

# Multidimensional Signal Processing and Detection for Storage Systems With Data-Dependent Transition Noise

Riccardo Pighi<sup>1</sup>, Riccardo Raheli<sup>1</sup>, and Umberto Amadei<sup>2</sup>

<sup>1</sup>Università degli Studi di Parma, Dipartimento di Ingegneria dell'Informazione, I-43100 Parma, Italy

<sup>2</sup>Selta S.p.A., 29010 Cadeo, Piacenza, Italy

**In the last decade, significant research on detection algorithms capable of mitigating the effects of colored Gaussian thermal noise and transition noise in storage systems, has been performed. In this paper, we present a new detection scheme based on a multidimensional detector front end and multidimensional linear prediction, applied to maximum *a posteriori* probability (MAP) sequence detection. This method improves the bit-error-rate (BER) performance with respect to previous approaches and makes the detector quite insensitive to transition noise. We show that the gain in terms of BER versus signal-to-noise ratio with our detector increases with the user density. The results obtained for a magnetic storage channel are extendable to optical storage systems as well.**

***Index Terms*—Longitudinal and perpendicular recording, magnetic storage systems, multidimensional linear prediction, multidimensional signal processing, optical storage systems, transition noise.**

## I. INTRODUCTION

**R**ECORDING densities in magnetic storage systems continue to increase at a considerable rate. These high recording densities require sophisticated detection schemes in order to preserve system reliability. High-density longitudinal and perpendicular magnetic recording systems based on thin-film media exhibit severe intersymbol interference, colored Gaussian thermal noise, and signal-dependent transition noise. The last kind of noise, also known as media noise, is due to the magnetic interaction between data transitions in the information sequence stored on the medium: therefore, transition noise is data-dependent and its power increases with the recording density. Transition noise comprises transition jitter and transition width fluctuations: they both occur in the writing process whenever a transition in the writing current polarity take place. Transition jitter is a fluctuation of a transition's position: this effect results in a shift of the pulse position at the output of the reading head. On the other hand, width variation is a fluctuation of the duration of a transition: its effect is that the pulse shape at the output of the channel is distorted in amplitude. Transition noise could account for as much as 90% of the total noise power in recording systems [1], [2].

In the literature, a few channel models have been proposed to enable the analysis and design of optimum detectors, among which the microtrack model [3], the signal-dependent autoregressive channel model [4], and the position jitter and width variation model [2], [5]. The latter is used in this paper.

After the definition of a suitable channel model, many authors proposed detection schemes based on signal processing algorithms to reduce the effects of noise in magnetic recording channels. In [6], a detection scheme based on linear prediction

was applied to the estimation of colored thermal noise and a modified Euclidean-distance branch metric computation in the Viterbi algorithm was proposed in order to incorporate linear prediction and enable maximum *a posteriori* probability (MAP) sequence detection. The performance improvement comes from an effective whitening of the noise samples, which exhibit correlation at the detector input due to partial response equalization [7]–[9].

The maximum likelihood sequence detector (MLSD) for a signal-dependent Gaussian noise has been first derived in [10] under the assumption that the noise can be modeled as an autoregressive (AR) process: the resulting structure is a Viterbi algorithm which incorporates signal-dependent noise prediction into the branch metric computation [11]. Finally, in [12], linear prediction extended to signal-dependent transition noise was applied to near-MLSD and other low-complexity sequence detection structures.

According to the model in [2] and [5], the observable can be viewed as conditionally Gaussian, given the data, and one is enabled to exploit the principle of linear predictive detectors proposed for fading channels [13]–[16]. As a matter of fact, transition noise, due to its multiplicate nature with respect to the information sequence, has some similarity with a fading process that usually affects wireless links.

In this paper, we extend the results in [12] and [16] to a detector based on a multidimensional front end for longitudinal and perpendicular magnetic recording channels and optical storage systems. In more detail, the presence of transition noise and the need for statistical sufficiency yield a detector front end with a number of filters proportional to the modeling order of the transition noise. Multidimensional linear prediction can be applied to estimate the realization of the transition noise process in order to incorporate its realization into a Viterbi algorithm and enable MAP sequence detection. Moreover, thanks to the proposed multidimensional detector front end, linear prediction

can operate over a wider set of samples. Numerical analysis demonstrates good improvements in terms of minimum mean square prediction error (MMSPE) with respect to the MMSPE achievable using a monodimensional front end, yielding an improved estimation of transition noise. Since magnetic recording channels are essentially communication systems characterized by a great amount of intersymbol interference, we used a bank of “whitening” filters matched to the impulse response of the channel in order to reduce the time dispersion of the impulse response of the information bearing signal as well as the length of the position jitter and width variation modeling impulses.

The state-complexity of the Viterbi algorithm used in the proposed multidimensional linear predictive detector was decoupled from the prediction order by means of state reduction techniques [17]–[19]. Specifically, we consider state reduction by memory truncation and apply per-survivor processing [20] (PSP) and tentative (preliminary) decision techniques. Simulations demonstrate good improvements, in terms of signal-to-noise ratio (SNR), of the proposed multidimensional linear predictive detector using a front end based on two or three filters, with respect to conventional detectors.

Finally, since the first-order channel model used to describe the transition noise process can be applied also to optical and magneto-optical recording systems [21]–[24], such as digital versatile disk (DVD) and compact disk (CD), the proposed multidimensional signal processing technique is extended to these systems, enabling us to outperform monodimensional detectors that previously appeared in the literature. This paper expands upon preliminary work presented in [25].

## II. CHANNEL MODEL

In order to describe the proposed multidimensional front end, we begin by considering a magnetic recording channel modeled by a first-order position jitter and width variation [5]. Our results can be also extended to optical and magneto-optical recording systems affected by transition noise in a straightforward manner. Let  $h(t, w)$  denote the response to an isolated transition recorded in magnetic or optical media, where  $t$  is time and  $w$  is a parameter characterizing the pulsewidth. Let  $a_k \in \{\pm 1\}$  be the information bits to be stored. Assuming that transition noise can be decomposed into position jitter and width variation, the read back waveform  $r(t)$  corrupted by additive white Gaussian thermal noise  $\eta(t)$  with power spectral density  $N_0/2$  can be expressed as

$$r(t) = \sum_k b_k h(t + \Delta t_k - kT, w + \Delta w_k) + \eta(t) \quad (1)$$

where  $b_k = a_k - a_{k-1} \in \{0, \pm 2\}$  denote transition symbols,  $\Delta t_k$  and  $\Delta w_k$ , modeled as independent Gaussian random variables with standard deviations  $\sigma_{\Delta t}$  and  $\sigma_{\Delta w}$ , represent the effect of position jitter and width variation noise, respectively, and  $T$  is the symbol period. Obviously, when  $\sigma_{\Delta t} = 0$  and  $\sigma_{\Delta w} = 0$ , the model reduces to a magnetic recording channel without transition noise. For the pulse response  $h(t, w)$ , we have adopted

the well-known Lorentzian approximation [26] for longitudinal recording, i.e.,

$$h(t, w) = \frac{1}{1 + (2t/PW_{50})^2}$$

where  $PW_{50} = 2w$  is the pulsewidth at half the maximum amplitude. For perpendicular recording systems, we adopt the approximation [27], [28]

$$h(t, w) = \operatorname{erf}\left(\frac{2\sqrt{\ln 2}t}{PW_{50}}\right) \quad (2)$$

where we have introduced the error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz$$

and defined the parameter  $PW_{50} = 2w$  as the pulsewidth at half the maximum amplitude of  $\partial h(t, w)/\partial t$ . We define the parameter  $D = PW_{50}/T$  as the normalized density.

According to [2] and [29], taking an  $n$ th-order Taylor series expansion of the read back impulse, the signal at the channel output can be approximated as a linear sum of the noise-free response and residual responses due to deviations around the nominal position and width of the pulse. Limiting the series expansion to the first order, the read back impulse associated to this (first-order) channel model can be approximated as

$$h(t + \Delta t_k, w + \Delta w_k) \simeq h(t, w) + \Delta t_k \frac{\partial h(t, w)}{\partial t} + \Delta w_k \frac{\partial h(t, w)}{\partial w}. \quad (3)$$

Defining the impulse response of the filters modeling the position jitter and width variation noise process as<sup>1</sup>

$$h_t(t) = \frac{\partial h(t, w)}{\partial t}$$

$$h_w(t) = \frac{\partial h(t, w)}{\partial w}$$

and using this first-order approximation (3) in (1), the continuous waveform at the output of the channel can be approximated as  $r(t) = y(t) + \eta(t)$ , where we have defined  $y(t)$  as

$$y(t) = \sum_k b_k [h(t - kT) + \Delta t_k h_t(t - kT) + \Delta w_k h_w(t - kT)]. \quad (4)$$

A block diagram descriptive of the first-order channel model is shown in Fig. 1.

## III. SUFFICIENT STATISTICS

We now derive a set of sufficient statistics for the considered digital storage channel affected by transition noise. The method can be applied to both longitudinal and perpendicular recording

<sup>1</sup>The subscript denotes the variable of differentiation.

