Information Rates of Multidimensional Front-Ends for Digital Storage Channels With Data-Dependent Transition Noise

Riccardo Pighi
Dept. of Information Engineering
University of Parma
43100 Parma, Italy
Email: pighi@tlc.unipr.it

Riccardo Raheli
Dept. of Information Engineering
University of Parma
43100 Parma, Italy
Email: raheli@unipr.it

Fabio Cappelletti
Dept. of Information Engineering
University of Parma
43100 Parma, Italy
Email: cappelletti@tlc.unipr.it

***Abstract*— A new simulation-based method to evaluate the information rates of multidimensional front-ends applied to digital storage channels with transition noise is presented. First, we propose an algorithm which extends recent work on the information rates of magnetic recording channels affected by colored Gaussian thermal noise, intersymbol interference and signal-dependent transition noise, by using linear prediction and state reduction techniques. Moreover, following a previous study on statistical sufficiency, we extend this algorithm to magnetic channels with a multidimensional front-end. The results suggest that significant gains may be achievable by multidimensional signal processing techniques in transition-noise limited digital storage channels.**

I. Introduction

The computation of the capacity of a magnetic recording channel has been an interesting challenge and open problem for some time. Storage systems, such as magnetic or optical recording channels, are essentially communication systems characterized by a great amount of intersymbol interference (ISI), colored Gaussian thermal noise and a kind of noise induced by the interaction between transitions in the information sequence stored on the medium. This kind of noise, also known as media noise or transition noise, increases with storage density and can be modeled as data-dependent. In [1], a simulation based method to compute the information rates of ISI channels with additive white Gaussian thermal noise (AWGN) is presented. In [2], this algorithm is extended to a channel with colored thermal noise and transition noise. The impact of transition noise on the information rates of magnetic longitudinal recording channels is also studied in [3] and [4].

In this paper, we first extend the algorithm in [2]. In order to evaluate the information rates of storage systems, e.g., of longitudinal and perpendicular magnetic recording channels, we propose a new method based on linear prediction and reduced-state techniques. In [5], it was shown that the presence of transition noise yields a multidimensional front-end with a number of filters proportional to the degree of description of the transition noise process. Assuming that transition noise can be characterized only by a jitter noise term, the need for sufficient statistics yields a front-end based on two matched filters only. For this channel model, the linear-prediction method is extended to a bidimensional scenario. The achievable information rates with a bidimensional front-end for longitudinal and perpendicular recording systems are investigated.

II. System Model

In order to describe the proposed algorithm, we consider a magnetic recording channel modeled by a first-order position jitter and width variation [6], but our results can also be extended to optical and magneto-optical recording systems affected by transition noise. Moreover, the proposed algorithm can be applied to a high-order channel model 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 pulse width. 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 [6], 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

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

where $b_k = a_k - a_{k-1} \in \{0, \pm 2\}$ denote transition symbols, Δt_k and Δw_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 recording channel without transition noise. For the pulse response $h(t, w)$ we adopt the well known Lorentzian approximation [7] for longitudinal magnetic 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

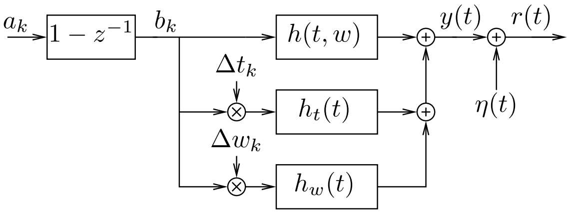

Fig. 1. Channel model with first-order media noise and additive white Gaussian thermal noise.

approximation [8], [9]

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

where PW_{50} is defined 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 and assume $T = 1$. According to the first-order channel model, the read back impulse can be approximated as

$$\begin{aligned} &h(t+\Delta t_k, w+\Delta w_k) \\ &\quad \simeq h(t,w) + \Delta t_k \frac{\partial h(t,w)}{\partial t} + \Delta w_k \frac{\partial h(t,w)}{\partial w}\,. \end{aligned} \tag{2}$$

Defining the impulse response of the filters modeling the position jitter and width variation noise processes as[1]

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

and using this first-order approximation (2) in (1), the waveform at the output of the channel due to transition noise can be written as

