (in this case, $\mathbf{a}$).

If the probability density function $p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_0^k)$, which appears in (8), satisfies the condition

$$p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_0^k) = p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_{k-C}^k) \tag{9}$$

where C is a suitable parameter, the channel has *finite memory* [46],[47]. For this reason, parameter C will be nicknamed

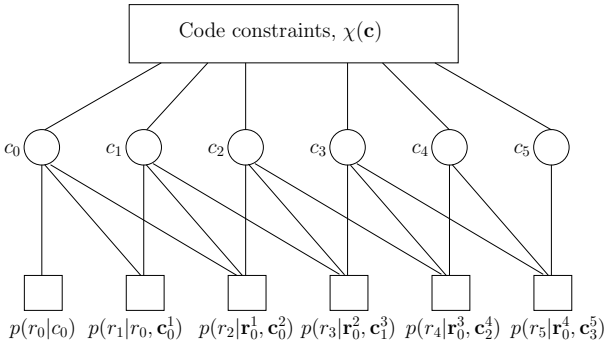


Fig. 1. Overall factor graph for $C = 2$.

finite memory parameter. Substituting (9) into (8), the joint a posteriori pmf of the information symbols $P(\mathbf{a}|\mathbf{r})$ may be expressed as

$$P(\mathbf{a}|\mathbf{r}) \sim \chi[\mathbf{c} = \eta_{\mathcal{C}}(\mathbf{a})] \prod_{k=0}^{K-1} p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_{k-C}^k). \qquad (10)$$

The corresponding FG is shown in Fig. 1 for $C = 2$, and represents both the code constraints (described by $\chi(\mathbf{c})$) and the channel behavior. In general, the code constraint function can be also represented by means of an FG. This graph is a portion of the overall FG and is connected with the graph representing the channel behavior. With respect to SPA-based decoding schemes for LDPC codes over a memoryless channel, *additional factor nodes* must be added at the bottom of the graph, as shown in Fig. 1. These additional factor nodes perform a marginalization, based on the channel model, without taking into account the code constraints. The complexity of this marginalization is, in general, exponential in C.

The finite memory condition (9) is exactly verified in the case of channels with known ISI. In fact, in this case it is

$$p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_0^k) = p(r_k|\mathbf{c}_{k-C}^k) \qquad (11)$$

where C is the length of the discrete-time channel impulse response. This case is not further considered in this paper since it is analyzed in depth in [38]. If the finite memory condition (9) is not verified in an exact sense as for a noncoherent or a fading channel (channels with *infinite* memory), an FG may be devised but the complexity of the message computation at the factor node $p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_0^k)$ modeling the channel grows exponentially with k and thus becomes impractical. For this reason, an approximation is introduced assuming that r_k depends on the R most recent observations and the most recent $C \geq R$ code symbols only. This *finite dependence assumption* may be expressed as

$$p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_0^k) \simeq p(r_k|\mathbf{r}_{k-R}^{k-1}, \mathbf{c}_{k-C}^k). \qquad (12)$$

This property, in general adopted in all practical detection schemes, is intuitive in the case of time-varying channels. In fact, in this case the conditional observations are asymptotically independent for increasing index difference. Also in this case, parameter C will be denoted to as memory parameter. The resulting (approximate) expression of $P(\mathbf{a}|\mathbf{r})$ becomes

$$P(\mathbf{a}|\mathbf{r}) \simeq \chi[\mathbf{c} = \eta_{\mathcal{C}}(\mathbf{a})] \prod_{k=0}^{K-1} p(r_k|\mathbf{r}_{k-R}^{k-1}, \mathbf{c}_{k-C}^k) \qquad (13)$$

and the corresponding FG is similar to that corresponding to (10) with $p(r_k|\mathbf{r}_0^{k-1}, \mathbf{c}_{k-C}^k)$ substituted by $p(r_k|\mathbf{r}_{k-R}^{k-1}, \mathbf{c}_{k-C}^k)$. The logarithm of the function $p(r_k|\mathbf{r}_{k-R}^{k-1}, \mathbf{c}_{k-C}^k)$ associated to the generic factor node modeling the channel is the same basic branch metric used in the Viterbi algorithm when MAP *sequence* detection is applied or in the BCJR algorithm implementing MAP *symbol* detection [47]. However, the structure and the characteristics of the derived algorithms are very different. As an example, as already mentioned the algorithms developed by using our framework are well suited for a fully parallel implementation of the detector/decoder. Other differences will be outlined in the next sections.

The quality of the convergence of the SPA to the exact marginal probabilities is in general determined by the girth of the graph. As an example, in designing LDPC codes, cycles of length 4 must be avoided to ensure decoding convergence. The FG derived from the proposed factorization has, in general, girth 4 (as one can see in Fig. 1). Surprisingly, we verified by computer simulations that these length-4 cycles involving two factor nodes which model the channel behavior often do not affect the convergence of the algorithm. If this is not the case, as for transmissions over ISI channels, FG transformations can be adopted [38].

IV. Complexity reduction

The complexity of the proposed algorithms may be reduced following a technique similar to that described in [48] for BCJR algorithms or to reduced-state sequence detection (RSSD) used for MAP sequence detection [49]–[51]. In fact, by choosing an integer $Q < C$, the updating rule (6) at factor nodes modeling the channel can be simplified as follows: the $C - Q$ symbols with highest reliability are hard-quantized based on the messages on the graph and the sum is performed over the Q symbols with lowest reliability.[2] In this way, the complexity becomes exponential in Q. With respect to the BCJR or the Viterbi algorithm, the complexity of these novel detection algorithms can be reduced more efficiently, due to the different structure of the algorithm. In fact, we are able to hard-quantize the symbols with highest reliability regardless of their position. On the contrary, when the BCJR or the Viterbi algorithm is used, we are constrained by the trellis structure to a truncation of the farthest symbols [48]–[51].