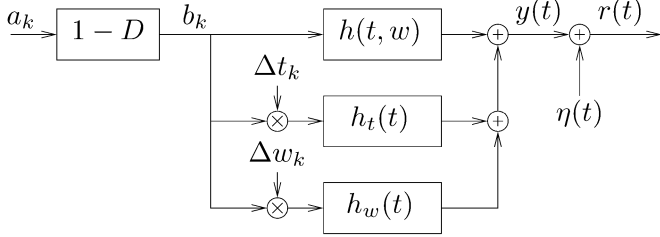


Fig. 1. Storage channel model with first-order media noise and additive white Gaussian thermal noise.

systems, as well as optical storage systems. The signal at the output of the channel can be expressed as<sup>2</sup>

$$r(t) = y(t, \mathbf{a}, \boldsymbol{\theta}) + \eta(t) \quad (5)$$

where  $\mathbf{a}$  is a data information vector,  $\boldsymbol{\theta}$  is a random vector collecting the unknown parameters affecting the observable, i.e., the sequences of random variables  $\{\Delta t_k\}$  and  $\{\Delta w_k\}$ ,  $\eta(t)$  is an additive white Gaussian thermal noise process with monolateral power spectral density  $N_0$  and  $y(t, \mathbf{a}, \boldsymbol{\theta})$  is defined as in (4). Given a probabilistic model of  $\boldsymbol{\theta}$  with realizations in a suitable space  $\Theta$  and noting that, for any finite number of transmitted bits, an information lossless discretization of signal  $r(t)$  by expansion over an orthonormal finite-dimensional basis can be achieved, the detection strategy can be formulated as

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} P(\mathbf{a}) f(\mathbf{r}|\mathbf{a}) \quad (6)$$

where  $P(\mathbf{a})$  is the *a priori* probability of the information sequence  $\mathbf{a}$  and  $f(\mathbf{r}|\mathbf{a})$  is the conditional probability density function (pdf) of the observation vector  $\mathbf{r}$ , given the information sequence  $\mathbf{a}$ . Under the assumption of statistical independence between  $\boldsymbol{\theta}$  and  $\mathbf{a}$ , the conditional probability density function in (6) can be expressed as

$$f(\mathbf{r}|\mathbf{a}) = \int_{\Theta} f(\mathbf{r}|\boldsymbol{\theta}, \mathbf{a}) f(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (7)$$

in which the integral is over the parameter space  $\Theta$  and  $f(\boldsymbol{\theta})$  is the pdf of vector  $\boldsymbol{\theta}$ . Given the sequences of random variables  $\{\Delta t_k\}$  and  $\{\Delta w_k\}$ , i.e., the parameter vector  $\boldsymbol{\theta}$ , and the data sequence  $\mathbf{a}$ , the observation vector  $\mathbf{r}$  is conditionally Gaussian. Therefore, it is possible to express the conditional pdf of the observation vector as

$$\begin{aligned} f(\mathbf{r}|\boldsymbol{\theta}, \mathbf{a}) &= \frac{1}{(\pi N_0)^{\frac{N}{2}}} e^{-\frac{1}{N_0} \|\mathbf{r} - \mathbf{y}(\mathbf{a}, \boldsymbol{\theta})\|^2} \\ &\sim \frac{1}{(\pi N_0)^{\frac{N}{2}}} e^{-\frac{1}{N_0} [\|\mathbf{y}(\mathbf{a}, \boldsymbol{\theta})\|^2 - 2\Re\{\mathbf{r}^T \mathbf{y}(\mathbf{a}, \boldsymbol{\theta})\}]} \end{aligned} \quad (8)$$

<sup>2</sup>The used notation highlights the dependence of  $y(t)$  on a random vector  $\boldsymbol{\theta}$  and an information vector  $\mathbf{a}$ .

where  $\|\cdot\|$  denotes the Euclidean norm, the quantity  $\|\mathbf{r}\|^2$  is irrelevant in the detection process and can be discarded,  $\mathbf{y}(\mathbf{a}, \boldsymbol{\theta})$  is a discretization of  $y(t, \mathbf{a}, \boldsymbol{\theta})$ ,  $N$  is the dimensionality of the observation vector and the symbol  $\sim$  denotes a monotonic relationship with respect to the variable of interest (i.e., the data sequence  $\mathbf{a}$ ). Using (8) in (7), we obtain

$$\begin{aligned} f(\mathbf{r}|\mathbf{a}) & \\ &\sim \frac{1}{(\pi N_0)^{\frac{N}{2}}} \int_{\Theta} e^{-\frac{1}{N_0} [\|\mathbf{y}(\mathbf{a}, \boldsymbol{\theta})\|^2 - 2\Re\{\mathbf{r}^T \mathbf{y}(\mathbf{a}, \boldsymbol{\theta})\}]} f(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

The discrete correlation between the observation vector  $\mathbf{r}$  and the signal vector  $\mathbf{y}(\mathbf{a}, \boldsymbol{\theta})$  can be equivalently expressed in the time domain, thanks to the optimal discretization procedure, as a correlation integral

$$\mathbf{r}^T \mathbf{y}(\mathbf{a}, \boldsymbol{\theta}) = \int_{-\infty}^{\infty} r(t) y(t; \mathbf{a}, \boldsymbol{\theta}) dt. \quad (9)$$

Similarly, the square Euclidean norm of  $\mathbf{y}(\mathbf{a}, \boldsymbol{\theta})$  is equal to the energy of the signal

$$\|\mathbf{y}(\mathbf{a}, \boldsymbol{\theta})\|^2 = \int_{-\infty}^{\infty} y^2(t; \mathbf{a}, \boldsymbol{\theta}) dt.$$

Using (4), the correlation integral in (9) can be expressed as

$$\begin{aligned} \int_{-\infty}^{\infty} r(t) y(t; \mathbf{a}, \boldsymbol{\theta}) dt &= \sum_k \left[ \int_{-\infty}^{\infty} b_k r(t) h(t - kT) dt \right. \\ &\quad + b_k \Delta t_k \int_{-\infty}^{\infty} r(t) h_t(t - kT) dt \\ &\quad \left. + b_k \Delta w_k \int_{-\infty}^{\infty} r(t) h_w(t - kT) dt \right]. \end{aligned} \quad (10)$$

Finally, defining the following quantities

$$\begin{aligned} x'_k &= \int_{-\infty}^{\infty} r(t) h(t - kT) dt \\ y'_k &= \int_{-\infty}^{\infty} r(t) h_t(t - kT) dt \\ z'_k &= \int_{-\infty}^{\infty} r(t) h_w(t - kT) dt \end{aligned}$$

Equation (10) becomes

$$\begin{aligned} \int_{-\infty}^{\infty} r(t) y(t; \mathbf{a}, \boldsymbol{\theta}) dt &= \sum_k b_k x'_k \\ &\quad + \sum_k b_k \Delta t_k y'_k \\ &\quad + \sum_k b_k \Delta w_k z'_k. \end{aligned} \quad (11)$$

Equation (11) shows that  $(x'_k, y'_k, z'_k)^T$  is a vector of sufficient statistics for the detection process in a storage channel with data-dependent transition noise, according to the model of Fig. 1.

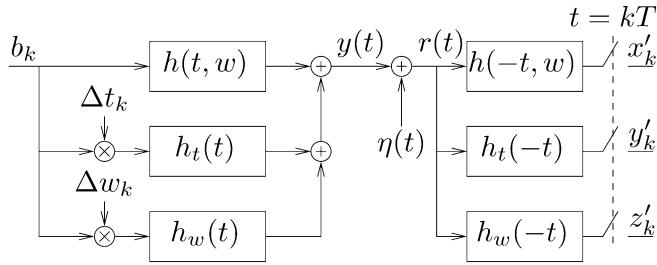


Fig. 2. Multidimensional detector front end for a storage channel model with first-order media noise and additive white Gaussian thermal noise.

These sufficient statistics can be obtained from the time-continuous received signal  $r(t)$  by means of a multidimensional detection front end, as shown in Fig. 2.

The proposed front end is based on a bank of filters, each followed by a sampler at the symbol rate: the first filter is the usual matched filter, whereas the second and the third filters are matched to the second and third impulse responses modeling the transition noise, respectively. We remark that commonly used front ends are based on the matched filter  $h(-t, w)$  only, in spite of the fact that in the presence of transition noise the discrete observation sequence  $\{x'_k\}$  is not a sufficient statistic. An intuitive explanation of the fact that  $\{x'_k\}$  is not a sufficient statistic in the presence of transition noise may be based on the multiplicative nature of this noise with respect to the transition sequence  $\{b_k\}$ . In fact, the transition noise waveform can be viewed as a noisy signal corrupting information bits, but also as another kind of information-bearing signal superimposed on the useful signal. This point of view is confirmed also by recent works in [30] and [31], where it is shown that for a magnetic recording channel with transition noise, the information rate at the output of a matched filter<sup>3</sup> is increased with respect to a channel model affected by colored thermal noise only. This means that the transition noise process is ultimately a signal carrying useful information. The proposed front end enables the extraction of this residual information from the observation of  $r(t)$  and its use to improve system reliability, i.e., to improve bit-error-rate (BER) performance.

It is worth highlighting that our derivation of the sufficient statistics can be straightforwardly extended to storage channel models with transition noise of higher order [2], [29]. The number of filters in the multidimensional front end is controlled by the degree of approximation of the transition noise process, i.e., the order of the Taylor series expansion of the read back impulse. For example, a second-order channel model would result in a detector front end composed of a bank of five matched filters.

Linear prediction can be applied to estimate the realization of the transition noise process in order to incorporate its realization into a Viterbi algorithm and enable MAP sequence or symbol detection. With respect to the algorithms proposed in the literature [6], [12], which deal with colored Gaussian thermal noise and transition noise, respectively, and are based on the observation of  $\{x'_k\}$  only, we can now operate on an augmented set of useful observations, allowing the described detector to outperform the previously proposed detection schemes.

<sup>3</sup>The first matched filter in the front end in Fig. 2.

