Non-Binary Message-Passing Algorithms for Magnetic Channels with Data-Dependent Noise

Alessandro Tomasoni, Luca Reggiani

Dip. di Elettronica ed Informazione

Politecnico di Milano, 20133 Milano, ITALY

E-mail: {tomasoni,reggiani}@elet.polimi.it

Abstract—The paper proposes an implementation of the message passing algorithm adapted to iterative channel detection. The algorithm uses soft messages associated to non binary symbols in order to cancel cycles in the equivalent Tanner graphs, achieving optimal performance after a low number of iterations. This architecture, suited to very fast channel detectors, is applied to magnetic recording channels and adapted to the non stationary nature of the magnetic media noise.

Index Terms—Iterative decoding, low-density parity-check (LDPC) codes, maximum a posteriori decoder, message-passing algorithm, partial-response channel.

I. INTRODUCTION

In the last ten years storage technology has played a fundamental role in the progress of all the software and hardware applications in information technology. In this context, magnetic recording is the fundamental technology of information storage. Hard disk drives provide the features essential to this technology, particularly in terms of capacity, cost, access time and reliability of the stored data. Capacity growth of hard disks has followed and, at the same time, promoted the increasing demand for "room" in modern computer applications. From a signal processing point of view, the problem is challenging for two main reasons: (i) the length of the discrete equivalent channel impulse response of the magnetic channel (i.e. with considerable amount of inter-symbol interference) and (ii) the noise statistics, which is colored and data dependent since generated more by the precision of the writing/reading heads rather than by the additive gaussian electronic noise.

This paper presents a class of detectors based on the message passing algorithm (MP) and characterized by very low latency and an attractive parallel implementation. This solution, originally implemented for additive white Gaussian noise (AWGN), is then improved by matching more precisely the noise autocorrelation at the input of the detector, affected by the necessity of a partial response pre-equalization and data-dependent noise.

The solution adopted in this work is inspired by the approach presented in [1] and used for decoding binary information passed through a channel characterized by a discrete equivalent response. In [1], the authors adapt the message-passing principle, conceived for Low Density Parity Check

(LDPC) codes [2] [3], to the computation of soft information associated to bits that are convolved with a channel response function. Tanner graphs associated to channels with memory show that performance of this detector is greatly affected by the presence of cycles and the numerical results confirm this aspect of the application. In fact presence of cycles is intrinsic to the structure of channel detection since the memory taps of the impulse response operate on successive samples of the input data (a random interleaver cannot be used as in error correcting codes). This problem is addressed in [1] by means of an alternative parallel implementation of the soft detector based on message passing of state information in a forward and backward way, like in the BCJR algorithm [4], [5]. This paper deals with an alternative view of the problem and uses a non-binary implementation of the message passing algorithm for mitigating and cancelling the cycle impact. The non-binary version of the algorithm increases complexity in the message generation, maintaining its convenient parallel structure, suitable for integration with turbo and LDPC codes. In addition, the soft messages associated to non-binary symbols will be derived taking into account the statistics of the colored data dependent noise improving substantially detection performance.

The paper is organized as it follows. Sect. II introduces the system model and Sect. III describes the principle and the computations involved in the message passing algorithm extended to the non-binary case. Then, in Sect. IV, we investigate the application of noise predictors in this detector either for stationary or data-dependent noise. Numerical results on some significative examples are reported for all the proposed receiver architectures in Sect. V. Finally, Sect. VI concludes the paper.

II. SYSTEM MODEL

A magnetic recording system can be assimilated to a transmission system with data-dependent noise. In a magnetic recording channel, a reading/writing head, sensitive to the media polarization, converts the magnetic signal into an electrical one and viceversa. The signal deriving from a single magnetization transition can be modelled by the Lorentzian pulse

---

[1]Annual Congress GTTI 2008 - Session: Digital Transmissions.

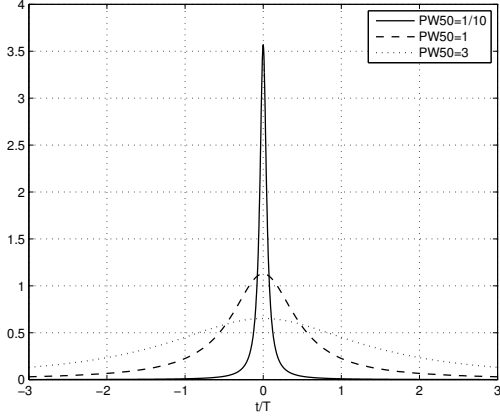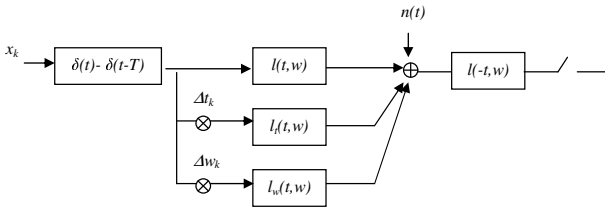Fig. 1. Lorentzian pulses with different PW50

Fig. 2. First order Taylor model of media noise, transmitter and matched filter.

$$l(t,w) = \sqrt{\frac{2}{\pi w}} \frac{1}{1 + \left(\frac{t}{w}\right)^2} \qquad (1)$$

where $w = \frac{PW50}{2}$, $PW50$ being the pulse width at 50% of the peak value. Fig. 1 depicts some Lorentzian pulses, with normalized energy and different $PW50$s with respect to the bit (or sampling) time T ($PW50/T$ is usually referred as D, the data density over the track).