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

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

III. Information Rates of ISI Channels with Transition Noise and Linear Prediction

In [1], the information rates of i.i.d. binary input intersymbol interference channel with additive white Gaussian thermal noise were computed by the forward recursion of a BCJR algorithm [10]. In [2], the same simulation method was applied to an ISI channel with colored Gaussian thermal noise and also signal-dependent transition noise. In this section, we begin by reviewing this method and introducing the used notation. We then extend the method applying linear prediction and state reduction techniques. Let us focus on a longitudinal magnetic channel model (similar considerations can be applied to perpendicular magnetic or optical recording): considering a front-end based on a matched filter $h(-t,w)$ and a sampler at symbol rate, the discrete time channel output can be written as

$$x_k = \sum_{i=-L_1}^{L_2} b_{k-i} g_k + m_k + n_k \tag{3}$$

where m_k is a transition noise sample, n_k is a sample of colored Gaussian thermal noise, $g_k = h(t,w) * h(-t,w)|_{t=kT}$ and L_1 and L_2 are the number of precursors and postcursors of the discrete time channel model.

The mutual information between the input information sequence and the discrete time channel output is given by [2], [11]

$$I(A;X) = h(X) - h(X|A) \tag{4}$$

where $h(X)$ and $h(X|A)$ are differential entropies of the r.v. a_k and x_k, respectively. In an AWGN scenario, the differential entropy $h(X|A) = (1/2)\log(2\pi e\,\sigma_w^2)$ is well known. Due to the presence of transition noise, however, $h(X|A)$ must be evaluated explicitly. Since $\{x_k\}$ is a stationary ergodic hidden-Markov process, the Shannon-McMillan-Breiman theorem holds [11] and it is possible to evaluate $h(X)$ and $h(X|A)$ using a reasonably long information sequence by means of simulation. In order to compute $h(X)$ and $h(X|A)$ in (4), we can express these two terms as

$$\begin{aligned} h(X) &= \lim_{n\to\infty} \frac{1}{n} h(\mathbf{x}_0^n) = -\lim_{n\to\infty} \frac{1}{n} E\left[\log(p(\mathbf{x}_0^n))\right] \\ &= -\lim_{n\to\infty} \frac{1}{n} \log\left[p(\mathbf{x}_0^n)\right] \\ h(X|A) &= \lim_{n\to\infty} \frac{1}{n} h(\mathbf{x}_0^n|\mathbf{a}_0^n) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{\tilde{\mathbf{a}}_0^n \in \Lambda^n} h(\mathbf{x}_0^n|\mathbf{a}_0^n = \tilde{\mathbf{a}}_0^n) \cdot P(\mathbf{a}_0^n = \tilde{\mathbf{a}}_0^n) \end{aligned}$$

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, Λ^n is the set of all possible input sequence of length $n+1$, $p(\mathbf{x}_0^n)$ is the probability density function (pdf) of the observation sequence $\mathbf{x}_0^n$ and $P(\mathbf{a}_0^n = \tilde{\mathbf{a}}_0^n)$ is the a priori probability of the sequence $\tilde{\mathbf{a}}_0^n$.

Let us define ζ_k and ζ_{k-1} as the current trellis state and the previous trellis state, at time epoch k and $k-1$, respectively. With these definitions, we are able to evaluate the pdf of a sequence of length $k+1$ as

$$p(\boldsymbol{x}_0^k) = \sum_{\zeta_k} p(\mathbf{x}_0^k|\zeta_k) P(\zeta_k) = \sum_{\zeta_k} \mu(\zeta_k)$$

where we have defined the "metric" $\mu(\zeta_k)$ as[2]

$$\mu(\zeta_k) = p(\mathbf{x}_0^k|\zeta_k) P(\zeta_k)\,. \tag{5}$$

Marginalizing over the previous states and using the chain factorization rule, we can rewrite (5) as

$$\begin{aligned} \mu(\zeta_k) &= p(\mathbf{x}_0^k|\zeta_k)P(\zeta_k) = \sum_{\zeta_{k-1}} p(\boldsymbol{x}_0^k|\zeta_k,\zeta_{k-1}) P(\zeta_k,\zeta_{k-1}) \\ &= \sum_{\zeta_{k-1}} p(x_k|\boldsymbol{x}_0^{k-1},\zeta_k,\zeta_{k-1}) p(\boldsymbol{x}_0^{k-1}|\zeta_k,\zeta_{k-1}) P(\zeta_k,\zeta_{k-1}) \end{aligned}$$

[1]The subscript denotes the variable of differentiation.