For equal energy signals, a modified version of the described FGs for noncoherent and flat fading channels may be devised. In fact, in the next Section we will show that for fading channels the function $p(\mathbf{r}|\mathbf{c})$ factors into the product of functions of two code symbols. For noncoherent channels this factorization is not exact but involves a simple approximation. The SPA on these modified graphs has a complexity *linear* in C, allowing a low-complexity receiver implementation for all practical values of C: in other words, no complexity reduction is needed.

[2] A partial representation by using set partitioning can be also adopted.

V. EXAMPLES OF APPLICATIONS

A. Noncoherent Channel

In this Section, we consider the application of the framework described in the previous Section to the case of a noncoherent channel. The system model is given by (3).

First algorithm — For the time being, we model the channel phase as a time-invariant random variable θ with uniform distribution in $[0, 2\pi)$. However, the finite dependence property (12) will lead to a detection algorithm that can be used for slowly-varying channels also. In this case, it is $R = C$ and the probability density function $p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k})$ which appears in (13) may be expressed, as

$$p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k}) = \frac{E_\theta\{p(\mathbf{r}_{k-C}^{k}|\mathbf{c}_{k-C}^{k}, \theta)\}}{E_\theta\{p(\mathbf{r}_{k-C}^{k-1}|\mathbf{c}_{k-C}^{k-1}, \theta)\}}. \quad (14)$$

The pdfs $p(\mathbf{r}_{k-C}^{k}|\mathbf{c}_{k-C}^{k}, \theta)$ and $p(\mathbf{r}_{k-C}^{k-1}|\mathbf{c}_{k-C}^{k-1}, \theta)$ which appear in (14) are now Gaussian pdfs. By averaging with respect to θ, after straightforward manipulations we obtain

$$p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k}) \sim \frac{\mathrm{I}_0\left(\frac{1}{\sigma^2}\left|\sum_{i=0}^{C} r_{k-i}c_{k-i}^{*}\right|\right)}{\mathrm{I}_0\left(\frac{1}{\sigma^2}\left|\sum_{i=1}^{C} r_{k-i}c_{k-i}^{*}\right|\right)} e^{-\frac{|c_k|^2}{2\sigma^2}} \quad (15)$$

where $\mathrm{I}_0(x)$ is the zeroth order modified Bessel function of the first kind. Unless the above mentioned technique for complexity reduction is adopted, at these factor nodes the SPA performs a marginalization whose computational burden grows exponentially with C.

The described algorithm is the graph-based version of the trellis-based algorithms in [52] and [19]. As previously mentioned, although the logarithm of the function associated to the generic factor node is the same basic metrics used in [52] and [19], the algorithm has different structure and characteristics.

Second algorithm — We now consider the case of a time-varying channel phase. A realistic model of phase noise is based on a discrete-time Wiener process $\theta_k = \theta_{k-1} + \Delta_k$, characterized by i.i.d. Gaussian increments Δ_k with mean zero and standard deviation σ_Δ, descriptive of the phase noise intensity. Hence[3]

$$\begin{aligned} &p(\theta_k|\theta_{k-1}, \theta_{k-2}, \ldots, \theta_0) = p(\theta_k|\theta_{k-1}) \\ &= \frac{1}{\sqrt{2\pi\sigma_\Delta^2}} e^{-\frac{(\theta_k - \theta_{k-1})^2}{2\sigma_\Delta^2}}, \; k = 1, \ldots, K-1 \quad (16) \\ &p(\theta_0) = \frac{1}{2\pi}, \; \theta_0 \in [0, 2\pi). \quad (17) \end{aligned}$$

Considering the channel model (3), the pdf $p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k})$ which appears in (13) may be expressed, as

$$p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k}) = \frac{E_{\boldsymbol{\theta}_{k-C}^{k}}\{p(\mathbf{r}_{k-C}^{k}|\mathbf{c}_{k-C}^{k}, \boldsymbol{\theta}_{k-C}^{k})\}}{E_{\boldsymbol{\theta}_{k-C}^{k-1}}\{p(\mathbf{r}_{k-C}^{k-1}|\mathbf{c}_{k-C}^{k-1}, \boldsymbol{\theta}_{k-C}^{k-1})\}} \quad (18)$$

having denoted by $\boldsymbol{\theta}_{k_1}^{k_2} = (\theta_{k_1}, \theta_{k_1+1}, \ldots, \theta_{k_2})$. In this case, however, an exact closed form expression of $E_{\boldsymbol{\theta}_{k-C}^{k}}\{p(\mathbf{r}_{k-C}^{k}|\mathbf{c}_{k-C}^{k}, \boldsymbol{\theta}_{k-C}^{k})\}$ which appears in (18), does not exist. However, a very good approximation can be found. By using the following approximate result [53]

$$\begin{aligned} &\frac{1}{\sqrt{2\pi\sigma_\Delta^2}} \int_0^{2\pi} e^{\mathrm{Re}[ze^{-jx}]} e^{-\frac{(x-y)^2}{2\sigma_\Delta^2}} dx \\ &\simeq \frac{\mathrm{I}_0(|z|)}{\mathrm{I}_0(\frac{z}{1+\sigma_\Delta^2|z|})} e^{\mathrm{Re}[\frac{|z|}{1+\sigma_\Delta^2|z|}e^{-jy}]} \end{aligned} \quad (19)$$

where z is a complex number and x and y are real numbers, it is possible to express

$$\begin{aligned} &E_{\boldsymbol{\theta}_{k-C}^{k}}\{p(\mathbf{r}_{k-C}^{k}|\mathbf{c}_{k-C}^{k}, \boldsymbol{\theta}_{k-C}^{k})\} \\ &\simeq \prod_{i=k-C}^{k} \mathrm{I}_0(|z_i|) e^{-\frac{|c_i|^2}{2\sigma^2}} \prod_{i=k-C+1}^{k} \frac{1}{\mathrm{I}_0(\frac{|z_i|}{1+\sigma_\Delta^2|z_i|})} \end{aligned} \quad (20)$$

where coefficients z_i can be recursively computed as

$$\begin{aligned} z_k &= \frac{r_k c_k^*}{\sigma^2} \\ z_{i-1} &= \frac{z_i}{1+\sigma_\Delta^2|z_i|} + \frac{r_{i-1}c_{i-1}^*}{\sigma^2} \\ &\quad i = k, k-1, \ldots, k-C+1. \end{aligned} \quad (21)$$