A single recorded bit over the track generates two opposite transitions, one for each edge of the magnetic portion of the polarized track and the resulting signal is named *dbit*,

$$dbit(t,w) = l(t,w) - l(t - T, w). \qquad (2)$$

The writing/reading process, affected by additive gaussian noise (AWGN), corresponds to the top path of Fig. 2, where a matched filter is implemented at the receiver side. Nevertheless, non linear phenomena related to the non-ideal polarization on the medium, generate additional impairments in the read signal that can be assimilated to some width and time jitter of the Lorentzian pulses [6]. These errors, either in timing or pulse width, are comprehensively referred as *media noise* and they are usually dominant w.r.t. the additive white gaussian noise. In the system model, media noise effects correspond to the other two paths in Fig. 2 since they can be approximated by a first order Taylor expansion [6]:

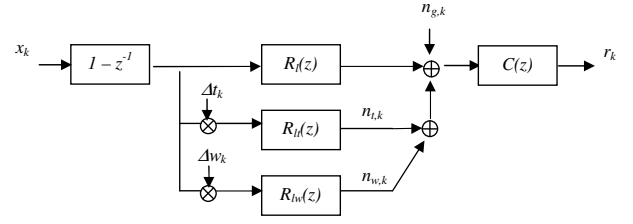$$l(t+\Delta t, w+\Delta w) \approx l(t,w) + \Delta t \cdot l_t(t,w) + \Delta w \cdot l_w(t,w) \qquad (3)$$

Fig. 3. Discrete equivalent channel model and equalizer $C(z)$.

where

$$
\begin{aligned}
l_t(t,w) &= \frac{d}{dt} l(t,w) \\
l_w(t,w) &= \frac{d}{dw} l(t,w).
\end{aligned}
$$

So, we can re-write the equivalent digital model, including the matched filter at the receiver into the three others pulse responses (Fig. 3). Fixing the parameter w (or D) we define

$$
\begin{aligned}
R_{l,k} &= l(t,w) * l(-t,w)|_{kT} \\
R_{lt,k} &= l_t(t,w) * l(-t,w)|_{kT} \\
R_{lw,k} &= l_w(t,w) * l(-t,w)|_{kT}. \qquad (4)
\end{aligned}
$$

In the sequel, for those simulations that includes the impact of media noise, we suppose the overall noise composed at 90% by the jitter and width noise (with equal power after the matched filter), and the remaining 10% by AWGN. The density D on the magnetic track is set to 2.8 and the signal-to-noise ratio (SNR) is defined according to [7].

After the matched filter and the sampler, the long discrete equivalent response of the system is usually equalized to a target response, or Partial Response (PR) [8], in order to reduce the complexity of the sequence detector (Fig. 3). It is well known that this approach is suboptimal, because noise is enhanced and it is no longer white. In the magnetic recording literature, the following target responses, that approximate the equivalent discrete response at low and high frequencies, are widespread:

$$
\begin{aligned}
h_{DI}(z) &= (1 - z^{-1}) \qquad (5) \\
h_{PR4}(z) &= (1 - z^{-1})(1 + z^{-1}) \\
h_{EPR4}(z) &= (1 - z^{-1})(1 + z^{-1})^2 \\
h_{EEPR4}(z) &= (1 - z^{-1})(1 + z^{-1})^3
\end{aligned}
$$

Denoting as L_{PR} the length of the partial channel response ($L_{PR} = 2, 3, 4, 5$), as $\mathbf{x}_{k_1}^{k_2}$ ($\mathbf{r}_{k_1}^{k_2}$) the column vector of input (received) symbols from discrete times k_1 to k_2, we can express the received samples as

$$\mathbf{r}_{k-N_c+1}^{k} = \begin{bmatrix} r_k \\ r_{k-1} \\ \vdots \\ r_{k-N_c+1} \end{bmatrix} = \mathbf{G}\mathbf{x}_{k-N_c-L_g+1}^{k} + \mathbf{n}_{k-N_c+1}^{k} \qquad (6)$$

where the matrix $\mathbf{G}$ ($N_c \times (L_g + N_c)$) implements the discrete convolution with the overall discrete equivalent channel and N_c is the equalizer length. The overall response after the equalizer $C(z)$ is forced to approximate a generic target

function $\mathbf{h}$ (5) by means of a Minimum Mean Square Error (MMSE) approach ($\mathbf{c}^T = \mathbf{h}^T \mathbf{R}_{xx} \mathbf{G}^T \left(\mathbf{G} \mathbf{R}_{xx} \mathbf{G}^T + \mathbf{R}'_{nn} \right)^{-1}$ with $\mathbf{R}_{xx}$ equal to the input autocorrelation).

This approach enhances the noise power and does not whiten it but it is attractive because it allows trellises with a smaller number of states and adapts the pre-equalizer formula to the input noise correlation matrix (the element $R'_{nn,k}$ is equal to $N_0/2 \cdot R_{l,k}$ (4) in case of presence of only AWGN noise or to the average of the autocorrelations from all the noise contributions $N_0/2 \cdot R_{l,k} + \overline{R}_{n_t n_t, k} + \overline{R}_{n_w n_w, k}$ in case of data-dependent media noise).

III. PARALLEL CHANNEL DETECTION

In a standard magnetic recording system, the pre-equalization stage is followed by a standard Maximum Likelihood Sequence Estimator (MLSE) or by a SISO (Soft Input Soft Output) algorithm for Maximum a Posteriori (MAP) detection as e.g. the BCJR algorithm [4] used extensively in soft and iterative decoding schemes. As anticipated in the introduction, here we propose a different structure of the sequence detector, based on a non-binary message passing algorithm, that solves the problem of short cycles over the inherent graph. In addition an enhanced version is capable of exploiting noise correlation, improving performance and allowing the possible employment of shorter PR responses, as it will be clear in the following sections. Nevertheless, a further theoretical research for PRs optimized for a better match with our equalizers can be an interesting research field in the future.

Message passing algorithms due their notoriety to Low-Density-Parity-Check (LDPC) codes [2], typically decoded by means of iterative algorithms over graphs [9], exchanging messages carrying likelihoods concerning variables. An analogous strategy can be applied to channel detection. Let us consider a block of N binary symbols $x_k \in \{-1, 1\}$ fed into a channel that can be viewed as a state machine returning the output samples y_k (in each binary representation the bit 0 corresponds to the level -1): the outputs belong to a finite alphabet and they are related to the inputs by

$$y_k = \sum_{i=0}^{L_{PR}-1} h_i \cdot x_{k-i} \tag{7}$$

where $\{h_i\}$ is the impulse response of a generic channel with memory $L_{PR} - 1$. We remark that, in our system model, the received samples $r_k = y_k + n_k$ contain the contribution of additive and media noise.