[2]In a mathematical sense, $\log \mu(\zeta_k)$ is a metric.

$$= \sum_{\zeta_{k-1}} p(x_k|\boldsymbol{x}_0^{k-1}, \zeta_k, \zeta_{k-1}) P(\zeta_k|\boldsymbol{x}_0^{k-1}, \zeta_{k-1}) \cdot \underbrace{p(\boldsymbol{x}_0^{k-1}|\zeta_{k-1}) P(\zeta_{k-1})}_{\mu(\zeta_{k-1})}$$

$$= \sum_{\zeta_{k-1}} p(x_k|\boldsymbol{x}_0^{k-1}, \zeta_k, \zeta_{k-1}) P(\zeta_k|\zeta_{k-1}) \mu(\zeta_{k-1}) \quad (6)$$

where we have introduced $\mu(\zeta_{k-1}) = p(\boldsymbol{x}_0^{k-1}|\zeta_{k-1})P(\zeta_{k-1})$ and assumed the probability of the transition from state ζ_{k-1} to state ζ_k, i.e., $P(\zeta_k|\zeta_{k-1})$, independent from previous output samples.

In order to limit the channel memory, we assume Markovianity of order ν in the observation sequence: as a consequence, we can express the conditional pdf as

$$p(x_k|\boldsymbol{x}_0^{k-1}, \zeta_k, \zeta_{k-1}) \approx p(x_k|\boldsymbol{x}_{k-\nu}^{k-1}, \zeta_k, \zeta_{k-1}) \,. \quad (7)$$

This pdf is completely specified by the conditional mean and variance

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

Keeping in mind that the observation sequence is conditionally Gaussian, given the data, $\hat{x}_k$ can be interpreted as a linear predictive estimate of x_k and $\hat{\sigma}_{x_k}^2$ as the relevant Minimum Mean Square Prediction Error (MMSPE) [12]. The linear prediction approach is computationally more efficient and perfectly equivalent to the method used in [2] for the evaluation of (7).

Since the prediction coefficients are data-dependent [5], it is necessary to extend the state definition ζ_k, at time epoch k, including not only precursors and postcursors of the information bearing signal but also of the transition noise signals, i.e.,

$$\zeta_k = (a_{k+\delta_1}, a_{k+\delta_1-1}, \ldots, a_k, a_{k-1}, \ldots, a_{k-\delta_2-\nu+1})$$

where $\delta_1 \geq L_1$ and $\delta_2 \geq L_2$ are the maximum numbers of precursors and postcursors of all the discrete time signals presented in the channel model. As in [2], the Markovianity assumption implies the computation of an upper bound on $h(X)$ and $h(X|A)$. As a consequence, the information rate in (4) is not a bound, in a strict sense, but only a good estimate. The computation of a lower bound to complement and confirm these results is currently under investigation [13][3].

The state-complexity of a linear prediction algorithm can be naturally decoupled from the prediction order ν by means of state-reduction techniques [14]–[16]. Let Q denote the memory parameter to be taken into account in the definition of a "reduced" trellis state

$$\omega_k = (a_{k+\delta_1}, a_{k+\delta_1-1}, \ldots, a_k, \ldots, a_{k+\delta_1-Q+1}) \,.$$

[3]This reference was brought to our attention by an anonymous reviewer.

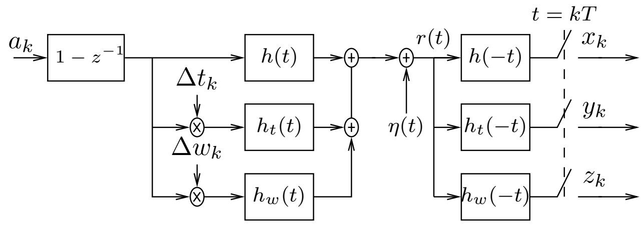

Fig. 2. Magnetic channel model with transition noise and the multidimensional front-end for sufficient statistics.