The term $E_{\boldsymbol{\theta}_{k-C}^{k-1}}\{p(\mathbf{r}_{k-C}^{k-1}|\mathbf{c}_{k-C}^{k-1}, \boldsymbol{\theta}_{k-C}^{k-1})\}$ which appears in (18) can be computed similarly. Even in this case, the computational burden grows exponentially with C and the complexity reduction technique described in the previous section becomes necessary.

Third algorithm — For a general time-varying phase process θ_k, assumed stationary, zero-mean and described by a given autocorrelation sequence of the phasor process $h_k = e^{j\theta_k}$, denoted by $R_h(n) = E\{e^{j\theta_{n+k}}e^{-j\theta_k}\}$, we can extend the approximate linear predictive approach described in [54] for Viterbi-based MAP *sequence* detection receivers.[4] In this case, the probability density function $p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k})$ may be approximated as (omitting irrelevant constant terms) [54]

$$\begin{aligned} &p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k}) \\ &\simeq \exp\left\{-\frac{1}{\sigma_e^2}\left|r_k - c_k \frac{\sum_{i=1}^{C} p_i \frac{r_{k-i}}{c_{k-i}}}{|\sum_{i=1}^{C} p_i \frac{r_{k-i}}{c_{k-i}}|}\right|^2\right\} \end{aligned} \quad (22)$$

where, in this case, C assumes the meaning of *prediction order*, $\{p_i\}_{i=1}^{C}$ are the prediction coefficients and σ_e^2 is the mean square prediction error. The prediction coefficients $\{p_i\}_{i=1}^{C}$ can be computed by solving a Wiener-Hopf linear system $\mathbf{R}\,\mathbf{p} = \mathbf{b}$, where $\mathbf{R}$ is a square $C \times C$ matrix whose elements have the following expression

$$[\mathbf{R}]_{\ell,m} = \begin{cases} R_h(|\ell - m|) & \text{if} \quad \ell \neq m \\ R_h(0) + \frac{2\sigma^2}{|c_{k-\ell}|^2} & \text{if} \quad \ell = m \end{cases} \quad (23)$$

[3]Note that, since the channel phase is defined modulus 2π, the pdf $p(\theta_k|\theta_{k-1})$ can be approximated as Gaussian only if $\sigma_\Delta \ll 2\pi$.

[4]For the above mentioned discrete-time Wiener process, it is easy to verify that $R_h(n) = e^{-|n|\sigma_\Delta^2/2}$.

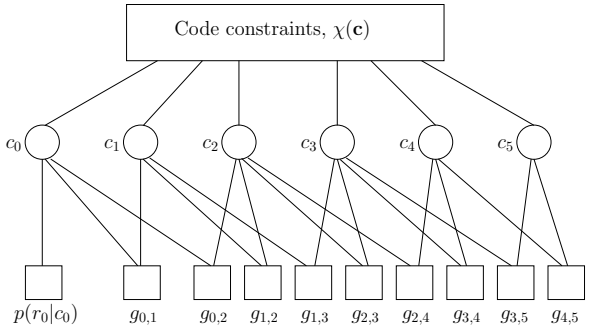


Fig. 2. Simplified overall factor graph for PSK signals and $C = 2$.

$\mathbf{p} \triangleq [p_1 \cdots p_C]^T$ is the unknown vector, and $\mathbf{b} = [R_h(1), R_h(2), \cdots, R_h(C)]^T$. The mean square prediction error may be expressed as [54]

$$\sigma_e^2 = R_h(0) + \frac{2\sigma^2}{|c_k|^2} - \sum_{i=1}^{C} p_i R_h(i). \tag{24}$$

For phase shift keying (PSK) signals, the prediction coefficients and the mean square prediction error become independent of the considered sequence. In addition, approximating $|\sum_{i=1}^{C} p_i \frac{r_{k-i}}{c_{k-i}}| \simeq |\sum_{i=1}^{C} p_i|$, and taking into account that $|c_k| = 1$, we have

$$\begin{aligned} &p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k}) \\ &\quad \sim \exp\left\{ \frac{2}{\sigma_e^2 |\sum_{i=1}^{C} p_i|} \mathrm{Re}\left[r_k c_k^* \sum_{i=1}^{C} p_i r_{k-i}^* c_{k-i} \right] \right\} \\ &\quad = \prod_{i=1}^{C} \exp\left\{ \frac{2\mathrm{Re}\left[p_i r_k c_k^* r_{k-i}^* c_{k-i} \right]}{\sigma_e^2 |\sum_{i=1}^{C} p_i|} \right\} \\ &\quad = \prod_{i=1}^{C} g_{k-i,k}(c_{k-i}, c_k) \end{aligned} \tag{25}$$

having defined

$$g_{k-i,k}(c_{k-i}, c_k) = \exp\left\{ \frac{2\mathrm{Re}\left[p_i r_k c_k^* r_{k-i}^* c_{k-i} \right]}{\sigma_e^2 |\sum_{i=1}^{C} p_i|} \right\}. \tag{26}$$

This further factorization has a direct impact on the graph structure. In fact, each factor node can be decomposed into C simpler degree-2 factor nodes. As an example, for $C = 2$, the corresponding FG is that in Fig. 2 (for brevity, the arguments of functions $g_{k-i,k}(c_{k-i}, c_k)$ are omitted). Hence, for increasing values of C, the number of factor nodes increases linearly but the computational burden at each factor node remains the same. In addition, in this modified FG there are no cycles of length 4 in the part of the graph modeling the channel. This approach can also be used when the phase is time-invariant. In this case, it is $p_i = 1/C$ and $\sigma_e^2 = 2\sigma^2$.