For illustrating our approach, Fig. 4 shows the equivalent Tanner graph for a channel with memory $L_{PR} - 1$ equal to 3: the triangles are the function (or channel) nodes where a posteriori likelihoods are updated at each iteration step and circles are the variable nodes (i.e. the input bits). The presence of cycles of length 4 or more is clear (bold lines in Fig. 4). Fig. 5 shows how a non-binary implementation of the same algorithm mitigates the presence of cycles with $M = 2$ and cancels all the cycles with $M = L_{PR} - 1 = 3$. At the same time, it is clear that the cardinality of the messages passed between the graph nodes grows correspondingly from 2 to 2^M. Hence the non binary algorithm has to manage soft messages

that are not real numbers (representing, e.g., the $P(x_k = 1)$) but vectors that contain the probability density functions of the symbols $\underline{x}_n$. At each iteration, data variable nodes and channel nodes simultaneously send new likelihood messages to their neighbors, computed exploiting the messages received at the previous step. By this procedure, agreeing messages will enforce their likelihoods overcoming the possible incorrect initial estimations.

Starting from the binary implementation (Fig. 4), we define with $R_{p \to n}$ the message going from the channel generic node p to the variable node n and viceversa for $Q_{n \to p}$. Focusing on channel nodes, the algorithm computes one different message for each connected destination variable node. At channel nodes, this is done because, with the aim of computing $R_{p \to n}$ to be sent to variable node n, the channel node gathers information coming from each connected node, with the exception of n itself (otherwise, variable nodes would self-influence
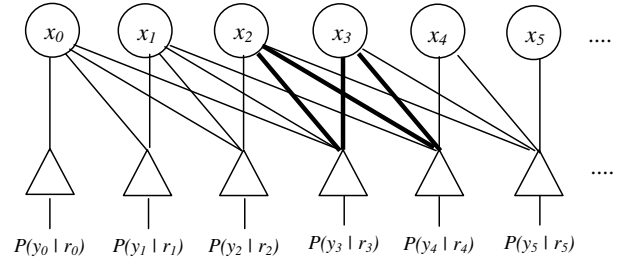

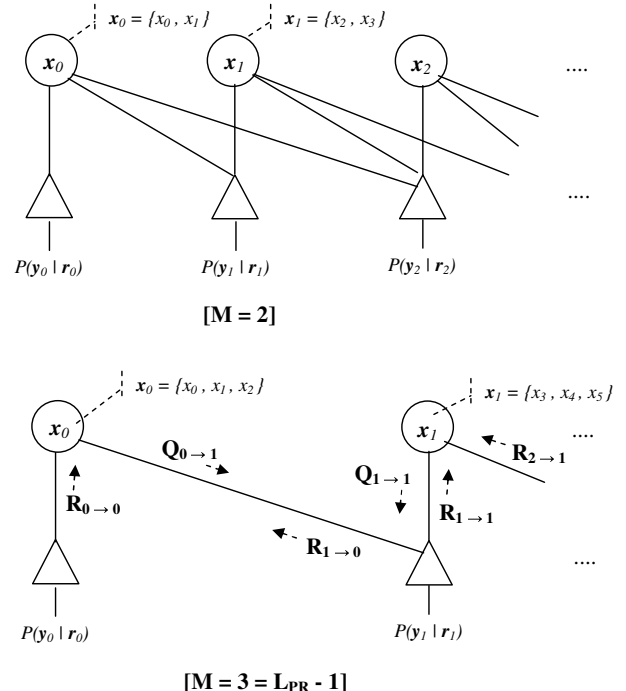Fig. 4. Binary message passing graph for a channel with memory equal to 3.

[M = 2]

[M = 3 = L_PR - 1]

Fig. 5. Non-binary ($M = 2, 3$) message passing graph for a channel with memory equal to 3.

themselves):

$$
\begin{aligned}
R_{p\to n}^{(1)} &= P\left(x_n = 1|r_p\right) \\
&= \sum_{j_1,\dots,j_\nu} P(x_n = 1, \mathbf{x}_{p-\nu}^{p\backslash n} = [j_1,\dots,j_\nu]\,|r_p) \\
&= \sum_{j_1,\dots,j_\nu} \frac{P(r_p, x_n = 1|\mathbf{x}_{p-\nu}^{p\backslash n} = [j_1,\dots,j_\nu])}{P(r_p)} \cdot \\
&\quad \cdot P(\mathbf{x}_{p-\nu}^{p\backslash n} = [j_1,\dots,j_\nu]) \\
&= \sum_{j_1,\dots,j_\nu} \frac{P(r_p|x_n = 1, \mathbf{x}_{p-\nu}^{p\backslash n} = [j_1,\dots,j_\nu])}{P(r_p)} \cdot \\
&\quad \cdot P(x_n = 1|\mathbf{x}_{p-\nu}^{p\backslash n} = [j_1,\dots,j_\nu]) \cdot \\
&\quad \cdot P(\mathbf{x}_{p-\nu}^{p\backslash n} = [j_1,\dots,j_\nu])
\end{aligned}
\tag{8}
$$

In (8) the sum is extended to all the possible binary symbols $[j_1,\dots,j_\nu]$ within the memory of the channel ($\nu = L_{PR}-1$) and sub-indexes $'p-\nu'$ and $'p\backslash n'$ denote that the sequence of symbols from $p-\nu$ to p does not include the n-th one. In AWGN the first term is computed by means of a gaussian probability density function, the second one is constant, while the third represents the likelihood information coming from all variable nodes, except for n, or

$$
P(\mathbf{x}_{p-\nu}^{p\backslash n} = [j_1,\dots,j_\nu]) = \prod_{k=p-\nu}^{p\backslash n} Q_{k\to p}^{(j_{k-p+\nu+1})}
$$