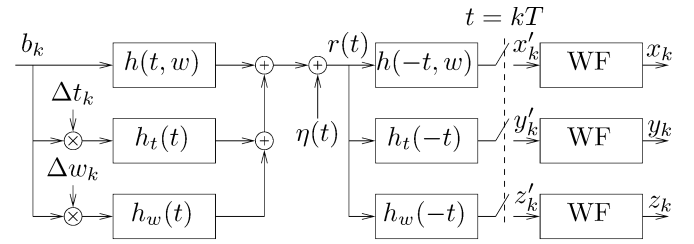


Fig. 3. Multidimensional detector front end with whitening filters for a channel model with first-order media noise and additive white Gaussian thermal noise.

We remark that the set of sufficient statistics is independent of the nature of the random variables modeling the transition noise. If we assume correlation between transition noise samples  $\Delta t_k$  and  $\Delta w_k$ , the set of sufficient statistics does not change, and linear prediction is still optimal. If we drop the Gaussianity assumption on  $\Delta t_k$  and  $\Delta w_k$ , the observation sequences at the output of the proposed multidimensional front end are sufficient statistics for a given approximation order, despite linear prediction is not optimal. As a consequence, under the non-Gaussian assumption, the linear predictive detector described in the following sections is conceptually suboptimal, although it may still offer good performance in practical cases. Finally, note that, at higher transition noise power, the observable is not Gaussian even because higher order terms in the Taylor series expansion may not be neglected. However, we expect that, at high SNR values, where the higher order effects become negligible, the performance of the linear prediction detector shows a limited SNR loss. These issues are the subject of current investigation.

Since the relevant impulse responses are characterized by a large amount of intersymbol interference, in order to reduce the complexity of the detector, instead of adopting a partial response equalizer with the purpose of channel shortening, a bank of thermal noise whitening filters (WF) matched to the impulse responses of the multidimensional front end can be used.

The proposed multidimensional detector front end with a bank of whitening filters is shown in Fig. 3: note that, from the reversibility theorem [32], [33], the set  $(x_k, y_k, z_k)^T$  at the output of whitening filters is still a vector of sufficient statistics. The whitening process reduces the dispersion of the impulse response of the information-bearing signal as well as the length of the position jitter and width variation modeling impulses. These whitening filters decorrelate the thermal noise samples in the time domain only, in the sense that the thermal noise samples at the output of the first, second, and third matched filters are independent from past or future samples at the output of the first, second, and third filters, respectively. The output samples, however, are still correlated with each other in "space." This means that a nonvanishing cross correlation still exists between noise samples, at different time indexes, at the output of the three matched filters. However, this is not a limiting factor because a linear predictive detector can cope with this correlation. We remark the fact that we use a bank of WF not for whitening the thermal noise samples (this could be alternatively done using linear prediction) but with the purpose of obtaining an overall discrete time channel model as close as possible to a minimum phase system for which the energy is delayed of a minimum amount among all systems having a given magnitude frequency response [34].

Having obtained a set of sufficient statistics, we can now investigate how these quantities can be used to perform sequence detection: this issue is addressed in the next section.

#### IV. DETECTION STRATEGY BASED ON LINEAR PREDICTION

Assuming a first-order channel model, we have shown that the quantities at the output of the whitening filters ( $x_k, y_k, z_k$ ) are sufficient statistics for sequence detection. Collecting these samples into suitable vectors  $\mathbf{x}, \mathbf{y}, \mathbf{z}$ , we can reformulate the detection strategy (6) in equivalent form as

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} P(\mathbf{a}) f(\mathbf{x}, \mathbf{y}, \mathbf{z} | \mathbf{a}). \quad (12)$$

Assuming causality and applying the chain factorization rule to the multidimensional conditional pdf, we obtain

$$f(\mathbf{x}, \mathbf{y}, \mathbf{z} | \mathbf{a}) = \prod_{k=0}^{K-1} f(x_k, y_k, z_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^{k-1}, \mathbf{z}_0^{k-1}; \mathbf{a}_0^k) \quad (13)$$

where  $\mathbf{x}_{k_1}^{k_2}$  is a shorthand notation for the vector collecting signal observations from time epoch  $k_1$  to  $k_2$  and  $K$  is the length of the transmission. The three-dimensional conditional pdf in (13) can be further factored as a product of three monodimensional conditional pdfs, i.e.,

$$\begin{aligned} f(x_k, y_k, z_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^{k-1}, \mathbf{z}_0^{k-1}; \mathbf{a}_0^k) \\ = f(x_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^k, \mathbf{z}_0^k; \mathbf{a}_0^k) \\ \cdot f(y_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^{k-1}, \mathbf{z}_0^k; \mathbf{a}_0^k) \\ \cdot f(z_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^{k-1}, \mathbf{z}_0^{k-1}; \mathbf{a}_0^k). \end{aligned} \quad (14)$$

We can observe that the first pdf in (14) is conditioned on the past output samples from the first channel, from time 0 to time  $k-1$ , and from the second and the third channel, from time 0 to current time  $k$ ; the second pdf is conditioned on the past output samples from the first and second channel, from time 0 to time  $k-1$ , and from the third channel, from time 0 to current time  $k$ ; the last pdf is conditioned on the past output samples from all channels, from time 0 to time  $k-1$ .

Using (14), we can now rewrite (13) as

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}, \mathbf{z} | \mathbf{a}) &= \prod_{k=0}^{K-1} [f(x_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^k, \mathbf{z}_0^k; \mathbf{a}_0^k) \\ &\cdot f(y_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^{k-1}, \mathbf{z}_0^k; \mathbf{a}_0^k) \\ &\cdot f(z_k | \mathbf{x}_0^{k-1}, \mathbf{y}_0^{k-1}, \mathbf{z}_0^{k-1}; \mathbf{a}_0^k)] \\ &\simeq \prod_{k=0}^{K-1} [f(x_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^k, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k) \\ &\cdot f(y_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k) \\ &\cdot f(z_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^{k-1}; a_k, \zeta_k)]. \end{aligned} \quad (15)$$

In the last step of (15), in order to limit the detector's memory, we have assumed Markovianity of order  $\nu$  in the conditional observation sequences. Moreover, we have defined a state of the

system accounting for the ‘‘postcursors’’ and ‘‘precursors’’ of the impulse responses and the order of Markovianity  $\nu$  as

$$\zeta_k = (a_{k-1}, a_{k-2}, a_{k-3}, \dots, a_{k-L}) \quad (16)$$

where  $L = \delta_1 + \delta_2 + \nu$ , with  $\delta_1$  and  $\delta_2$  denoting the number of precursors and postcursors in the impulse responses. The assumed Markovianity results in an approximation whose quality increases with  $\nu$ .

Since the thermal and transition noise processes have Gaussian distribution, the observation is Gaussian, given the data. The application of the chain factorization rule in (15) allows us to factor the multidimensional conditional pdf in (12) as a product of  $3K$  one-dimensional conditionally Gaussian pdfs, completely defined by the conditional means

$$\begin{aligned} \hat{x}_k &= E\{x_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^k, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k\} \\ \hat{y}_k &= E\{y_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k\} \\ \hat{z}_k &= E\{z_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^{k-1}; a_k, \zeta_k\} \end{aligned}$$

and the conditional variances

$$\begin{aligned} \hat{\sigma}_{x_k}^2 &= E\{[x_k - \hat{x}_k]^2 | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^k, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k\} \\ \hat{\sigma}_{y_k}^2 &= E\{[y_k - \hat{y}_k]^2 | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k\} \\ \hat{\sigma}_{z_k}^2 &= E\{[z_k - \hat{z}_k]^2 | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^{k-1}; a_k, \zeta_k\}. \end{aligned}$$

It should be now clear that  $\hat{x}_k$ ,  $\hat{y}_k$ , and  $\hat{z}_k$  can be interpreted as linear predictive estimates of  $x_k$ ,  $y_k$ , and  $z_k$ , respectively, and  $\hat{\sigma}_{x_k}^2$ ,  $\hat{\sigma}_{y_k}^2$ , and  $\hat{\sigma}_{z_k}^2$  as the relevant MMSPEs [35]. It is also possible to express explicitly the conditional means as<sup>4</sup>

$$\begin{aligned} \hat{x}_k &= E\{x_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^k, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k\} = s_{1,k}(a_k, \zeta_k) \\ &+ \sum_{i=1}^{\nu} p_{1,1,i}(a_k, \zeta_k)[x_{k-i} - s_{1,k-i}(a_k, \zeta_k)] \\ &+ \sum_{i=0}^{\nu} p_{1,2,i}(a_k, \zeta_k)[y_{k-i} - s_{2,k-i}(a_k, \zeta_k)] \\ &+ \sum_{i=0}^{\nu} p_{1,3,i}(a_k, \zeta_k)[z_{k-i} - s_{3,k-i}(a_k, \zeta_k)] \end{aligned} \quad (17)$$

$$\begin{aligned} \hat{y}_k &= E\{y_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k\} = s_{2,k}(a_k, \zeta_k) \\ &+ \sum_{i=1}^{\nu} p_{2,1,i}(a_k, \zeta_k)[x_{k-i} - s_{1,k-i}(a_k, \zeta_k)] \\ &+ \sum_{i=1}^{\nu} p_{2,2,i}(a_k, \zeta_k)[y_{k-i} - s_{2,k-i}(a_k, \zeta_k)] \\ &+ \sum_{i=0}^{\nu} p_{2,3,i}(a_k, \zeta_k)[z_{k-i} - s_{3,k-i}(a_k, \zeta_k)] \end{aligned} \quad (18)$$

$$\begin{aligned} \hat{z}_k &= E\{z_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^{k-1}; a_k, \zeta_k\} = s_{3,k}(a_k, \zeta_k) \\ &+ \sum_{i=1}^{\nu} p_{3,1,i}(a_k, \zeta_k)[x_{k-i} - s_{1,k-i}(a_k, \zeta_k)] \\ &+ \sum_{i=1}^{\nu} p_{3,2,i}(a_k, \zeta_k)[y_{k-i} - s_{2,k-i}(a_k, \zeta_k)] \\ &+ \sum_{i=1}^{\nu} p_{3,3,i}(a_k, \zeta_k)[z_{k-i} - s_{3,k-i}(a_k, \zeta_k)] \end{aligned} \quad (19)$$