B. Flat Fading Channel

We now consider the case of a transmission over a flat Rayleigh fading channel. The system model is given by (4). In this case also, it is $R = C$. Omitting irrelevant constant terms, the probability density function $p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k})$ which appears in (13) may be expressed, based on linear prediction [28], [55]–[58], as[5]

$$\begin{aligned} &p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k}) \\ &\qquad \sim \exp\left\{ -\frac{1}{\sigma_e^2} \left| r_k - c_k \sum_{i=1}^{C} p_i \frac{r_{k-i}}{c_{k-i}} \right|^2 \right\}. \end{aligned} \tag{27}$$

As for the third algorithm in Section V-A, parameter C can be interpreted as the *prediction order*, $\{p_i\}_{i=1}^{C}$ are the prediction coefficients, and σ_e^2 represents the mean square prediction error. Coefficients $\{p_i\}$ and the mean square prediction error can be computed by solving a Wiener-Hopf linear system, and in general depend on the considered code sequence.

For PSK signals the prediction coefficients and the mean square prediction error become independent of the considered sequence. Taking into account that $|c_k| = 1$, after straightforward manipulations we have

$$\begin{aligned} &p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k}) \\ &\quad \sim \exp\left\{ \frac{2}{\sigma_e^2} \sum_{i=1}^{C} \mathrm{Re}[p_i r_k r_{k-i}^* c_k^* c_{k-i}] \right\} \\ &\quad \cdot \exp\left\{ -\frac{2}{\sigma_e^2} \sum_{i=1}^{C} \sum_{\ell=i+1}^{C} \mathrm{Re}[p_i p_\ell r_{k-i} r_{k-\ell}^* c_{k-i}^* c_{k-\ell}] \right\}. \end{aligned} \tag{28}$$

Substituting (28) into (13), it can be easily shown, by grouping the exponential terms which depend on the same argument $r_{k-i} r_{k-\ell}^* c_{k-i}^* c_{k-\ell}$, that the resulting (approximated for the finite dependence assumption only) joint a posteriori pmf of the information symbols becomes

$$P(\mathbf{a}|\mathbf{r}) \simeq \chi[\mathbf{c} = \eta_{\mathcal{C}}(\mathbf{a})] \prod_{k=0}^{K-1} \prod_{i=1}^{C} e^{\frac{2}{\sigma_e^2}\mathrm{Re}[q_i r_k r_{k-i}^* c_k^* c_{k-i}]} \tag{29}$$

where $q_i = p_i - \sum_{\ell=1}^{C-i} p_\ell p_{\ell+i}$. The corresponding FG is similar to that depicted in Fig. 2 and hence the complexity is *linear* in C. This is a fundamental difference with respect to trellis-based linear predictive receivers [28], [55]–[58], whose complexity is *exponential* in the prediction order, and suggests that by using the tool represented by FGs and SPA new computationally efficient algorithms can be derived. In addition, as already mentioned, the derived algorithms are also well suited for a fully parallel implementation of the detector/decoder.

VI. Numerical Results and Comparison with Existing Algorithms

In this Section, the performance of the proposed detection schemes is assessed by computer simulations in terms of bit error rate (BER) versus E_b/N_0, E_b being the received signal energy per information bit. The considered code is a (3,6)-regular LDPC code with codewords of length 4000. The binary PSK (BPSK) modulation is used and a maximum of 200 iterations of the SPA on the overall graph, by using the

[5]The case of a different flat fading model such as, for example, the Nakagami fading [59], can be easily managed by simply changing the expression of $p(r_k|\mathbf{r}_{k-C}^{k-1}, \mathbf{c}_{k-C}^{k})$.

flooding schedule, is allowed. A pilot symbol every 19 code bits is added to avoid ambiguity problems. This corresponds to a decrease in the effective transmission rate of a factor 19/20, resulting in an increase in the required signal-to-noise ratio of 0.223 dB which has been introduced artificially in the curves labeled "perfect channel state information" (CSI) for the sake of comparison. Hence, the gap between the "perfect CSI" curves, which refer to the case of perfect knowledge of the channel phase/fading, and the others is uniquely due to the need for phase/fading estimation/compensation, and not to the rate decrease due to pilot symbols insertion.

A. Noncoherent Channel

In Fig. 3 and Fig. 4, we show the performance of the first and third detection algorithm of Section V-A for different values of C in the case of a noncoherent time-invariant channel. In Fig. 3, the first algorithm is considered and compared with the ideal coherent receiver. To demonstrate that, in this case, the presence of cycles of length 4 does not affect the convergence of the algorithm, we also show, for $C = 3$ and $C = 5$, the performance obtained when the detector is implemented by using the BCJR-type algorithm in [19]. In this case a BCJR step is run, by using the classical forward-backward schedule, every LDPC decoder iteration. It can be easily shown that this algorithm can be derived starting from the proposed graph-based algorithm, introducing hidden variables representing $\mathbf{c}_{k-C}^{k-1}$, and changing the adopted schedule [6]. It can be observed that, although the proposed algorithm employs a fully parallel schedule and operates on a graph with girth 4, the performance loss is negligible. The complexity reduction is also considered to increase the phase memory C without an increase in complexity. As intuitively expected, the performance of the ideal coherent receiver is approached with limited complexity. In fact, a value of $Q = 1$ is in practice sufficient to attain the performance of the full-complexity receiver. For the third algorithm of Section V-A, which for BPSK has a complexity linear in C, similar considerations do not hold. It can be observed that larger values of C are required with respect to the first algorithm to obtain a given performance and, in addition, from Fig. 4 it seems that this algorithm is not able to reach the optimal "perfect CSI" performance, due to the approximations introduced to obtain an algorithm with complexity linear in C (see Section V-A).