Metric (6) can be obtained by defining a "pseudo state" [17]

$$\tilde{\zeta}(\omega_k) = \Big(\underbrace{\overbrace{a_{k+\delta_1}, \ldots, a_{k+\delta_1-Q+1}}^{\omega_k}}_{Q \text{ bits}}, \underbrace{\breve{a}_{k+\delta_1-Q}(\omega_k), \ldots, \breve{a}_{k-\delta_2-\nu+1}(\omega_k)}_{\delta_1+\delta_2+\nu-Q=P \text{ bits}} \Big) \quad (8)$$

where $\breve{a}_{k+\delta_1-Q}(\omega_k), \ldots, \breve{a}_{k-\delta_2-\nu+1}(\omega_k)$ are the information bits associated with the survivor of ω_k, according to a per-survivor processing technique [18]. Finally, the metric $\tilde{\mu}(\omega_k)$ in the reduced-state trellis can be defined in terms of the pseudo state (8) as

$$\tilde{\mu}(\omega_k) = \mu(\tilde{\zeta}(\omega_k)) \,.$$

Note that this approach is related to the state reduction techniques presented in [13].

IV. Information Rates of Multidimensional Front-Ends for Magnetic Recording Channels With Data-Dependent Transition Noise

In [5], the authors have shown that in the presence of transition noise, the need for statistical sufficiency yield a front-end with a number of filters proportional to the modeling order of transition noise. Considering the first-order media noise model, a multidimensional front-end[4] is shown in Fig. 2. It is now possible to evaluate the information rate for this system, extending the algorithm described in Sec. III to a multidimensional scenario. In order to make the presentation simple, we will focus on a bidimensional front-end, assuming that transition noise can be characterized by the jitter noise term only.

The information rate for a bidimensional system can be defined as [11]

$$I(A; X, Y) = h(X, Y) - h(X, Y|A).$$

Since $\{y_k\}$, whose expression is similar to (3), is also a stationary ergodic hidden-Markov process, the Shannon-McMillan-Breiman theorem still holds: therefore we can evaluate, by simulation, the joint differential entropy on the basis of

$$h(X, Y) = - \lim_{n \to +\infty} \frac{1}{n} \log[p(\boldsymbol{x}_0^n, \boldsymbol{y}_0^n)] \,.$$

[4]For simplicity, we omit the parameter w in the impulse response of the channel.

Similarly to the monodimensional scenario, the bidimensional pdf can be expressed as

$$p(\boldsymbol{x}_0^n, \boldsymbol{y}_0^n) = \sum_{\zeta_k} \mu(\zeta_k)$$

where, we now define

$$\mu(\zeta_k) = p(\boldsymbol{x}_0^n, \boldsymbol{y}_0^n | \zeta_k) P(\zeta_k) \,. \tag{9}$$

Marginalizing over the previous states and using chain factorization, we can express (9), at time k, as