<sup>4</sup>We note that the prediction process is not homogeneous [36].

in which the linear prediction coefficients  $p_{\ell,m,i}(a_k, \zeta_k)$  at time  $i$  are the solution of the Wiener–Hopf matrix equation [37], with index  $m$  denoting the branch number in the channel model of Fig. 1 (from top to bottom) and  $\ell$  denoting the branch number in the detector front end. The observation sequences have nonzero conditional means, given the data,  $s_{1,k}(a_k, \zeta_k)$ ,  $s_{2,k}(a_k, \zeta_k)$ , and  $s_{3,k}(a_k, \zeta_k)$ , defined as

$$\begin{aligned} s_{1,k}(a_k, \zeta_k) &= E\{x_k | a_k, \zeta_k\} = \sum_{i=-\delta_1}^{\delta_2} a_i g_{k-i}^{(1)} \\ s_{2,k}(a_k, \zeta_k) &= E\{y_k | a_k, \zeta_k\} = \sum_{i=-\delta_1}^{\delta_2} a_i g_{k-i}^{(2)} \\ s_{3,k}(a_k, \zeta_k) &= E\{z_k | a_k, \zeta_k\} = \sum_{i=-\delta_1}^{\delta_2} a_i g_{k-i}^{(3)} \end{aligned}$$

where  $g_k^{(1)}$ ,  $g_k^{(2)}$  and  $g_k^{(3)}$  are the discrete-time impulse responses of the noiseless signal components at the output of the first, second, and third front end filter, respectively.

Note also that, for a given value of  $\nu$ , the number of prediction coefficients  $p_{\ell,m,i}(a_k, \zeta_k)$  used in (17)–(19) changes with respect to the number of past samples defined in the conditioning events and that prediction coefficients are state dependent. For instance, from (17) one can see that  $\hat{x}_k$  depends on  $\nu$  samples at the output of the first whitening filter, and on  $\nu + 1$  samples at the output of the second and third whitening filters. On the other hand, from (18), it should be noted that  $\hat{y}_k$  depends on  $\nu$  samples at the output of the first and second whitening filters and on  $\nu + 1$  samples at the output of the third whitening filter, and, from (19), one can note that  $\hat{z}_k$  depends on  $\nu$  samples at the output of all whitening filters. The solution of the Wiener–Hopf matrix equation for multidimensional linear prediction will be presented in Section V.

The detection strategy (12), the factorization (15), and the linear prediction (17)–(19) allow us to derive the branch metric to be used for sequence detection in a Viterbi detector. Taking the logarithm, we can express these branch metrics as

$$\begin{aligned} \lambda_k(a_k, \zeta_k) &= \ln f(x_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^k, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k) \\ &\quad + \ln f(y_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^k; a_k, \zeta_k) \\ &\quad + \ln f(z_k | \mathbf{x}_{k-\nu}^{k-1}, \mathbf{y}_{k-\nu}^{k-1}, \mathbf{z}_{k-\nu}^{k-1}; a_k, \zeta_k) \\ &\quad + \ln P[a_k(\zeta_k)]. \end{aligned}$$

Assuming that the information bits are independent and identically distributed, and discarding irrelevant terms, the resulting branch metrics can be finally expressed as

$$\begin{aligned} \lambda_k(a_k, \zeta_k) &= -\frac{[x_k - \hat{x}_k]^2}{\hat{\sigma}_{x_k}^2} - \ln \hat{\sigma}_{x_k}^2 \\ &\quad -\frac{[y_k - \hat{y}_k]^2}{\hat{\sigma}_{y_k}^2} - \ln \hat{\sigma}_{y_k}^2 \\ &\quad -\frac{[z_k - \hat{z}_k]^2}{\hat{\sigma}_{z_k}^2} - \ln \hat{\sigma}_{z_k}^2. \end{aligned}$$

The state-complexity of a linear predictive detector can be naturally decoupled from the prediction order  $\nu$  by means of

state-reduction techniques [17]–[20]. Let  $Q < L$  denote the memory parameter to be taken into account in the definition of a “reduced” trellis state

$$\omega_k = (a_{k-1}, a_{k-2}, \dots, a_{k-Q}).$$

The branch metric can be obtained by defining a “pseudostate” [38]

$$\tilde{\zeta}_k(\omega_k) = \left( \underbrace{a_{k-1}, \dots, a_{k-Q}}_{Q \text{ bits}}, \underbrace{\check{a}_{k-Q-1}(\omega_k), \dots, \check{a}_{k-Q-P}(\omega_k)}_{P \text{ bits}}, \underbrace{\hat{a}_{k-Q-P-1}, \dots, \hat{a}_{k-L}}_{U=L-Q-P \text{ bits}} \right) \quad (20)$$

where  $P$  bits may be chosen by a per-survivor processing technique [20], and the last  $U = L - Q - P$  bits can be defined as tentative (or preliminary) decisions  $\hat{a}_k$  at the detector output. Note that  $\check{a}_{k-Q-1}(\omega_k), \dots, \check{a}_{k-Q-P}(\omega_k)$  are the information bits associated with the survivor of  $\omega_k$ . The branch metric  $\tilde{\lambda}_k(a_k, \omega_k)$  in the reduced-state trellis can be defined in terms of this pseudostate (20) according to

$$\tilde{\lambda}_k(a_k, \omega_k) = \lambda_k(a_k, \tilde{\zeta}_k(\omega_k)).$$

## V. MULTIDIMENSIONAL LINEAR PREDICTION

In this section, we describe how linear prediction can be applied to a multidimensional observation vector collecting the sufficient statistics  $(x_k, y_k, z_k)^T$  and how to obtain an estimate of the transition noise samples at the output of each whitening filter. As in a monodimensional scenario, we start defining a cost function  $J(\mathbf{P})$  which represents the mean square error between the transition noise samples at the output of the multidimensional detector and a possible set of estimates of the media noise process.

Defining the vectors  $\boldsymbol{\chi}_k = [x_k, y_k, z_k]^T$  and  $\mathbf{s}_k = [s_{1,k}(a_k, \zeta_k), s_{2,k}(a_k, \zeta_k), s_{3,k}(a_k, \zeta_k)]^T$  collecting the channel and the noiseless signal components at the output of the whitening filters, respectively, it is possible to express the cost function as

$$J(\mathbf{P}) = E \left\{ \left\| (\boldsymbol{\chi}_k - \mathbf{s}_k) - \sum_i^{\nu} \mathbf{P}_i [\boldsymbol{\chi}_{k-i} - \mathbf{s}_{k-i}] \right\|^2 \middle| a_k, \zeta_k \right\} \quad (21)$$

where  $\mathbf{P}$  is a matrix collecting all prediction coefficients, which must be determined in order to minimize the cost function  $J(\mathbf{P})$ . The quantity

$$\boldsymbol{\chi}_k - \mathbf{s}_k = [x_k - s_{1,k}(a_k, \zeta_k), y_k - s_{2,k}(a_k, \zeta_k), z_k - s_{3,k}(a_k, \zeta_k)]^T$$

denotes the transition noise samples we want to predict, and the quantities  $\mathbf{x}_{k-i} - \mathbf{s}_{k-i}$  represent the *data* [35] (i.e., the past samples of transition noise) to be used to perform linear prediction. Note that the sum index  $i$  in (21) is not explicitly initialized, because it depends on the number of the transition noise samples available at the output of the whitening filters.

The cost function (21) can be expressed explicitly as

$$\begin{aligned}
J(\mathbf{P}) = E \left\{ \left[ \begin{aligned} & [x_k - s_{1,k}(a_k, \zeta_k)] \\ & - \sum_{i=1}^{\nu} p_{1,1,i}(a_k, \zeta_k) [x_{k-i} - s_{1,k-i}(a_k, \zeta_k)] \\ & - \sum_{i=0}^{\nu} p_{1,2,i}(a_k, \zeta_k) [y_{k-i} - s_{2,k-i}(a_k, \zeta_k)] \\ & - \sum_{i=0}^{\nu} p_{1,3,i}(a_k, \zeta_k) [z_{k-i} - s_{3,k-i}(a_k, \zeta_k)] \end{aligned} \right]^2 \right. \\
& + \left[ \begin{aligned} & [y_k - s_{2,k}(a_k, \zeta_k)] \\ & - \sum_{i=1}^{\nu} p_{2,1,i}(a_k, \zeta_k) [x_{k-i} - s_{1,k-i}(a_k, \zeta_k)] \\ & - \sum_{i=1}^{\nu} p_{2,2,i}(a_k, \zeta_k) [y_{k-i} - s_{2,k-i}(a_k, \zeta_k)] \\ & - \sum_{i=0}^{\nu} p_{2,3,i}(a_k, \zeta_k) [z_{k-i} - s_{3,k-i}(a_k, \zeta_k)] \end{aligned} \right]^2 \\
& + \left[ \begin{aligned} & [z_k - s_{3,k}(a_k, \zeta_k)] \\ & - \sum_{i=1}^{\nu} p_{3,1,i}(a_k, \zeta_k) [x_{k-i} - s_{1,k-i}(a_k, \zeta_k)] \\ & - \sum_{i=1}^{\nu} p_{3,2,i}(a_k, \zeta_k) [y_{k-i} - s_{2,k-i}(a_k, \zeta_k)] \\ & - \sum_{i=1}^{\nu} p_{3,3,i}(a_k, \zeta_k) [z_{k-i} - s_{3,k-i}(a_k, \zeta_k)] \end{aligned} \right]^2 \left. \right| a_k, \zeta_k \Big\}. \quad (22)
\end{aligned}$$