We now consider (see Fig. 5) a time-varying noncoherent channel. The phase noise is modeled as a discrete-time Wiener process described by

$$\theta_k = \theta_{k-1} + \Delta_k \tag{30}$$

where $\{\Delta_k\}$ is a white real Gaussian process with zero mean and variance σ_Δ^2. The performance of the algorithms described in Section V-A, with the corresponding values of C optimized by computer simulations, is shown for $\sigma_\Delta = 6$, 12, and 16 degrees. All the proposed detection algorithms are very robust, especially those designed taking into account the channel statistics. A wide range of possibilities are available to the designer to choose the desired trade-off between performance and complexity.

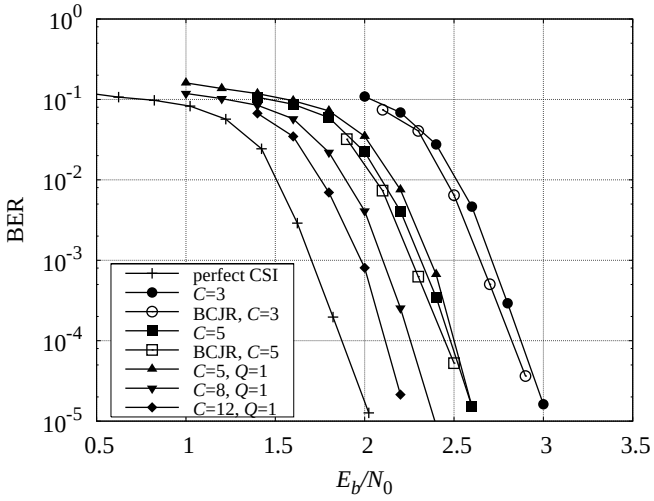


Fig. 3. Performance of the first algorithm in Section V-A.

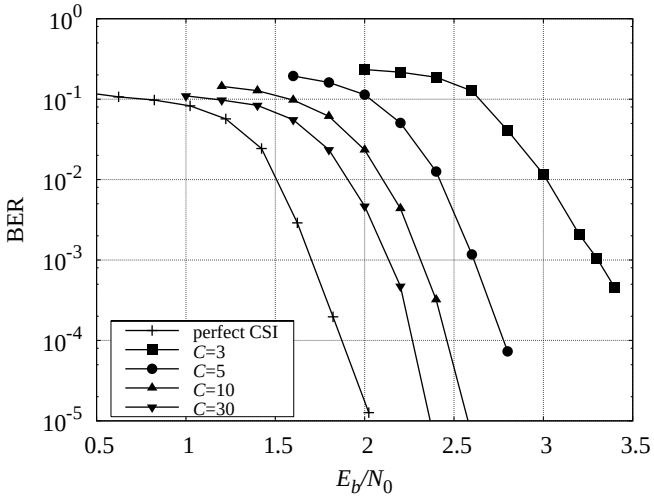


Fig. 4. Performance of the third algorithm in Section V-A.

For a Wiener phase model with $\sigma_\Delta = 6$ degrees, in Fig. 6 we compare the performance of our best algorithm (in this case, the algorithm based on linear prediction) with four other algorithms described in the technical literature. The first and the second algorithms are based on the canonical distribution approach in [33]. Namely, in the first case we consider a discretization of the parameter space which corresponds to letting the canonical distribution be a weighted sum of impulses. This approach has been adopted for Viterbi- and BCJR-like receivers in [60] and [17], respectively. The channel phase θ_k is assumed to take on the following L values: $\Theta = \{0, 2\pi/L, \ldots, 2\pi(L-1)/L\}$. Obviously, the corresponding receiver becomes "optimal" (in the sense that it approaches the performance of the exact SPA) for a sufficiently large number of discretization levels, at the expenses of an increasing computational complexity. We found that $L = 16$ values are sufficient to have no performance loss. The second algorithm is based on a Tikhonov canonical distribution [36], [37]. This algorithm is computationally very efficient since a couple of complex parameters only must

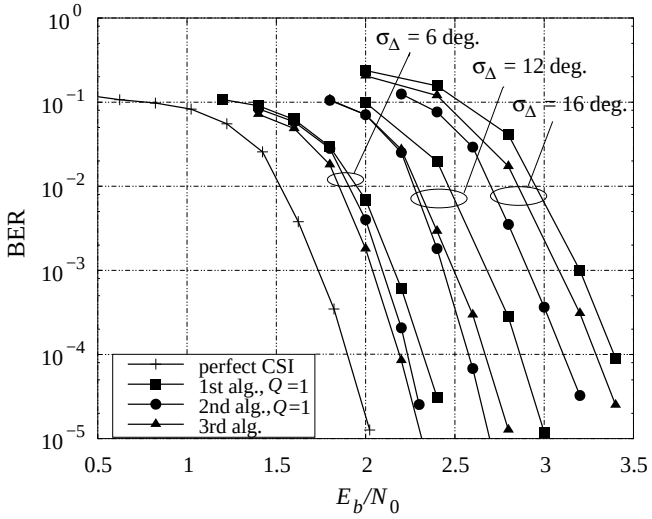


Fig. 5. Performance in the case of a time-varying channel phase. The value of C is optimized in each individual case.