A similar processing is required at variable nodes where all the a-posteriori probabilities are combined for the following iteration step. The messages

$$
\begin{aligned}
Q_{n\to p}^{(j_{k-p+\nu+1})} &= \frac{\prod_{k=n}^{n+\nu\backslash p} P\left(x_n = j_{k-p+\nu+1}|r_k\right)}{\prod_k P\left(x_n = 1|r_k\right) + \prod_k P\left(x_n = 0|r_k\right)} \\
&= \frac{\prod_{k=n}^{n+\nu\backslash p} R_{k\to n}^{(j_{k-p+\nu+1})}}{\prod_{k=n}^{n+\nu\backslash p} R_{k\to n}^{(1)} + \prod_{k=n}^{n+\nu\backslash p}(1 - R_{k\to n}^{(1)})}
\end{aligned}
$$

are the resulting likelihoods, as products of the incoming likelihoods (except for the p-th), divided by a normalization term. This binary algorithm, although attractive, leads to consistent remarkable degradations, as shown in [1]. These losses are due to the short loops in the graph that, correlating the exchanged messages, enable wrong decisions to self-influence themselves.

In the non-binary implementation of the message passing algorithm, M consecutive input information bits or output samples are grouped into symbols that we represent as column vectors $\underline{\mathbf{x}}_n = \mathbf{x}_{M\cdot n}^{M\cdot(n+1)-1} = \{x_{M\cdot n}, x_{M\cdot n+1}, \cdots, x_{M\cdot(n+1)-1}\}^T$, $\underline{\mathbf{y}}_n = \mathbf{y}_{M\cdot n}^{M\cdot(n+1)-1}$, $\underline{\mathbf{r}}_n = \mathbf{r}_{M\cdot n}^{M\cdot(n+1)-1}$. Input symbols $\underline{\mathbf{x}}_n$ can assume 2^M values denoted $\underline{\mathbf{x}}^{(0)}, \cdots, \underline{\mathbf{x}}^{(2^M-1)}$ and corresponding to all the combinations of the input binary symbols. Now the notation $\underline{\mathbf{x}}^{(j)}$ ($j = 0, \cdots, 2^M-1$) corresponds to the M-bit binary representation of j with the most significant bit equal to $x_{M\cdot n}$ and the least one equal to $x_{M\cdot(n+1)-1}$. So the exchanged likelihoods does not refer to the single bit but to each possible combination of variables in the cluster. In addition, with $M = L_{PR}-1$, loops in the graphs are avoided and messages always propagate in the same direction. The computation of soft messages that go

from the p-th 'triangle' to the n-th 'circle' of the channel graph, may be re-formulated in the following way (first stage of the MP algorithm). Defining the general $\mathbf{R}_{p\to n}$ as

$$
\mathbf{R}_{p\to n} = \{R_{p\to n}^{(0)}, R_{p\to n}^{(1)}, \cdots, R_{p\to n}^{(2^M-1)}\}^T
\tag{9}
$$

with

$$
R_{p\to n}^{(j)} = P(\underline{\mathbf{x}}_n = \underline{\mathbf{x}}^{(j)}|\mathbf{r}_p),
\tag{10}
$$

the binary computation (8) is extended, in the cycle-free graph (Fig. 5), to

$$
\begin{aligned}
R_{n\to n}^{(j)} &= \frac{1}{P(\mathbf{r}_n)} \cdot \sum_{j_1} P(\underline{\mathbf{x}}_n = \underline{\mathbf{x}}^{(j)}|\underline{\mathbf{x}}_{n-1} = \underline{\mathbf{x}}^{(j_1)}) \cdot \\
&\quad \cdot P(\mathbf{r}_n|\underline{\mathbf{x}}_n = \underline{\mathbf{x}}^{(j)}, \underline{\mathbf{x}}_{n-1} = \underline{\mathbf{x}}^{(j_1)}) Q_{(n-1)\to n}^{(j_1)}
\end{aligned}
\tag{11}
$$

where the sum is performed to all possible combinations of input symbols; in this case the index j_1 is not limited to the binary values $0, 1$ but it is extended till to the cardinality of the non-binary symbols $0, \cdots, 2^M - 1$. Of course the other vector $\mathbf{R}_{n\to(n-1)}$ is computed similarly. Then the general terms $Q_{k\to p}^{(j)}$ are generated by the 'circle' node k and passed to the 'triangle' node p. Also these values are organized similarly in vectors $\mathbf{Q}_{k\to p} = \{Q_{k\to p}^{(0)}, Q_{k\to p}^{(1)}, \cdots, Q_{k\to p}^{(2^M-1)}\}^T$ and they represent the a-priori probabilities $P(\underline{\mathbf{x}}_k = \underline{\mathbf{x}}^{(j)})$ used in the next iteration step (11). It is straightforward to see that they are obtained by the simple forwarding (second stage of the MP algorithm)

$$
\begin{aligned}
\mathbf{Q}_{(n-1)\to n} &= \mathbf{R}_{(n-1)\to(n-1)} \tag{12} \\
\mathbf{Q}_{n\to n} &= \mathbf{R}_{(n+1)\to n}. \tag{13}
\end{aligned}
$$