Since the cost function defined in (22) is a sum of three positive functions of disjoint sets of variables, the minimization can be performed separately on each function. In the following, we show how to obtain the prediction coefficients for the first

channel (i.e.,  $\{p_{1,1,i}\}$ ,  $\{p_{1,2,i}\}$  and  $\{p_{1,3,i}\}$ ) only. Defining a prediction vector  $\mathbf{p}_1$  for the first channel as<sup>5</sup>

$$\mathbf{p}_1 = [\mathbf{p}_{1,1}^T; \mathbf{p}_{1,2}^T; \mathbf{p}_{1,3}^T]$$

and a *data* vector

$$\mathbf{d}_{k-\nu}^k = \begin{bmatrix} \underbrace{\left( \mathbf{x}_{k-\nu}^{k-1} - \mathbf{s}_{1,k-\nu}^{k-1}(a_k, \zeta_k) \right)^T}_{\boldsymbol{\phi}_{k-\nu}^{k-1}}; \\ \underbrace{\left( \mathbf{y}_{k-\nu}^k - \mathbf{s}_{2,k-\nu}^k(a_k, \zeta_k) \right)^T}_{\boldsymbol{\psi}_{k-\nu}^k}; \\ \underbrace{\left( \mathbf{z}_{k-\nu}^k - \mathbf{s}_{3,k-\nu}^k(a_k, \zeta_k) \right)^T}_{\boldsymbol{\mu}_{k-\nu}^k} \end{bmatrix}$$

where  $\boldsymbol{\phi}_{k-\nu}^{k-1}$  are  $\nu$  transition noise samples at the output of the first whitening filter,  $\boldsymbol{\psi}_{k-\nu}^k$  and  $\boldsymbol{\mu}_{k-\nu}^k$  are  $\nu + 1$  media noise samples at the output of the second and third whitening filter, we can express the cost function for the first channel as

$$J_1(\mathbf{p}_1) = E \left\{ \left[ \boldsymbol{\phi}_k - \mathbf{p}_1^T \cdot \mathbf{d}_{k-\nu}^k \right] \left[ \boldsymbol{\phi}_k - \mathbf{p}_1^T \cdot \mathbf{d}_{k-\nu}^k \right]^T \middle| a_k, \zeta_k \right\}.$$

Applying the gradient operator with respect to the prediction vector  $\mathbf{p}_1$ , we have

$$\begin{aligned}
\nabla_{\mathbf{p}_1} J_1(\mathbf{p}_1) = 2E \left\{ \mathbf{d}_{k-\nu}^k \cdot \left( \mathbf{d}_{k-\nu}^k \right)^T \middle| a_k, \zeta_k \right\} \cdot \mathbf{p}_1 \\
- 2E \left\{ \boldsymbol{\phi}_k \cdot \mathbf{d}_{k-\nu}^k \middle| a_k, \zeta_k \right\}. \quad (23)
\end{aligned}$$

We are now able to formulate the Wiener-Hopf equation as

$$\mathbf{R}(a_k, \zeta_k) \cdot \mathbf{p}_1 = \mathbf{q}(a_k, \zeta_k) \quad (24)$$

where the system matrix is defined as in (25) at the bottom of the page, and the vector of known terms is

$$\mathbf{q}(a_k, \zeta_k) = \begin{bmatrix} E \{ \boldsymbol{\phi}_k \boldsymbol{\phi}_{k-\nu}^{k-1} \} \\ E \{ \boldsymbol{\phi}_k \boldsymbol{\psi}_{k-\nu}^k \} \\ E \{ \boldsymbol{\phi}_k \boldsymbol{\mu}_{k-\nu}^k \} \end{bmatrix}.$$

<sup>5</sup>In order to simplify the notation, we omit the dependence of  $\mathbf{p}_1$  on  $(a_k, \zeta_k)$ .

$$\mathbf{R}(a_k, \zeta_k) = \begin{bmatrix} E \left\{ \left( \boldsymbol{\phi}_{k-\nu}^{k-1} \right) \cdot \left( \boldsymbol{\phi}_{k-\nu}^{k-1} \right)^T \right\} & E \left\{ \left( \boldsymbol{\phi}_{k-\nu}^{k-1} \right) \cdot \left( \boldsymbol{\psi}_{k-\nu}^k \right)^T \right\} & E \left\{ \left( \boldsymbol{\phi}_{k-\nu}^{k-1} \right) \cdot \left( \boldsymbol{\mu}_{k-\nu}^k \right)^T \right\} \\ E \left\{ \left( \boldsymbol{\psi}_{k-\nu}^k \right) \cdot \left( \boldsymbol{\phi}_{k-\nu}^{k-1} \right)^T \right\} & E \left\{ \left( \boldsymbol{\psi}_{k-\nu}^k \right) \cdot \left( \boldsymbol{\psi}_{k-\nu}^k \right)^T \right\} & E \left\{ \left( \boldsymbol{\psi}_{k-\nu}^k \right) \cdot \left( \boldsymbol{\mu}_{k-\nu}^k \right)^T \right\} \\ E \left\{ \left( \boldsymbol{\mu}_{k-\nu}^k \right) \cdot \left( \boldsymbol{\phi}_{k-\nu}^{k-1} \right)^T \right\} & E \left\{ \left( \boldsymbol{\mu}_{k-\nu}^k \right) \cdot \left( \boldsymbol{\psi}_{k-\nu}^k \right)^T \right\} & E \left\{ \left( \boldsymbol{\mu}_{k-\nu}^k \right) \cdot \left( \boldsymbol{\mu}_{k-\nu}^k \right)^T \right\} \end{bmatrix} \quad (25)$$

The system matrix  $\mathbf{R}(a_k, \zeta_k)$  is based on the autocorrelation sequence of the transition noise samples at the output of the first whitening filter and the crosscorrelation between these samples and the samples at the output of the other whitening filters. Thanks to the multidimensional front end, we can exploit the crosscorrelation between media noise samples to improve the prediction process. We remark that the noise samples  $\phi_{k-\nu}^{k-1}$ ,  $\psi_{k-\nu}^k$ , and  $\mu_{k-\nu}^k$  are not available at the detector: they must be evaluated through the observation of the output of the multidimensional front end and a reconstruction of noiseless signal components associated with the survivor path leading to state  $\zeta_k$ .

The linear system defined in (24) can now be solved using Cholesky factorization [35], obtaining the prediction coefficient vector

$$\mathbf{p}_1 = \mathbf{R}^{-1}(a_k, \zeta_k) \cdot \mathbf{q}(a_k, \zeta_k).$$

Since the transition noise sample we wish to estimate is data dependent, the system matrix  $\mathbf{R}(a_k, \zeta_k)$  and the vector  $\mathbf{q}(a_k, \zeta_k)$  are also data dependent: it should be clear that the prediction coefficients are also data dependent. The prediction coefficients can be precomputed, given the state  $\zeta_k$  and the current information bit  $a_k$ , and stored in a lookup table.

Rewriting the cost function  $J_1(\mathbf{p}_1)$  as an explicit function of the predictor vector  $\mathbf{p}_1$ , we obtain

$$\begin{aligned} J_1(\mathbf{p}_1) &= E \left\{ \left[ \phi_k - \mathbf{p}_1^T \cdot \mathbf{d}_{k-\nu}^k \right] \cdot \left[ \phi_k - \mathbf{p}_1^T \cdot \mathbf{d}_{k-\nu}^k \right]^T \right\} \\ &= E \{ |\phi_k|^2 \} - 2E \left\{ \mathbf{p}_1^T \cdot \mathbf{d}_{k-\nu}^k \cdot \phi_k \right\} \\ &\quad + E \left\{ \mathbf{p}_1^T \cdot \mathbf{d}_{k-\nu}^k \cdot \left( \mathbf{d}_{k-\nu}^k \right)^T \cdot \mathbf{p}_1 \right\} \\ &= \sigma_\phi^2 - 2 \cdot \mathbf{p}_1^T \cdot E \left\{ \mathbf{d}_{k-\nu}^k \cdot \phi_k \right\} \\ &\quad + \mathbf{p}_1^T \cdot E \left\{ \mathbf{d}_{k-\nu}^k \cdot \left( \mathbf{d}_{k-\nu}^k \right)^T \right\} \cdot \mathbf{p}_1 \end{aligned}$$

where  $\sigma_\phi^2$  is the variance of the transition noise. Using the definition of  $\mathbf{p}_1$ ,  $\mathbf{R}(a_k, \zeta_k)$ , and  $\mathbf{q}(a_k, \zeta_k)$ , we can express the MMSPE as

$$\begin{aligned} J_1(\mathbf{p}_1) &= \sigma_\phi^2 - 2 \cdot \mathbf{p}_1^T \cdot \mathbf{q}(a_k, \zeta_k) + \mathbf{p}_1^T \cdot \mathbf{R}(a_k, \zeta_k) \cdot \mathbf{p}_1 \\ &= \sigma_\phi^2 - \mathbf{q}^T(a_k, \zeta_k) \cdot \mathbf{R}^{-1}(a_k, \zeta_k) \cdot \mathbf{q}(a_k, \zeta_k) \\ &= \sigma_\phi^2 - \mathbf{p}_1^T \cdot \mathbf{q}(a_k, \zeta_k). \end{aligned}$$

The same procedure can be used for the computation of the prediction coefficients and the relevant MMSPEs associated with the second and third front end filter.