$$\begin{aligned}
\mu(\zeta_k) \;=\; & p(\boldsymbol{x}_0^k, \boldsymbol{y}_0^k | \zeta_k) P(\zeta_k) \\
=& \sum_{\zeta_{k-1}} p(\boldsymbol{x}_0^k, \boldsymbol{y}_0^k | \zeta_k, \zeta_{k-1}) P(\zeta_k, \zeta_{k-1}) \\
=& \sum_{\zeta_{k-1}} p(x_k | \boldsymbol{x}_0^{k-1}, \boldsymbol{y}_0^k, \zeta_k, \zeta_{k-1}) p(y_k | \boldsymbol{x}_0^{k-1}, \boldsymbol{y}_0^{k-1}, \zeta_k, \zeta_{k-1}) \\
& \cdot p(\boldsymbol{x}_0^{k-1}, \boldsymbol{y}_0^{k-1} | \zeta_k, \zeta_{k-1}) P(\zeta_k, \zeta_{k-1}) \\
=& \sum_{\zeta_{k-1}} p(x_k | \boldsymbol{x}_0^{k-1}, \boldsymbol{y}_0^k, \zeta_k, \zeta_{k-1}) p(y_k | \boldsymbol{x}_0^{k-1}, \boldsymbol{y}_0^{k-1}, \zeta_k, \zeta_{k-1}) \\
& \cdot \underbrace{P(\zeta_k | \boldsymbol{x}_0^{k-1}, \boldsymbol{y}_0^{k-1}, \zeta_{k-1})}_{P(\zeta_k|\zeta_{k-1})} \underbrace{p(\boldsymbol{x}_0^{k-1}, \boldsymbol{y}_0^{k-1} | \zeta_{k-1}) P(\zeta_{k-1})}_{\mu(\zeta_{k-1})} \\
\approx& \sum_{\zeta_{k-1}} p(x_k | \boldsymbol{x}_{k-\nu}^{k-1}, \boldsymbol{y}_{k-\nu}^k, \zeta_k, \zeta_{k-1}) P(\zeta_k | \zeta_{k-1}) \mu(\zeta_{k-1}) \\
& \cdot p(y_k | \boldsymbol{x}_{k-\nu}^{k-1}, \boldsymbol{y}_{k-\nu}^{k-1}, \zeta_k, \zeta_{k-1})
\end{aligned} \tag{10}$$

where, in the last step, we have assumed Markovianity of order ν in both observation sequences and defined the conditional means and variances of the pdfs in (10) as

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

The quantities $\hat{x}_k$, $\hat{y}_k$, $\hat{\sigma}_{x_k}^2$ and $\hat{\sigma}_{y_k}^2$ can be interpreted as linear predictive estimates of x_k, y_k and the relevant MMSPEs, respectively.

Similarly, it is possible to apply multidimensional linear prediction in order to estimate $h(X, Y|A)$. Accordingly, we have

$$H(X, Y|A) = -\lim_{n\to\infty} \frac{1}{n} \log_2[p(\boldsymbol{u}_0^n, \boldsymbol{v}_0^n | \boldsymbol{a}_0^n)] \tag{11}$$

where $\boldsymbol{u}_0^n$ and $\boldsymbol{v}_0^n$ denote the overall noise sequences at the output of the bidimensional front-end (first and second branch, respectively). In particular, with the knowledge of the input sequence, we can estimate the probability in (11) using, at time k, Markovianity of order ν and chain factorization, i.e.

$$\begin{aligned}
p(\boldsymbol{u}_0^k, \boldsymbol{v}_0^k | \boldsymbol{a}_0^k) \;\approx\; & p(\boldsymbol{u}_{k-\nu}^k, \boldsymbol{v}_{k-\nu}^k | \boldsymbol{a}_0^k) \\
=\; & p(u_k | \boldsymbol{u}_{k-\nu}^{k-1}, \boldsymbol{v}_{k-\nu}^k, \boldsymbol{a}_0^k) p(v_k | \boldsymbol{u}_{k-\nu}^{k-1}, \boldsymbol{v}_{k-\nu}^{k-1}, \boldsymbol{a}_0^k) \\
& \cdot \underbrace{p(\boldsymbol{u}_{k-\nu-1}^{k-1}, \boldsymbol{v}_{k-\nu-1}^{k-1} | \boldsymbol{a}_0^{k-1})}_{(k-1)\text{-st step}} \,.
\end{aligned}$$

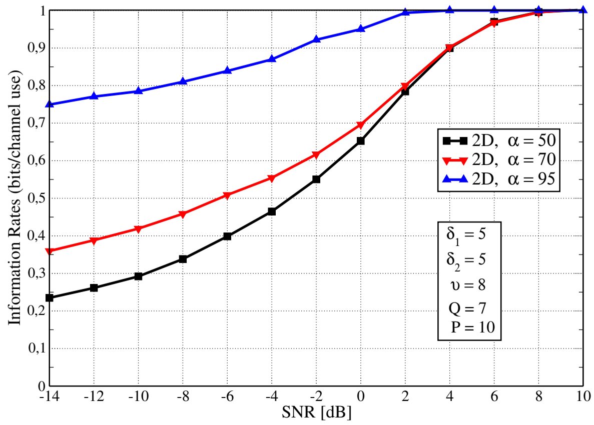

Fig. 3. Information rates for longitudinal bidimensional magnetic recording channel with different values of α and $D = 2.0$.