A. Computation of the Non-Binary Message Vector $\mathbf{R}_{n\to n}$

It is interesting to note that the $R_{n\to n}^{(j)}$ messages (with $j = 0, \cdots, 2^{L_{PR}-1} - 1$) in (11) can be computed with an efficient algorithm in $L_{PR} - 1$ steps. The algorithm is based on a trellis structure that is determined by the channel memory. Skipping here the details, we observe that $P(\mathbf{r}_n|\underline{\mathbf{x}}_n = \underline{\mathbf{x}}^{(j)}, \underline{\mathbf{x}}_{n-1} = \underline{\mathbf{x}}^{(j_1)})$ is given by the product of the single terms $P(r_{n\cdot M+i}|\underline{\mathbf{x}}_n = \underline{\mathbf{x}}^{(j)}, \underline{\mathbf{x}}_{n-1} = \underline{\mathbf{x}}^{(j_1)})$ for $i = 0, \cdots, L_{PR} - 2$, that are reused as j varies in (11). So the computation of the entire vector $\mathbf{R}_{n\to n}$ is accomplished with an algorithm that receives the vector $\mathbf{Q}_{(n-1)\to n}$ as input and solves all the input/output combinations in a number of steps equal to the non-binary block length ($M = L_{PR} - 1$), just in the common trellis representation of the channel response (Fig. 6). In fact, the non-binary message passing detector inherits, from the ideal (MAP) one, the regularity of the required operations while introducing a remarkable degree of parallelism. Table I reports the required number of sums and multiplications per data variable in the binary and non-binary cases. Of course all the operations may be performed in the Log-domain as usual in MAP and MP detectors.

	Mul	Sum
Binary	$N_{iter}(\nu+1)\left(\nu 2^{\nu+1}-1\right)$	$N_{iter}(2^\nu+\nu)$
Non-Binary	$N_{iter}2^{\nu+2}$	$N_{iter}2^{\nu+1}+\nu(2^\nu-2)$

TABLE I
COMPUTATIONAL COSTS (PER DATA VARIABLE) FOR THE MESSAGE
PASSING ALGORITHM, IN THE BINARY AND NON-BINARY CYCLE-FREE
CASES (ν IS THE CHANNEL MEMORY LENGTH AND N_{iter} IS THE NUMBER
OF MP ITERATIONS).

IV. ENHANCED DETECTORS FOR COLORED NOISE

The detector presented in Sect. III solves the latency problems realizing a very fast version of a MAP (MLSE) detector. Nevertheless, the noise correlation at the detector input is neglected and this leads to suboptimal performance. We remark that this problem arises even without media-noise, because of the adoption of PR pre-equalization. The key point for mitigating these effects is to take into account the past noise samples in order to predict the future noise samples, and to subtract them by the observations, as depicted in Fig. 7 [11]. This operation can be included into the algorithm that returns the soft messages $\mathbf{R}_{p\to n}$ (Sect. III-A) because it affects only the computation of the terms $P(\mathbf{r}_n|\underline{\mathbf{x}}_n=\mathbf{x}^{(j)},\underline{\mathbf{x}}_{n-1}=\mathbf{x}^{(j_1)})$. The evaluation of these terms is modified, w.r.t. the AWGN assumption, by subtracting the noise portion that is predictable from the past ones. This operation would increase the detector memory from $L_{PR}-1$ to $L_{PR}+L_p-1$ and the single term
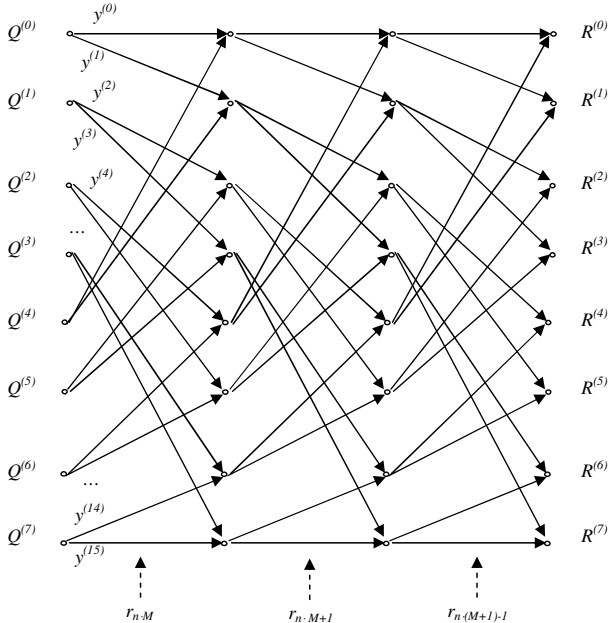

Fig. 6. Algorithm for computing the vector $\mathbf{R}_{n\to n}$ for M-bits blocks. At each step, a single sample $r_{n\cdot M+i}$ ($i=0,\cdots,M-1$) is used for computing all the channel probabilities.
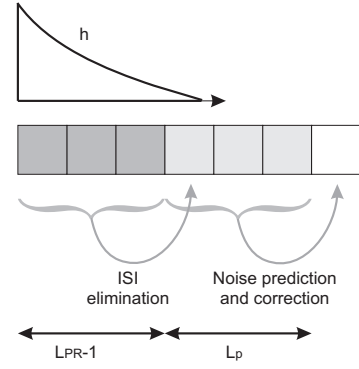

Fig. 7. Per-state noise predictors scheme. Inter-Symbol Interference is subtracted from each sample and then the last 3 noise samples are used for subtracting the noise predictable part from the current sample (the rightmost one).

may be expressed now by

$$P\left(r_{n\cdot M+i}|\underline{\mathbf{x}}_n=\mathbf{x}^{(j)},\underline{\mathbf{x}}_{n-1}=\mathbf{x}^{(j_1)},\underline{\mathbf{x}}^{(p)},\mathbf{r}_{n\cdot M+i-L_p}^{n\cdot M+i-1}\right)=$$
$$\frac{1}{\sqrt{2\pi\sigma_{NP}^2}}e^{\left(-\frac{\left(r_{n\cdot M+i}-y_{n\cdot M+i}-NP(\mathbf{x},\mathbf{r}_{n\cdot M+i-L_p}^{n\cdot M+i-1})\right)^2}{2\sigma_{NP}^2}\right)} \quad (14)$$

where $\mathbf{x}=\{\underline{\mathbf{x}}^{(j)},\underline{\mathbf{x}}^{(j_1)},\underline{\mathbf{x}}^{(p)}\}$, σ_{NP}^2 is the reduced noise power and $NP(\cdots)$ is the predictor function, that can be derived through the Yule-Walker equations as it will be shown in the next two subsections, dedicated to the stationary and non-stationary (data dependent) cases. The vector $\underline{\mathbf{x}}^{(p)}$ represents the portion of past inputs necessary for completing the memory required by the predictor operations. The increase of the detector memory can be treated in the following ways:

1) Effective extension of the memory in the trellis algorithm necessary for computing the soft messages (Sect. III-A). The number of states is increased from $2^{L_{PR}-1}$ to $2^{L_p+L_{PR}-1}$.
2) Average of $P\left(r_{n\cdot M+i}|\mathbf{x},\mathbf{r}_{n\cdot M+i-L_p}^{n\cdot M+i-1}\right)$ w.r.t. to all the past symbols $\underline{\mathbf{x}}^{(p)}$ for obtaining $P\left(r_{n\cdot M+i}|\underline{\mathbf{x}}_n=\mathbf{x}^{(j)},\underline{\mathbf{x}}_{n-1}=\mathbf{x}^{(j_1)},\mathbf{r}_{n\cdot M+i-L_p}^{n\cdot M+i-1}\right)$. In this case the a-priori probabilities $\mathbf{Q}$ would be taken in the adjacent variable nodes.
3) Use of hard decisions in the memory portion $\underline{\mathbf{x}}^{(p)}$ according to the the a-priori probabilities $\mathbf{Q}$ in the adjacent variable nodes.

In this paper we have chosen the first, full complexity, option for achieving the best performance improvement.

A. Stationary Colored Noise

If we denote as $\mathbf{n}_{k-L_p}^{k-1}$ the sequence of L_p past noise samples, the predictor $\mathbf{p}$ of length L_p which minimizes the mean square prediction error is given by

$$\mathbf{p}=\mathbf{R}_{nn}^{-1}\mathbf{q}_n \quad (15)$$

where $\mathbf{R}_{nn}=E\left[\mathbf{n}_{k-L_p}^{k-1}\left(\mathbf{n}_{k-L_p}^{k-1}\right)^T\right]$ is the noise stationary (Toeplitz) correlation matrix and $\mathbf{q}_n=E\left[\mathbf{n}_{k-L_p}^{k-1}n_k\right]$ is the

correlation vector between the current noise sample and the past ones. This noise predictor whitens partially the noise and it cancels its predictable part, reducing the noise power of a factor $(1 - \mathbf{p}^T\mathbf{q}_n)$.

We underline that the noise prediction procedure is common to each iteration and can be implemented in an efficient way. Here we are adopting the same prediction filter at each state and, consequently, it becomes convenient to exchange the order of the operations

$$
\begin{aligned}
NP(\mathbf{x}, \mathbf{r}_{j-L_p}^{j-1}) &= \mathbf{p}^T \left(\mathbf{r}_{j-L_p}^{j-1} - \mathbf{H}\mathbf{x} \right) \\
&= \mathbf{p}^T \mathbf{r}_{j-L_p}^{j-1} - \mathbf{p}^T \mathbf{H}\mathbf{x} \qquad (16)
\end{aligned}
$$

where j is a generic time index ($j = n \cdot M + i$ in (14)), $\mathbf{H}$ implements the partial response convolution with the generic channel $\mathbf{h}$ and $\mathbf{x} = \{\underline{\mathbf{x}}^{(j)}, \underline{\mathbf{x}}^{(j_1)}, \underline{\mathbf{x}}^{(p)}\}$ is the binary sequence including all the memory necessary for the predictor. We observe that the product $\mathbf{p}^T \mathbf{r}_{j-L_p}^{j-1}$ can be computed once for all the detector operations, while $\mathbf{p}^T \mathbf{H}\mathbf{x}$ can be pre-computed, as it does not depend on the received signal.

B. Data-Dependent (non-Stationary) Colored Noise

In the previous section, the noise correlation matrix has been considered independent from the data sequence. In the case of magnetic channels, affected by media noise, this is only an approximation. In order to enhance performances, it is possible to write per-state predictors, taking into account a more precise correlation matrix, according to the state of the algorithm for $\mathbf{R}_{n \to n}$ computation. This opportunity has been already investigated for MLSE detectors [11], [12] and MP binary detectors [13]. Here we extend this solution to the MP non-binary detector investigating the additional performance improvement. Per-state predictors require a different formulation of the Yule-Walker equations, given the non-stationary nature of the noise. We obtain a solution

$$
\mathbf{p}(\mathbf{x}) = (\mathbf{R}_{nn}(\mathbf{x}))^{-1} \mathbf{q}_n(\mathbf{x}) \qquad (17)
$$

with $\mathbf{x} = \{\underline{\mathbf{x}}^{(j)}, \underline{\mathbf{x}}^{(j_1)}, \underline{\mathbf{x}}^{(p)}\}$, that depends on the specific memory content. Notice that $\mathbf{R}_{nn}(\mathbf{x})$ is no longer a Toeplitz matrix and also the whitened noise variance

$$
\sigma_{NP}^2(\mathbf{x}) = \sigma^2 \left(1 - \mathbf{p}(\mathbf{x})^T \mathbf{q}_n(\mathbf{x}) \right) \qquad (18)
$$

depends on the past symbols. As a consequence, the probability computation (14) will be re-formulated accordingly.