## VI. SIMULATION RESULTS

We start considering magnetic longitudinal storage channels. In Fig. 4, we show the prediction error obtained both for a monodimensional front end (indicated by 1D), for a bidimensional front end (2D) and for a three-dimensional front end (3D), for a normalized density  $D = 2.50$ . In the bidimensional case, the second front end filter was selected to be matched to either the time derivate of the Lorentzian pulse ( $2D_t$  curves)

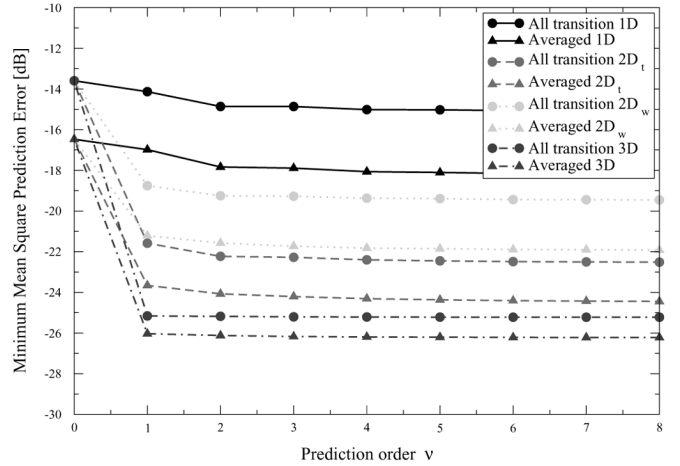


Fig. 4. MMSPE versus number of prediction coefficients  $\nu$  for longitudinal magnetic recording with  $D = 2.50$ ,  $\alpha = 95\%$ ,  $\lambda = 50\%$ , and  $\text{SNR}_{95} = 10$  dB.

or the width derivate of the Lorentzian pulse ( $2D_w$  curves). Since media noise arises in transitions, a bit pattern characterized by continuous changes of the writing current's polarity, the so-called "all-transition" pattern  $\{1, -1, 1, -1, \dots\}$ , was considered. An average MMSPE is also shown, by averaging over all possible bit patterns defining a trellis branch  $(a_k, \zeta_k)$  ("averaged" pattern).

The SNR with transition noise [39] is defined at the input of the Viterbi detector for a one-dimensional front end as

$$\text{SNR}_\alpha = \frac{P_s}{\sigma_n^2 + \sigma_m^2}$$

where  $P_s$  is the signal power,  $\sigma_n^2$  is the thermal noise power,  $\sigma_m^2$  is the transition noise power, and  $\alpha = 100 \times [\sigma_m^2 / (\sigma_n^2 + \sigma_m^2)]$  denotes the percentage of transition noise with respect to total noise. In order to evaluate the MMSPE, the signal-to-noise ratio was fixed at  $\text{SNR}_{95} = 10$  dB, i.e., assuming a 95% transition noise consisting of 50% position jitter and 50% width variation (in a more compact form, we can write  $\lambda = 50\%$ , where  $\lambda$  is the fraction of transition noise power due to position jitter [39]). Fig. 4 shows that, using a bidimensional front end and a prediction order  $\nu = 2$ , it is possible to obtain gains between 4.5 and 7.0 dB for the all-transition bit pattern, with respect to the MMSPE obtained by a monodimensional detector front end (1D curves). Finally, the MMSPE obtained for the three-dimensional detector, further improves over the results of the bidimensional front end by at least 1.5 dB.

Fig. 5 shows the BER obtained for a monodimensional front end without transition noise (ISI + AWGN curve), with transition noise but without linear prediction (ISI + Transition Noise curve) and with both transition noise and linear prediction (Linear Prediction curve), for a density  $D = 2.50$  and a 95% transition noise with  $\lambda = 50\%$ . All the BER curves are obtained by Monte Carlo simulations with random data sequences: the number of transmitted bits is set to  $128 \cdot 10^8$ , which guarantees good numerical accuracy. In Fig. 5 we set  $\delta_1 = 3$  precursors,  $\delta_2 = 8$  postcursors, a prediction order  $\nu = 2$ , state-reduction parameters  $P = 4$  and  $U = L - P - Q = 4$ . Therefore, the



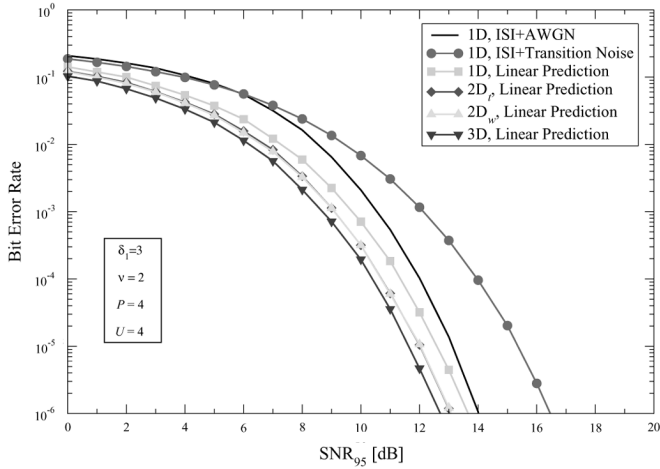


Fig. 5. BER for the Lorentzian magnetic channel model with transition noise at  $D = 2.5$ ,  $\alpha = 95\%$ , and  $\lambda = 50\%$ .

Viterbi algorithm searches a trellis diagram with  $2^{\delta_1 + \nu} = 32$  states. The BER curve obtained with linear prediction and a monodimensional front end shows an SNR gain of almost 2.5 dB, with respect to the one obtained without linear prediction. With a bidimensional detector front end (Linear Prediction 2D curves), the BER outperforms the 1D linear prediction curve by approximately 0.6 dB (note that the  $2D_t$  curve lays upon the  $2D_w$  one), while with the multidimensional detector the gain, with respect to the ISI + Transition Noise curve, is nearly 3.5 dB.

Fig. 6 shows the BER curves for a monodimensional front end, applied to a channel affected by transition noise, without linear prediction (ISI + Transition Noise curve) and with linear prediction (1D Linear Prediction curve), and the performance of a three-dimensional detector with linear prediction (3D Linear Prediction curve), for a density  $D = 2.50$  and a 70% transition noise with  $\lambda = 50\%$ . The simulations have been carried out assuming  $\delta_1 = 3$  precursors, a prediction order  $\nu = 2$ , state-reduction parameters  $P = 4$  and  $U = L - P - Q = 4$ . Therefore, the Viterbi algorithm searches a trellis diagram with 32 states. With respect to Fig. 5, one can observe that: i) the actual “ISI + Transition Noise” curve is moved to the left due to a reduced transition noise power; ii) the “1D Linear Prediction curve” is shifted to the right because the quality of the linear prediction is limited by the increased thermal noise; and iii) the “3D Linear Prediction curve” does not substantially change its position. This finding confirms the robustness of the proposed multidimensional detector, i.e., while for relatively low media noise percentage  $\alpha$  a monodimensional detector has a BER performance limited by a decreased quality in the estimate of the transition noise sample (the prediction coefficients have amplitude that decreases when the power of the white thermal noise increases), the multidimensional detector is able to extract all the information available at the output of the channel and use it for the computation of the prediction coefficients.

Fig. 7 also presents the BER obtained using both the monodimensional and the three-dimensional detector with linear prediction, operating with different values of the predictor order  $\nu$ . The user density is set to  $D = 2.5$  and a 95% transition noise

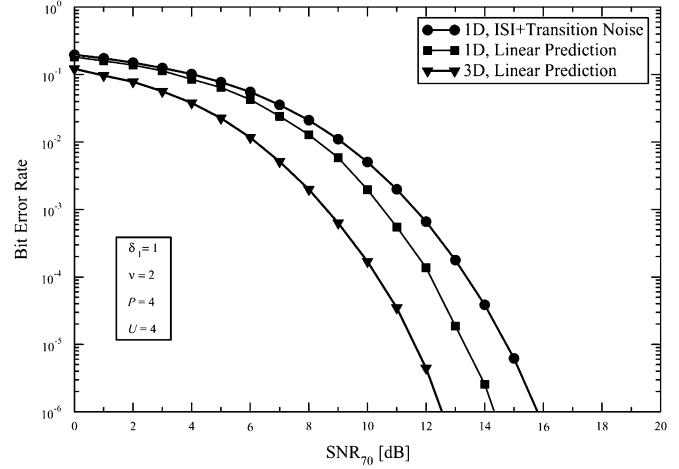


Fig. 6. BER for the Lorentzian magnetic channel model with transition noise at  $D = 2.5$ ,  $\alpha = 70\%$  and  $\lambda = 50\%$ .