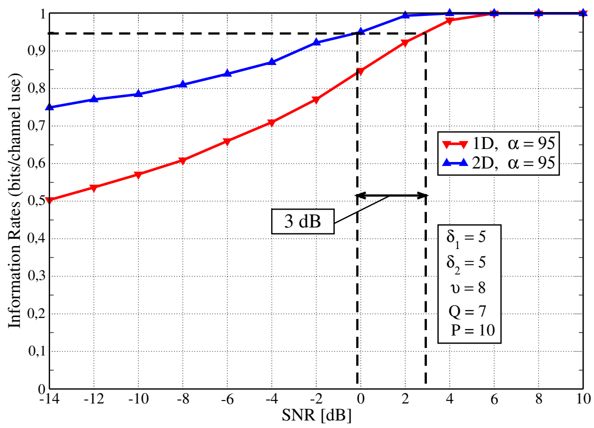

Fig. 4. Information rates for longitudinal magnetic recording channel using a monodimensional (1D) or bidimensional (2D) front-end, $\alpha = 95$, $D = 2.0$.

The state definition must now include the precursors and postcursors of the discrete time signals present in the bidimensional channel model: finally, in order to limit the numerical complexity of the algorithm, reduced-state techniques can be used as in the monodimensional scenario.

V. Numerical Results

Fig. 3 shows the binary-i.i.d.-input information rates obtained with a bidimensional front-end for a transition noise percentage [19] α equal to $50, 70$ and 95. The signal-to-noise-ratio is defined as in [5], [19]. Fig. 4 presents a comparison between the information rates achievable using a monodimensional (1D curve) or bidimensional front-end (2D curve). These curves are obtained with a state complexity of 2^Q, with $Q = 7$, $P = 10$ and a prediction order $\nu = 8$. Assuming that an error correcting code with rate $16/17$ [2] is used, the gain in terms of SNR needed for an information rate of $16/17$ bits per channel use is nearly 3 dB, showing that with a bidimensional front-end we are able to extract more information from the output of the bidimensional channel and use it in order to increase the system reliability.

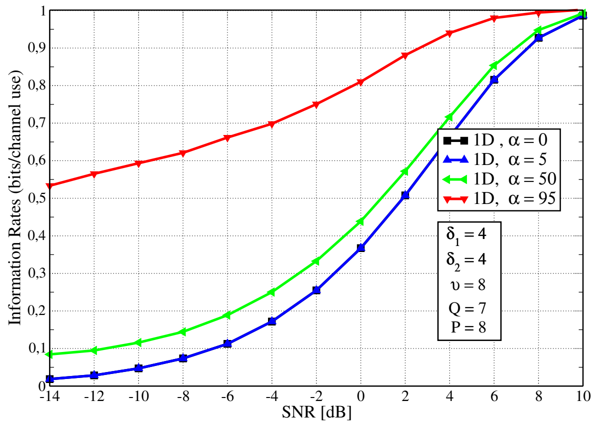

Fig. 5. Information rates for perpendicular magnetic recording channel using a monodimensional front-end and $D = 1.50$.

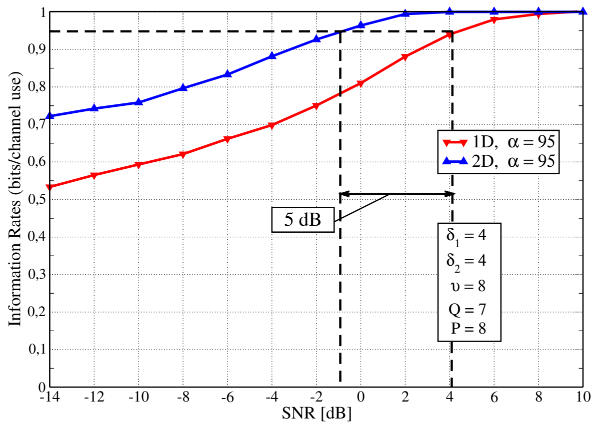

Fig. 6. Information rates for perpendicular magnetic recording channel using a monodimensional (1D) or bidimensional (2D) front-end, $\alpha = 95$, $D = 1.50$.