V. SIMULATION RESULTS

This section reports the simulation results for the different proposed detection strategies. As mentioned in Sect. II, we assume a data density D equal to 2.8, with a media noise fraction equal to 90%. In the conventional MLSE detector, the number of steps is equal to the length of the data block, 4096 bits, while in the MP detector the bits are processed simultaneously in a number of iterations that is equal or smaller than 6. So we compare performance of our parallel non-binary detector with the corresponding serial MLSE implementation. We have simulated the targets h_{DI} and h_{EPR4}, reported in (5), with three different noise assumptions at the receivers:
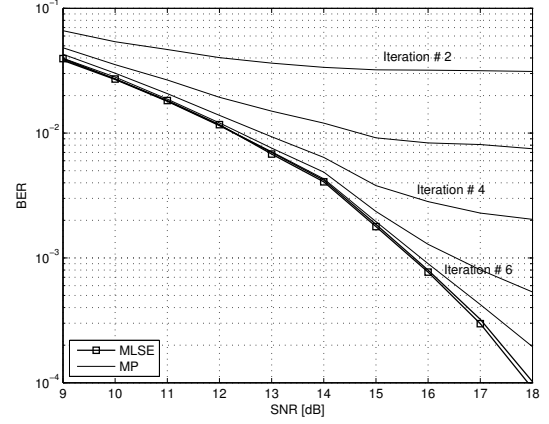

Fig. 8. MLSE (serial implementation) and MP (parallel implementation) performance, with $L_{PR} = 2$.
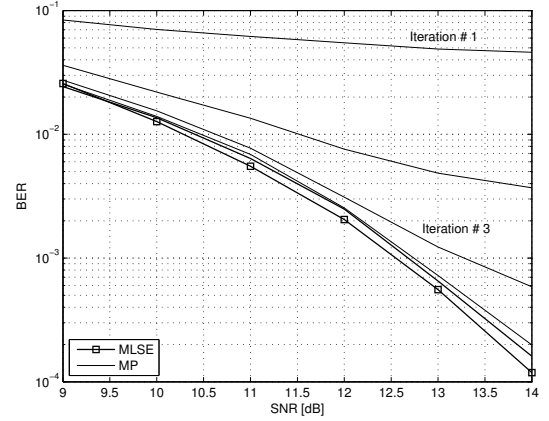

Fig. 9. MLSE (serial implementation) and MP (parallel implementation) performance, with $L_{PR} = 4$.

fully additive white noise, colored stationary noise and colored data-dependent noise.

When the noise is assumed additive and white, optimal MLSE performance is achieved in a number of iterations that is independent and much lower than the block length, as can be seen in Figs. 8 - 9.

When the receiver takes into account the noise correlation, the use of predictors reduces the overall noise impact on performance: Figs. 10 and 11 report the Bit Error Rate (BER) of MP detectors embedding different predictors (at the fifth iteration), with $L_{PR} = 2$ (the dicode channel) and $L_{PR} = 4$ (EPR4 class) respectively. Fig. 11 is restricted to even predictors, because the resulting autocorrelation function in this case has the samples in $k = \pm 1$ close to zero and performance of odd predictors is very close to the shorter even ones. We can observe that performance of the former (shorter) PR is about 0.75 dB from the latter (at $BER = 10^{-4}$), a performance gap that is much smaller than 4 dB, obtained without predictors. This could suggest new PR design criteria that, differently from [10], take into account both the message passing algorithm and the inherent noise whitening process.

Fig. 10. MP performance (after the fifth iteration) with different noise predictors and $L_{PR} = 2$.
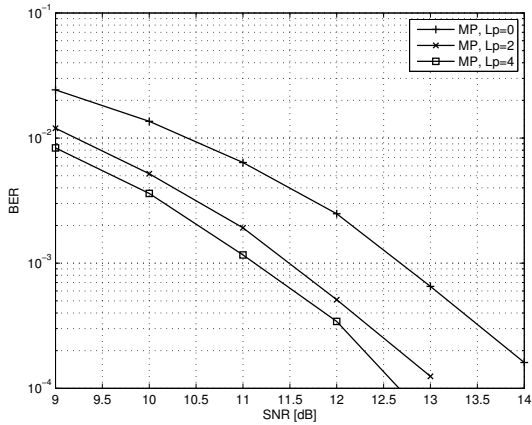

Fig. 11. MP performance (after the fifth iteration) with different noise predictors and $L_{PR} = 4$.

The final goal should be to effectively reduce the overall memory length, equal to $L_{PR} + L_p - 1$, jointly determined by the PR equalization and the noise predictor length.

Then Figs. 12 and 13 adopt per-state predictors, thus taking into account the actual data dependent noise nature. This more sophisticated approach leads to higher computational costs since the simplification in (16) no longer holds, but it guarantees additional gains (about 1 dB for $L_{PR} = 2$ and 0.5 dB for $L_{PR} = 4$) w.r.t. the previous solutions. Furthermore, noise prediction reduces the number of iterations required to approach the ideal detector performance: in Fig. 14 we can observe that predictors of order $L_p = 4$ allow a reduction of the number of iterations from 6 ($L_p = 0$) or 3 ($L_p = 2$) to 2. Finally we have also verified that including also future samples into the prediction process provides a negligible additional advantage (Fig. 15).

VI. CONCLUSIONS

In the context of channel detection, the paper presents an extension of the message-passing algorithm to the non-binary case, showing a way for cancelling cycles that limit perfor-
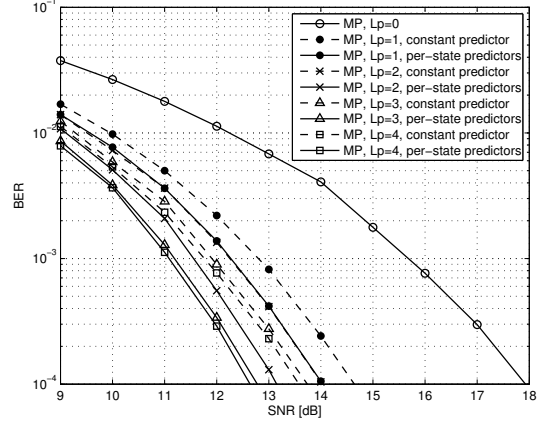