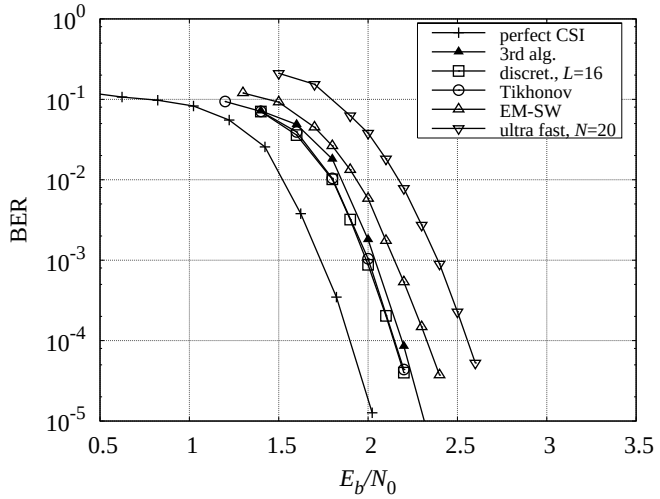


Fig. 6. Performance in the case of a time-varying channel phase. Comparison with other schemes.

be updated in a forward-backward fashion [36], [37]. The third algorithm is the "ultra fast" algorithm with overlapped windows described in [23], with the value of N optimized by computer simulation. Finally, the fourth one is based on the EM algorithm [22], [24]–[27]. In order to adapt this algorithm to a time-varying channel phase, different phase estimates are computed for each code symbol, taking into account the contribution of the adjacent symbols belonging to a window whose dimension is optimized by computer simulation. For this reason the algorithm is denoted by EM with sliding window (EM-SW). We found that the optimal window has width of 60 symbols for the considered phase noise.

In the third and fourth case, the performance loss is due to the fact that these two algorithms are designed for a different phase model, i.e., a block-constant phase. The more efficient algorithm, both in terms of performance and complexity, is that based on the Tikhonov canonical distribution. This algorithm is however specifically tailored for the Wiener phase model and cannot be extended to variables $\{\theta_k\}$ which do not form a Markov chain. In addition it is not suited for a fully parallel detector/decoder implementation since a forward-backward schedule is necessary to assure a good performance. On the contrary, the proposed third algorithm, which exhibits an almost optimal performance (and the slope of the curve is the same of the practically optimal approach based on discretization) with limited complexity, is expected to work well also for different phase noise models, since it uses the information of the phase autocorrelation function.

B. Flat Fading Channel

Finally, we consider the performance of the algorithm described in Section V-B for a flat correlated Rayleigh fading channel with normalized Doppler rate $f_DT = 10^{-2}$. The performance in the case of perfect CSI is also considered for comparison. Obviously, for increasing values of the memory parameter C there is a performance improvement. A similar behavior is observed for different values of the normalized Doppler rate. In addition, due to the linear complexity of the detection algorithm, it is possible to implement receivers with values of C higher than those used in [28], [55]–[58] for Viterbi- or BCJR-based algorithms, thus closely approaching the performance of the receiver with perfect CSI.

The flat fading channel is a case in which our framework is favorable with respect to the framework in [33]. In fact, the explicit representation of the channel parameters into the FG and a fading model autoregressive of order N (AR(N)) leads to an algorithm for joint decoding and estimation in which the fading estimate is obtained through a Kalman smoother [6], [32]. The value of $N = 1$ is the only one considered in the literature on graph-based algorithms [32], yielding the so-called *scalar Kalman smoother* which, however, is not sufficient to attain a good performance, as shown in Fig. 7. Values of $N \simeq C \geq 5$ should be adopted to obtain a good performance comparable with that of the proposed algorithm. In this case, however, the number of parameters to be recursively updated in the forward and backward recursions [32] is $N + N(N+1)/2$ and therefore the complexity is *quadratic* in N [61]. Hence, for a value of $N \simeq C$, the proposed algorithm is less complex since its complexity is *linear* in C.[6] In addition, a fully parallel schedule, such that adopted by the proposed algorithm, would havily degrade the performance of the Kalman smoother.

VII. CONCLUSIONS

In this paper, the problem of joint detection and decoding of LDPC codes transmitted over channels with memory has been considered. A general method to build a factor graph, taking into account both the code constraints and the channel behavior has been described. The application of the sum-product algorithm to this graph allows the (approximate) computation

[6]From a theoretical point of view, at least in the absence of noise, the inverse of the innovation filter [62] generating an AR(N) process is a finite impulse response filter whose coefficients are the prediction coefficients. Hence the prediction order is $C = N$. In case of isotropic scattering, the fading is not exactly AR, and in the considered case noise is present.

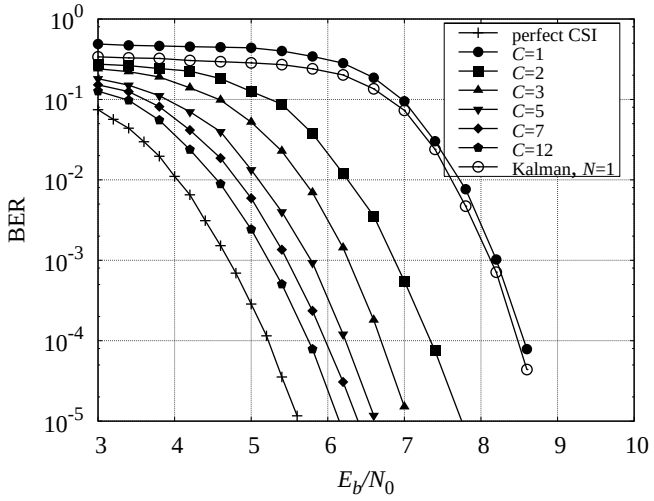


Fig. 7. Performance in the case of a flat correlated Rayleigh fading channel with $f_D T = 10^{-2}$.

of the marginal a posteriori probabilities of the information symbols, thus implementing maximum a posteriori *symbol* detection.