Fig. 5 presents information rates for a perpendicular magnetic recording channel using a monodimensional front-end. As in the longitudinal channel, the signal-dependent transition noise provides more information about the stored signal compared with the white noise case, i.e., from an information rate viewpoint, transition noise is preferable.

Finally, Fig. 6 shows a comparison between the information rates achievable for a perpendicular recording system using a monodimensional (1D curve) or bidimensional front-end (2D curve), both at density $D = 1.50$ and $\alpha = 95$. These curves are obtained with a state complexity of 2^Q, with $Q = 7$, $P = 8$ and a prediction order $\nu = 8$. Assuming, as in the longitudinal channel, that a code rate of $16/17$ is used in a perpendicular recording system, the gain in terms of SNR is nearly 5 dB.

VI. Conclusion

Using linear prediction and state reduction techniques, we have extended the method in [2] for the computation of the information rates of binary-input ISI channels with signal dependent noise. This method has been applied to the computation of the information rates of magnetic recording channels. Numerical results have shown that the information rates of longitudinal and perpendicular magnetic recording channels can be significantly increased by multidimensional signal processing and detection techniques, with respect to the rates achievable in conventional monodimensional systems. This conclusion may be viewed as the information theoretic counterpart of recent results in [5] and suggests that multi-dimensional signal processing may be an effective approach to enable a steady increase in the capacity of digital storage devices.

References

[1] D. Arnold and H. A. Loeliger, “On the Information Rate of Binary-Input Channels with Memory,” in *Proc. Int. Conf. Communications*, Vol. 9, June 2001, pp. 2692–2695.

[2] Z. Zhang, T. M. Duman, and E. M. Kurtas, “Information Rates of Binary-Input Intersymbol Interference Channel with Signal-Dependent Media Noise,” *IEEE Trans. Magnetics*, vol. 39, no. 1, pp. 599–607, January 2003.

[3] T. Oenning and J. Moon, “The Effect of Jitter Noise on Binary Input Interference Channel Capacity,” in *Proc. Int. Conf. Communications*, Helsinki,Finland, June 2001, pp. 2416–2420.

[4] Z. N. Wu, S. Lin, and J. M. Cioffi, “Numerical Results on Capacity Bounds for Magnetic Recording Channels,” in *Proc. Globecom Conf.*, Vol. 6, 1998, pp. 3385–3390.

[5] 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, October 2004, pp. 885–890.

[6] T. Oenning and J. Moon, “Modeling the Lorentzian Mmagnetic Rrecording Channel With Transition Noise,” *IEEE Trans. Magnetics*, vol. 37, no. 1, pp. 583–591, January 2001.

[7] J. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[8] P. Kovintavewat, I. Ozgunes, E. Kurtas, J. R. Barry, and S. McLaughlin, “Generalized Partial-Response Target for Perpendicular Recording With Jitter Noise,” *IEEE Trans. Magnetics*, vol. 38, no. 5, pp. 2340–2342, September 2002.

[9] 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.

[10] L. R. Bahl, J. Coke, F. Jelinek, and J. Raviv, “Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate,” *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, March 1974.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[12] S. Haykin, *Adaptive filter theory*. Prentice-Hall International Editions, 1991.

[13] D. Arnold, A. Kavcic, H. A. Loeliger, P. O. Vontobel, and W. Zeng, “Simulation-Based Computation on Information Rates: Upper and Lower Bounds,” in *Proc. IEEE Symposium on Information Theory*, Yokohama, Japan, June 2003, p. 119.

[14] 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, January 1988.

[15] A. Duel-Hallen and C. Heegard, “Delayed Decision-Feedback Sequence estimation,” *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.

[16] 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, July 1989.

[17] G. Ferrari, G. Colavolpe, and R. Raheli, *Detection algorithms for wireless communications, with applications to wired and storage systems*. London: John Wiley & Sons, 2004.

[18] 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, February 1995.

[19] J. Moon, “Signal-to-Noise Ratio Definition for Magnetic Recording Channels with Transition Noise,” *IEEE Trans. Magnetics*, vol. 36, no. 5, pp. 3881–3883, September 2000.