Fig. 12. MP performance, adopting the constant predictor or per-state predictors, with $L_{PR} = 2$.
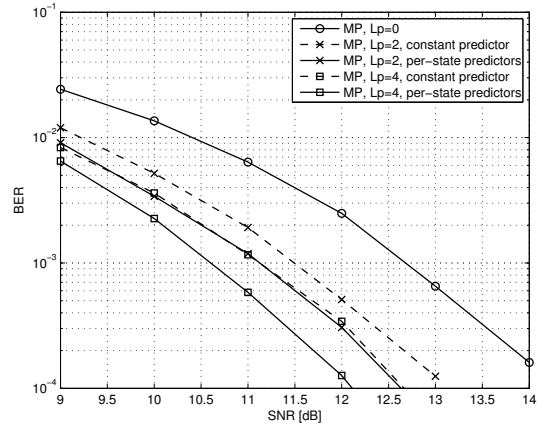

Fig. 13. MP performance, adopting the constant predictor or per-state predictors, with $L_{PR} = 4$.
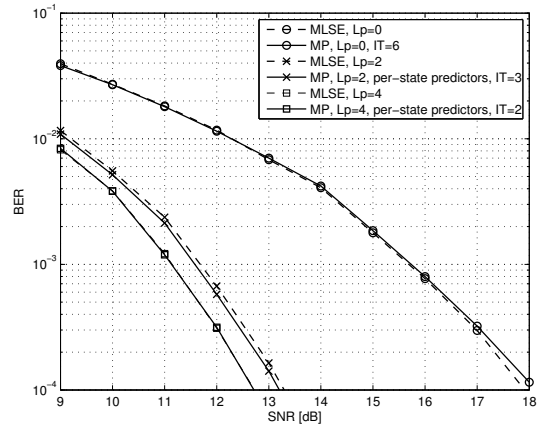

Fig. 14. MP performance: required number of iterations for approaching MLSE performance, when adopting per-state predictors (either in MLSE or MP). The channel has $L_{PR} = 2$ and IT is the number of iterations.
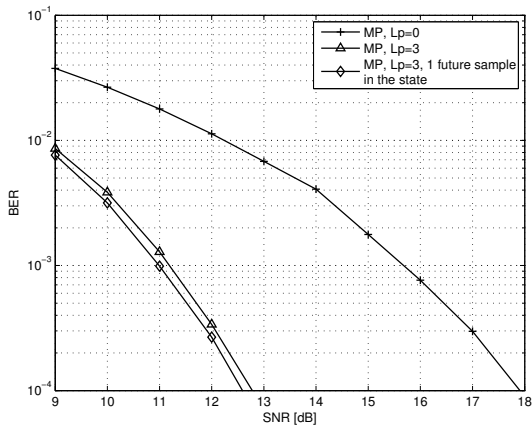
Fig. 15. Message passing performance when considering future samples, with $L_{PR} = 2$.

mance. The algorithm adopts a non-binary message passing structure, suited to a very low-latency parallel implementation, as an alternative to conventional serial detectors such as MAP or MLSE. In order to enhance performance in presence of data dependent media noise, predictors are included into the message passing procedure, achieving remarkable performance gains. All the simulated schemes show that noise prediction reduces also the number of iterations required by the message passing algorithm for achieving MLSE performance, with a further reduction of the latency. Moreover, performance with short partial responses do not decrease dramatically, differently from the architectures without noise predictors. This suggests an investigation on new partial response pre-equalization design criteria, that minimize the overall detector

memory induced by the partial response equalization and by noise prediction, for a prescribed BER. Finally, we remark that the proposed algorithms can output soft extrinsic information, useful when the channel detector is followed by LDPC or turbo codes.

REFERENCES

[1] B. M. Kurkoski, P. H. Siegel, J. K. Wolf, "Joint Message-Passing Decoding of LDPC Codes and Partial-Response Channels," *IEEE Trans. on Information Theory*, Vol. 48, No. 6, pp. 1410-1422, June 2002.
[2] R. G. Gallager, *Low-Density Parity-Check Codes*, MIT Press, Cambridge, MA, 1963.
[3] M. C. Davey, D. MacKay, "Low-Density Parity Check Codes over GF(q)," *IEEE Communications Letters*, Vol. 2, No. 6, pp. 165-167, June 2002.
[4] L.R. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. on Information Theory*, vol. 20, pp. 284-287, March 1974.
[5] P. Robertson, P. Hoeher, and E. Villebrun, "A Comparison of Optimal and Sub-Optimal MAP Decoding Algorithms Operating in the Log Domain," *Proc. IEEE Int. Conf. Communications*, 1995, pp. 1009-1013.
[6] J. Moon, "Discrete-Time Modelling of Transition-Noise-Dominant Channels and Study of Detection Performance," *IEEE Trans. on Magnetics*, pp. 4573-4578, 1991.
[7] J. Moon, "Signal-to-Noise Definition for Magnetic Recording Channels with Transition Noise," *IEEE Trans. on Magnetics*, pp. 3881-3883, 2000.
[8] R. Hermann and W. Hirt and R. D. Cideciyan and F. Dolivo and W. Schott, "A PRML System for Digital Magnetic Recording," *IEEE J. Select. Areas Commun.*, pp. 38-56, 1992.
[9] R. M. Tanner, "A Recursive Approach to Low Complexity Codes," *IEEE Trans. on Information Theory*, vol. IT-27, pp.533-547, Sept. 1981.
[10] H. Sun and G. Mathew, "Detection Techniques for High-Density Magnetic Recording," *IEEE Trans. on Magnetics*, pp. 1193-1199, 2005.
[11] J. Moon and J. Park, "Pattern-Dependent Noise Prediction in Signal-Dependent Noise," *IEEE J. Select. Areas Commun.*, pp. 730-743, 2001.
[12] A. Kavcic, J. M. F. Moura, "Signal-Dependent Correlation-Sensitive Branch Metrics for Viterbi-like Sequence Detectors," *Proc. IEEE Int. Conf. Communications*, pp. 657-661, 1998.
[13] M. N. Kaynak, T.M. Duman, E.M. Kurtas, "Noise Predictive Belief Propagation," *IEEE Trans. on Magnetics*, Vol. 41, No. 12, pp. 4427-4434, Dec. 2005.