The function associated to factor nodes modeling the channel is shown to be the same basic branch metric used in the Viterbi algorithm when maximum a posteriori *sequence* detection is applied or in the BCJR algorithm implementing maximum a posteriori *symbol* detection. This fact suggests that all the previously proposed solutions for those algorithms simply extend to LDPC codes and graphical models. We demonstrated that this approach leads to algorithms with good performance and acceptable complexity. In addition, in some significant cases, the complexity of the graph-based derived algorithms is linear and not exponential in the channel memory, as in the trellis-based algorithms occurs, thus allowing decoding even for large values of the channel memory parameter.

The proposed framework is alternative to that proposed in [33]. It cannot be concluded that one framework is better than the other since this evaluation must be performed for the particular considered channel model. Nevertheless, the computational advantages and the good performance of the proposed framework, along with the possibility of a fully parallel implementation represent some important aspects of clear interest in practical applications.

REFERENCES

[1] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399–431, Feb. 1999.

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

[3] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann, "Practical loss-resilient codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 569–584, February 2001.

[4] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[5] S.-Y. Chung, G. D. Forney, T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Letters*, vol. 5, pp. 58–60, February 2001.

[6] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[7] T. V. Souvignier, M. Öberg, P. H. Siegel, R. E. Swanson, and J. K. Wolf, "Turbo decoding for partial-response channels," *IEEE Trans. on Commun.*, vol. 48, pp. 1297–1308, Aug. 2000.

[8] T. Mittelholzer, A. Dholakia, and E. Eleftheriou, "Reduced-complexity decoding of low-density parity check codes for generalized partial response channels," *IEEE Trans. on Magnetics*, vol. 37, pp. 721–728, March 2001.

[9] B. M. Kurkoski, P. H. Siegel, and J. K. Wolf, "Joint message-passing decoding of LDPC codes and partial-response channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1410–1422, June 2002.

[10] A. Thangaraj and S. W. McLaughlin, "Thresholds and scheduling for LDPC-coded partial response channels," *IEEE Trans. on Magnetics*, vol. 38, pp. 2307–2309, Sept. 2002.

[11] N. Varnica and A. Kavćić, "Optimized low-density parity-check codes for partial response channels," *IEEE Commun. Letters*, vol. 7, pp. 168–170, Apr. 2003.

[12] A. Kavćić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1636–1652, July 2003.

[13] 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. 20, pp. 284–287, Mar. 1974.

[14] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *European Trans. Telecommun.*, vol. 6, pp. 507–511, September/October 1995.

[15] A. Picart, P. Didier, and A. Glavieux, "Turbo-detection: A new approach to combat channel frequency selectivity," in *Proc. IEEE Intern. Conf. Commun.*, (Montreal, Canada), June 1997.

[16] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. on Commun.*, vol. 55, pp. 754–767, May 2002.

[17] M. Peleg, S. Shamai (Shitz), and S. Galán, "Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channel," *IEE Proc. Commun.*, vol. 147, pp. 87–95, Apr. 2000.

[18] I. D. Marsland and P. T. Mathiopoulos, "On the performance of iterative noncoherent detection of coded M-PSK signals," *IEEE Trans. on Commun.*, vol. 48, pp. 588–596, Apr. 2000.

[19] G. Colavolpe, G. Ferrari, and R. Raheli, "Noncoherent iterative (turbo) detection," *IEEE Trans. on Commun.*, vol. 48, pp. 1488–1498, Sept. 2000.

[20] A. Anastasopoulos and K. M. Chugg, "Adaptive iterative detection for phase tracking in turbo coded systems," *IEEE Trans. on Commun.*, vol. 49, Dec. 2001.

[21] R. Nuriyev and A. Anastasopoulos, "Pilot-symbol-assisted coded transmission over the block-noncoherent AWGN channel," *IEEE Trans. on Commun.*, vol. 51, pp. 953–963, June 2003.

[22] H. Steendam, N. Noels, and M. Moeneclaey, "Iterative carrier phase synchronization for low-density parity-check coded systems," in *Proc. IEEE Intern. Conf. Commun.*, (Anchorage, AK, U.S.A.), pp. 3120–3124, June 2003.

[23] I. Motedayen-Aval and A. Anastasopoulos, "Polynomial-complexity noncoherent symbol-by-symbol detection with application to adaptive iterative decoding of turbo-like codes," *IEEE Trans. on Commun.*, vol. 51, pp. 197–207, Feb. 2003.

[24] V. Lottici and M. Luise, "Carrier phase recovery for turbo-coded linear modulations," in *Proc. IEEE Intern. Conf. Commun.*, (New York, NY, USA), pp. 1541–1545, Apr. 2002.

[25] V. Lottici and M. Luise, "Embedding carrier phase recovery into iterative decoding of turbo-coded linear modulations," *IEEE Trans. on Commun.*, vol. 52, pp. 661–669, Apr. 2004.

[26] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "Turbo synchronization: an EM algorithm interpretation," in *Proc. IEEE Intern. Conf. Commun.*, (Anchorage, AK, U.S.A.), pp. 2933–2937, June 2003.

[27] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "A theoretical framework for soft information based synchronization in iterative (turbo) receivers." submitted to IEEE Trans. on Commun., 2003.

[28] P. Hoeher and J. Lodge, ""Turbo DPSK": Iterative differential PSK demodulation and channel decoding," *IEEE Trans. on Commun.*, vol. 47, pp. 837–843, June 1999.