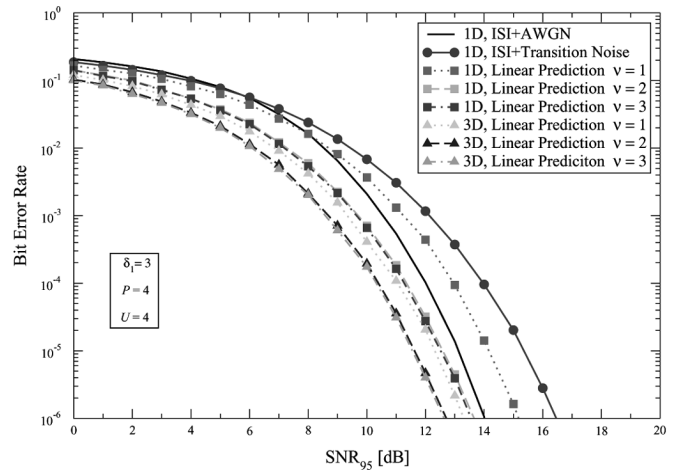


Fig. 7. BER for the Lorentzian channel model with transition noise at  $D = 2.5$ ,  $\alpha = 95\%$ , and  $\lambda = 50\%$  as a function of the predictor order  $\nu$ .

with  $\lambda = 50\%$  is assumed. As already observed, in order to obtain a good quality in the estimation of transition noise, it is possible to operate with a small value of  $\nu$ , i.e.,  $\nu = 2$ .

Fig. 8 shows the BER obtained using the proposed multidimensional detector operating at a user density  $D = 3.25$  and assuming a 95% transition noise with  $\lambda = 50\%$ . These curves are obtained with  $\delta_1 = 3$  precursors,  $\delta_2 = 9$  postcursors, a prediction order  $\nu = 2$ , state-reduction parameters  $P = 5$  and  $U = 5$ . The gain obtained with the multidimensional detector with respect to the ISI + Transition Noise curve is about 5 dB, showing that, as the density increases, the proposed detector can cope very well with an increased transition noise power.

We now extend these results to perpendicular magnetic recording systems. Fig. 9 shows the performance of a monodimensional detector without linear prediction (ISI + Transition Noise curve) and mono (1D) and bidimensional ( $2D_t$ ) detectors operating at user density  $D = 2.50$ , assuming a 95% transition noise consisting of only the position jitter contribution, i.e.,  $\lambda = 100\%$ . In order to obtain these curves, we used a Viterbi processor which searches a trellis diagram with 16 states, with  $\delta_1 = 1$  precursors,  $\delta_2 = 7$  postcursors, a prediction order

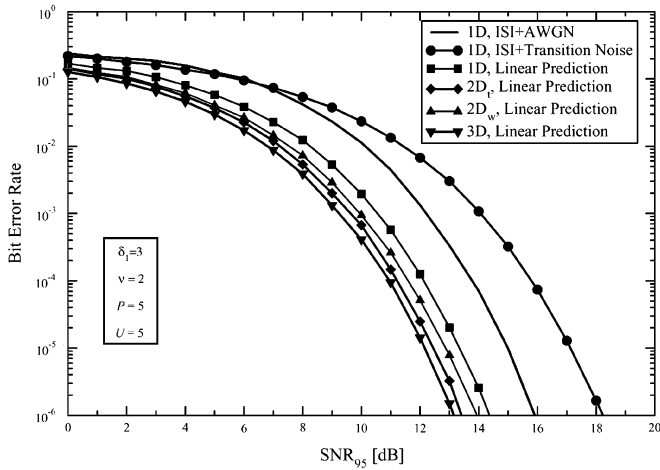


Fig. 8. BER for the Lorentzian channel model with transition noise at  $D = 3.25$ ,  $\alpha = 95\%$ , and  $\lambda = 50\%$ .

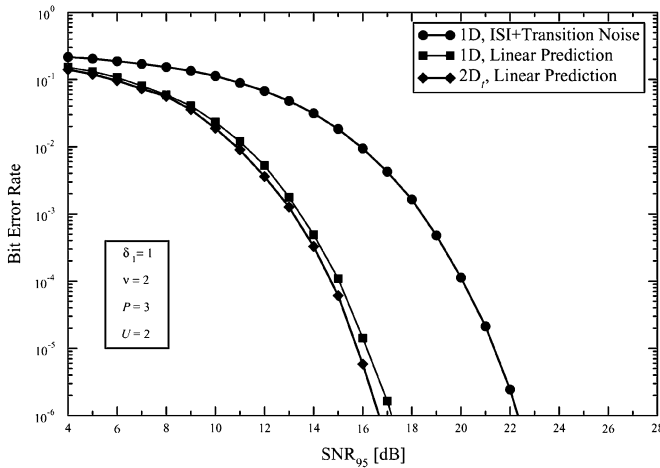


Fig. 9. BER for a perpendicular magnetic channel model with transition noise consisting of position jitter noise only, at  $D = 2.50$ ,  $\alpha = 95\%$ , and  $\lambda = 100\%$ .

$\nu = 2$ , state-reduction parameters  $P = 3$  and  $U = 2$ . Fig. 10 shows similar results for a user density of  $D = 3.25$ . In this case, the used parameters were  $\delta_1 = 1$  precursors,  $\delta_2 = 6$  postcursors, a prediction order  $\nu = 2$ , per-survivor processing with state-reduction parameter  $P = 3$  and  $U = 3$ . The SNR gain obtained by the bidimensional detector with respect to the ISI + Transition Noise curve is about 5 dB in both cases (Figs. 9 and 10). Note that a reduced SNR gain with respect to the 1D curve is now present. However, by increasing the user density  $D$  (compare Figs. 9 and 10), the proposed detector shows a larger improvement.

Finally, we consider optical recording systems affected by transition noise. The origins of the media noise include non-evenness, roughness, and birefringence in the substrate, which are substantially determined by the stamp and the mastering process [40]. As a consequence, although for such kind of systems the nature of transition noise is related to optical phenomena, it is possible to apply the first-order channel model as well. These systems can be modeled by a transition response  $h(t, w)$  similar to the transition response used in perpendicular magnetic recording systems [21]–[24]. As a consequence, it is

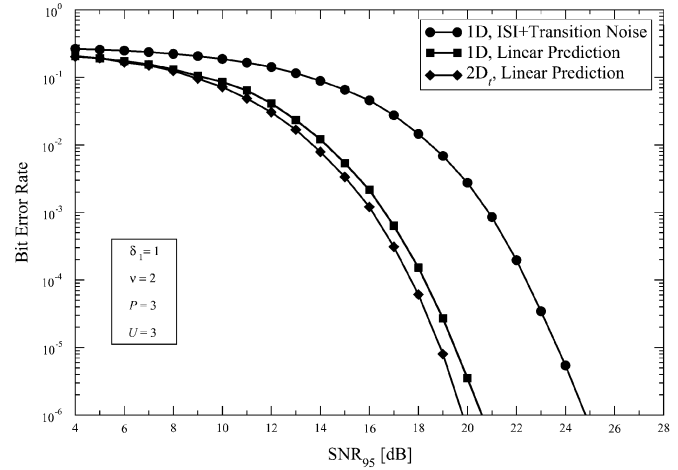


Fig. 10. BER for the perpendicular magnetic channel model with transition noise consisting of position jitter noise only, at  $D = 3.25$ ,  $\alpha = 95\%$ , and  $\lambda = 100\%$ .

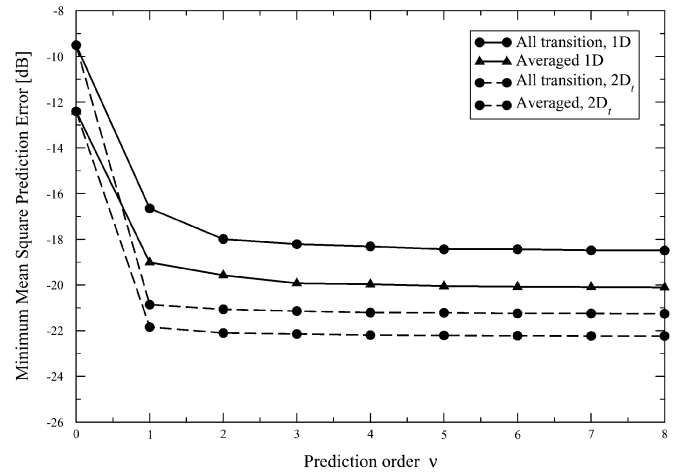


Fig. 11. MMSPE versus number of prediction coefficients  $\nu$  for an optical recording system (DVD) at  $D = 4.6$ ,  $\alpha = 95\%$ ,  $\text{SNR}_{95} = 10$  dB, and  $\lambda = 100\%$ .

possible to extend the proposed multidimensional detector to these storage systems.

In Fig. 11, we present curves of prediction error obtained for a monodimensional front end (indicated by 1D) and for a bidimensional front end ( $2D_t$ ), for an optical recording system with a normalized density  $D = 4.6$  (DVD density). These results are obtained using a transition response  $h(t, w)$  defined as in (2): as a consequence, an optical digital system is equivalent to a perpendicular recording system working at higher user density.

In the bidimensional case, the second front end filter was selected to be matched to the time derivate of the  $\text{erf}(x)$  pulse ( $2D_t$  curves), since we assumed that the 95% transition noise power is due only to position jitter ( $\lambda = 100\%$ ). A bit pattern characterized by continuous changes of the writing current's polarity, i.e., the “all-transition” bit pattern, and an “averaged” bit pattern is considered.

Fig. 12 shows the gain, in terms of SNR, between a monodimensional detector without linear prediction (ISI + Transition Noise curve) and a mono (1D) or a bidimensional ( $2D_t$ ) detector. We used a Viterbi processor searching a trellis diagram

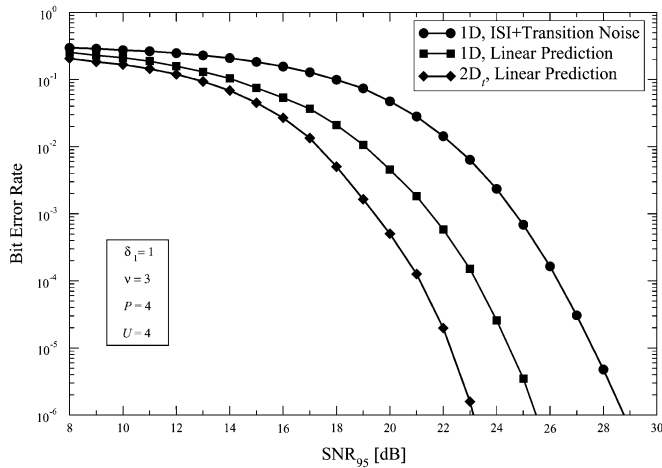


Fig. 12. BER for the optical channel model with transition noise consisting of position jitter noise only ( $\lambda = 100\%$ ), at  $D = 4.6$  and  $\alpha = 95\%$ .

with 32 states, with  $\delta_1 = 1$  precursors,  $\delta_2 = 9$  postcursors, a prediction order  $\nu = 3$ , state-reduction parameters  $P = 4$  and  $U = 4$ . The gain obtained with the multidimensional detector with respect to the ISI + Transition Noise curve is about 6 dB, whereas the gain with respect to the 1D curve is 2.5 dB, confirming that, as the density increases, the proposed detector can cope very well with an increased transition noise power.

## VII. CONCLUSION

A set of sufficient statistics for the digital storage channel subject to data-dependent transition noise has been proposed. These sufficient statistics can be obtained through a multidimensional detector front end. Multidimensional linear prediction can be used to modify the branch metrics of a Viterbi detector in order to improve the performance and make it quite insensitive to transition noise. The gain in terms of BER versus SNR achievable using the proposed detector, with multidimensional front end and linear prediction, increases with the user density  $D$ . The proposed multidimensional signal processing technique was applied to longitudinal and perpendicular magnetic recording systems, as well as optical storage system, such as DVD, demonstrating a good capability of coping with increasing storage density.

## REFERENCES

- [1] T. D. Howell, D. P. McCown, Y.-S. Tang, K. R. Hense, and R. L. Gee, "Error rate performance of experimental gigabit per square inch recording components," *IEEE Trans. Magn.*, vol. 26, no. 5, pp. 2298–2302, Sep. 1990.
- [2] J. Moon, "Discrete-time modeling of transition-noise-dominant channels and study of detection performance," *IEEE Trans. Magn.*, vol. 27, no. 6, pp. 4573–4578, Nov. 1991.
- [3] J. Caroselli and J. W. Wolf, "Application of a new simulation model for media noise limited magnetic recording channels," *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 3917–3919, Sep. 1996.
- [4] A. Kavcic and A. Patapoutian, "A signal-dependent autoregressive channel model," *IEEE Trans. Magn.*, vol. 35, no. 5, pp. 2316–2318, Sep. 1999.
- [5] T. Oenning and J. Moon, "Modeling the Lorentzian magnetic recording channel with transition noise," *IEEE Trans. Magn.*, vol. 37, no. 1, pp. 583–591, Jan. 2001.

- [6] E. Eleftheriou and W. Hirt, "Improving performance of PRML/EPRML through noise prediction," *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 3968–3970, Sep. 1996.
- [7] J. M. Cioffi, W. L. Abbott, H. K. Thapar, C. M. Melas, and K. D. Fisher, "Adaptive equalization in magnetic-disk storage channels," *IEEE Commun. Mag.*, vol. 28, no. 2, pp. 14–29, Feb. 1990.
- [8] R. D. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Schott, "APRML system for digital magnetic recording," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 1, pp. 38–56, Jan. 1992.
- [9] S. A. Altekari and J.K. Wolf, "Improvements in detectors based upon colored noise," *IEEE Trans. Magn.*, vol. 34, no. 1, pp. 94–97, Jan. 1998.
- [10] A. Kavcic and J. F. Moura, "Correlation-sensitive adaptive sequence detection," *IEEE Trans. Magn.*, vol. 34, no. 3, pp. 763–771, May 1998.
- [11] —, "The Viterbi algorithm and Markov noise memory," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 291–301, Jan. 2000.
- [12] J. Moon and J. Park, "Pattern-dependent noise prediction in signal-dependent noise," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 730–743, Apr. 2001.
- [13] J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 38, no. 6, pp. 787–794, June 1990.
- [14] D. Bouras, P. T. Mathiopoulos, and D. Makrakis, "Optimal decoding of coded PSK and QAM signals in correlated fast fading channels and AWGN: A combined envelope, multiple differential and coherent detection approach," *IEEE Trans. Commun.*, vol. 42, no. 1, pp. 63–75, Jan. 1994.
- [15] X. Yu and S. Pasupathy, "Innovations-based MLSE for rayleigh fading channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1534–1544, Feb./Mar./Apr. 1995.
- [16] G. Vitetta and U. Mengali, "Double-filtering receivers for psk signals transmitted over rayleigh frequency-flat fading channels," *IEEE Trans. Commun.*, vol. 44, no. 6, pp. 686–695, June 1996.
- [17] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13–20, Jan. 1988.
- [18] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback dequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.
- [19] P. R. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. Commun.*, vol. 37, no. 7, pp. 669–676, Jul. 1989.
- [20] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 41, no. 2, pp. 354–364, Feb. 1993.
- [21] H. Song, B. V. K. V. Kumar, E. Kurtas, Y. Yuan, L. L. McPheters, and S. W. McLaughlin, "Iterative decoding for partial response (PR), equalized, magneto-optical (MO) data storage channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 774–782, Apr. 2001.
- [22] B. Steingrimsson, J. Moon, and T. Oenning, "Signal space detection for DVD optical recording," *IEEE Trans. Magn.*, vol. 37, no. 2, pp. 670–675, Mar. 2001.
- [23] L. L. McPheters and S. W. McLaughlin, "Turbo-coded optical recording channels with DVD minimum mark size," *IEEE Trans. Magn.*, vol. 38, no. 1, pp. 298–302, Jan. 2002.
- [24] H. Cho, C. Woo, and D. Hong, "Simplified signal space detector for optical recording," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 3566–3569, Sep. 2002.
- [25] R. Pighi, R. Raheli, and U. Amadei, "Multidimensional receiver front-end for storage systems with data-dependent transition noise," in *Proc. IEEE Int. Symp. Inform. Theory and its Appl.*, Parma, Italy, Oct. 2004, pp. 885–890.
- [26] J. Proakis, *Digital Communications*. New York: McGraw-Hill, 2001.
- [27] P. Kovintavewat, I. Ozgunes, E. Kurtas, J. R. Barry, and S. McLaughlin, "Generalized partial-response target for perpendicular recording with jitter noise," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2340–2342, Sep. 2002.
- [28] T. A. Roscamp, E. D. Boerner, and G. J. Parker, "Three-dimensional modeling of perpendicular reading with a soft underlayer," *J. Appl. Phys.*, vol. 91, no. 10, pp. 8366–8368, May 2002.
- [29] J. Moon and T. R. Oenning, "Low density parity check coding for magnetic recording channels with media noise," in *Proc. Int. Conf. Communications*, Jun. 2001, vol. 7, pp. 2189–2193.
- [30] D. Arnold and E. Eleftheriou, "On the information-theoretic capacity of magnetic recording systems in the presence of medium noise," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2319–2321, Sep. 2002.

- [31] Z. Zhang, T. M. Duman, and E. M. Kurtas, "Information rates of binary-input intersymbol interference channel with signal-dependent media noise," *IEEE Trans. Magn.*, vol. 39, no. 1, pp. 599–607, Jan. 2003.
- [32] H. L. Van Trees, *Detection Estimation and Modulation Theory, Part I*. New York: Wiley, 1996.
- [33] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1969.
- [34] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [35] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [36] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.
- [37] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–579, Apr. 1975.
- [38] G. Ferrari, G. Colavolpe, and R. Raheli, *Detection Algorithms for Wireless Communications, With Applications to Wired and Storage Systems*. London, U.K.: Wiley, 2004.
- [39] J. Moon, "Signal-to-noise ratio definition for magnetic recording channels with transition noise," *IEEE Trans. Magn.*, vol. 36, no. 5, pp. 3881–3883, Sep. 2000.
- [40] M. Mansuripur and C. Peng, "Various sources of noise in optical data storage systems," in *Optical Data Storage, Conf. Dig.*, May 2000, pp. 144–146.

Manuscript received February 3, 2005; revised February 22, 2006. Corresponding author: R. Raheli (e-mail: raheli@unipr.it).

**Riccardo Pighi** was born in Piacenza, Italy, in November 1976. He received the Dr. Ing. degree (Laurea) in telecommunication engineering and the Ph.D. degree in information technology from the University of Parma, Parma, Italy, in 2002 and 2006, respectively.

Since 2003, he has been involved in the project of a multicarrier system for Power Line Communication (PLC) in collaboration with Selta S.p.A., Cadeo (PC), Italy. His main research interests are in the area of digital communication systems design, adaptive and multirate signal processing, storage systems, information theory and power line communications.

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

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

He served on the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS as an Editor for Detection, Equalization, and Coding from 1999 to 2003. He also served as a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, Special Issue on Differential and Noncoherent Wireless Communications, published in September 2005. Since 2003, he has been on the Editorial Board of the EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS as an Editor for Communication Theory.

**Umberto Amadei** was born in Parma, Italy, in August 1978. He received the Dr. Ing. degree (Laurea) in telecommunications engineering from the University of Parma in 2004.

From January 2005 to November 2005, he was with the Dipartimento di Ingegneria dell'Informazione, University of Parma. Since December 2005, he has been with Selta S.p.A., Cadeo, Piacenza, Italy. His research interests include advanced signal processing techniques for storage systems, power line communications, and digital communication systems design.