[29] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1706–1717, Sept. 2001.
[30] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1697–1705, Sept. 2001.
[31] R.-R. Chen, R. Koetter, U. Madhow, and D. Agrawal, "Joint noncoherent demodulation and decoding for the block fading channel: a practical framework for approaching Shannon capacity," *IEEE Trans. on Commun.*, vol. 51, pp. 1676–1689, October 2003.
[32] H. Niu, M. Shen, J. A. Ritcey, and H. Liu, "Iterative channel estimation and LDPC decoding over flat-fading channels: a factor graph approach," in *Proc. Conf. on Inform. Sciences and Systems*, (Johns Hopkins University), Mar. 2003.
[33] A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 843–849, Feb. 2001.
[34] J. Dauwels and H.-A. Loeliger, "Joint decoding and phase estimation: an exercise in factor graphs," in *Proc. IEEE Symposium on Information Theory and its Applications*, (Yokohama, Japan), p. 231, July 2003.
[35] J. Dauwels and H.-A. Loeliger, "Phase estimation by message passing," in *Proc. IEEE Intern. Conf. Commun.*, (Paris, France), pp. 523–527, June 2004.
[36] G. Colavolpe, A. Barbieri, G. Caire, and N. Bonneau, "Bayesian and non-Bayesian methods for iterative joint decoding and detection in the presence of phase noise," in *Proc. IEEE Symposium on Information Theory and its Applications*, (Chichago, IL, USA), p. 131, June-July 2004.
[37] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise." IEEE J. Selected Areas in Communications, to appear, 2005. Available at *http://www.tlc.unipr.it/people/colavolpe/public.html*.
[38] G. Colavolpe and G. Germi, "On the application of factor graphs and the sum-product algorithm to ISI channels," *IEEE Trans. on Commun.*, vol. 53, May 2005.
[39] H. Meyr, M. Oerder, and A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers," *IEEE Trans. on Commun.*, vol. 42, pp. 3208–3214, Dec. 1994.
[40] R. Clark, "A statistical theory of mobile radio reception," *Bell System Tech. J.*, vol. 47, pp. 957–1000, Aug. 1968.
[41] T. Richardson and R. Urbanke, "The capacity of low density parity check codes under message passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.
[42] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. on Commun.*, vol. 44, pp. 1261–1271, October 1996.
[43] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inform. Theory*, vol. 46, pp. 325–343, Mar. 2000.
[44] J. Boutros and G. Caire, "Iterative multiuser joint decoding: unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1772–1793, July 2002.
[45] F. Kschischang and B. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE J. Select. Areas Commun.*, pp. 219–231, February 1998.
[46] G. Ferrari, G. Colavolpe, and R. Raheli, "On trellis-based truncated-memory detection," in *Proc. IEEE Global Telecommun. Conf.*, (San Francisco, CA, U.S.A.), pp. 2218–2222, Nov. 2003.
[47] G. Ferrari, G. Colavolpe, and R. Raheli, "Finite memory: optimality and reality," in *Proc. Intern. Symp. on Turbo Codes & Relat. Topics*, (Brest, France), pp. 1–8, Sept. 2003.
[48] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR-type algorithms," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 848–859, May 2001.
[49] M. V. Eyuboğlu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. on Commun.*, vol. 38, pp. 13–20, Jan. 1988.
[50] A. Duel-Hallen and C. Heegard, "Delayed decision feedback estimation," *IEEE Trans. on Commun.*, vol. 37, pp. 428–436, May 1989.
[51] P. R. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. on Commun.*, vol. 36, pp. 669–676, July 1989.
[52] G. Colavolpe and R. Raheli, "Noncoherent sequence detection," *IEEE Trans. on Commun.*, vol. 47, pp. 1376–1385, Sept. 1999.
[53] A. Barbieri, G. Colavolpe, and G. Caire, "Joint iterative detection and decoding in the presence of phase noise and frequency offset," in *Proc. IEEE Intern. Conf. Commun.*, May 2005.
[54] G. Ferrari, G. Colavolpe, and R. Raheli, "Linear predictive receivers for phase uncertain channels," in *Intern. Symp. on Signal Processing and its Applications, (ISSPA'03)*, (Paris, France), pp. 3493–3496, July 2003.
[55] J. Lodge and M. Moher, "Maximum likelihood estimation of CPM signals transmitted over Rayleigh flat fading channels," *IEEE Trans. on Commun.*, vol. 38, pp. 787–794, June 1990.
[56] D. Makrakis, P. T. Mathiopoulos, and D. Bouras, "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. on Commun.*, vol. 42, pp. 63–75, January 1994.
[57] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. on Commun.*, vol. 43, pp. 1534–1544, February-April 1995.
[58] 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. on Commun.*, vol. 43, pp. 2750–2758, Nov. 1995.
[59] E. Biglieri, J. Proakis, and S. S. (Shitz), "Fading channels: information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
[60] O. Macchi and L. Scharf, "A dynamic programming algorithm for phase estimation and data decoding on random phase channels," *IEEE Trans. Inform. Theory*, pp. 581–595, September 1981.
[61] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice-Hall, 2000.
[62] A. Papoulis, *Probability, Random Variables and Sthocastic Processes*. New York, NY: McGraw-Hill, 1991.

PLACE PHOTO HERE

Giulio Colavolpe was born in Cosenza, Italy, in 1969. He received the Dr. Ing. degree in Telecommunications Engineering (cum laude) from the University of Pisa, Italy, in 1994 and the Ph.D. degree in Information Technology from the University of Parma, Italy, in 1998. Since 1997, he has been at the University of Parma, Italy, where he is now an Associate Professor of Telecommunications. In 2000, he was Visiting Scientist at the Institut Eurécom, Valbonne, France.

His main research interests include digital transmission theory, adaptive signal processing, channel coding and information theory. His research activity has led to more than seventy scientific publications in leading international journals and conference proceedings and a few industrial patents. He is also co-author of the book *Detection Algorithms for Wireless Communications, with Applications to Wired and Storage Systems* (New York: John Wiley & Sons, 2004). Dr. Colavolpe is also the principal investigator of several research projects funded by the European Space Agency (ESA-ESTEC) and important telecommunications companies.