RZ 3410 (# 93561) 04/22/02
Computer Science 43 pages

Research Report

Computing Information Rates of Magnetic Recording Systems with Media Noise

Dieter Arnold

IBM Research
Zurich Research Laboratory
8803 Rüschlikon
Switzerland
arn@zurich.ibm.com

Computing Information Rates of Magnetic Recording Systems with Media Noise

Dieter Arnold

IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

Abstract

The compound behavior of the magnetic recording channel is modelled by combining a media noise model, the Lorentzian read-back pulse, and additive white Gaussian noise (AWGN). The media noise model is used to imitate the random zig-zag transition effects that take place when storing a change in magnetic flux on thin film media. The Lorentzian pulse models the frequency-dispersive nature of the read-back head, and the AWGN models electronics noise. By noting that at the output of the magnetic recording channel the read-back signal is cyclostationary, the average autocorrelation function and corresponding power spectral density over one period are computed. The average power spectral density is then used to characterize achievable information rates of the magnetic recording channel for various linear-density and media noise scenarios by using the conjectured Shamai–Laroia bound.

Contents

List of Figures

List of Tables

1 Introduction

It is expected that in future high-density magnetic recording devices media noise will be one of the limiting factors in achieving ultra-high areal densities [1]-[4]. The main sources of media noise at such high densities are pulse jitter and partial signal erasure [5, 6]. We will focus on these two noise sources, and call them "media noise" although they only occur at the transitions and are therefore sometimes termed "transition noise" in the literature.

There are two main difficulties associated with determining the effects of media noise on the information-theoretic capacity of magnetic recording systems. First, even in the absence of media noise, the capacity of magnetic recording channels is unknown. Second, it is difficult to derive a simple channel model that combines the effects of media noise, intersymbol interference (ISI), and electronics noise at high linear densities.

French and Wolf computed upper and conjectured lower bounds on the capacity for the magnetic recording channel for various noise scenarios (including media noise) by assuming Gaussian inputs and physically motivated channel models [7]. However, the Gaussian assumption fails — in particular at high rates, where our main interest resides. Moreover, the channel models used are not easily describable and are difficult to use for signal processing such as coding.

Recently, the effect of pulse jitter on the capacity of binary input ISI channels was investigated with the Arimoto–Blahut algorithm in [8]. To this end, the channel output was quantized to three levels, resulting in considerable quantization loss even for the additive white Gaussian noise (AWGN) channel without memory and jitter.

The purpose of this report is to present a simple information-theoretic approach to study the limiting effects of media noise. It is based on a model for media noise and the conjectured Shamai–Laroia lower bound (SLLB) on the capacity of ISI channels [9]. Three different media noise models are considered, namely: the microtrack model [10], the 2nd-order model (a variation of the microtrack model), and the Nair–Moon Model [11]. The SLLB was applied in [12, 13] to magnetic recording systems in which partial-response (PR) polynomials without media noise were assumed as underlying channel model. In similar spirit, the SLLB is applied here to the compound magnetic recording channel including media noise. The pivotal observation behind this approach is that the read-back signal is cyclostationary at the output of the magnetic recording channel. This allows us to derive the average power spectral density, from which in turn the achievable information rates are computed by means of the SLLB.

The motivation for this report is the question whether for a fixed noise power media noise is preferable to AWGN from an information-theoretic viewpoint. Our approach to this question is first to derive a stationary channel model, secondly to compute achievable information rates with the assumption of Gaussian inputs, and thirdly to translate these numerical results to binary inputs by means of the SLLB. Because media noise is, in contrast to AWGN, shaped like the channel, we found that for the same noise power media noise leads to higher information rates in certain cases.

The original contributions of this report are:

- A qualitative and quantitative comparison of three commonly used models for media noise.
- Derivation of the average power spectral density of the microtrack model. Because of the generality of the microtrack model, this average power spectral density can be adapted to the 2nd-order and the Nair–Moon model, thereby providing a unified framework to study media noise.
- Extension of the SLLB to Markov input processes.
- Numerical results of information rates with a 0.5-Bernoulli and optimized memory-one Markov input process for various noise scenarios.

The structure of this report is as follows: In Section II different models for media noise are presented. In Section III, the average power spectral density is described and discussed. Section IV is devoted to a description of the Shamai–Laroia bound. Numerical results are provided in Section V, and final remarks are given in Section VI.

2 Models for the magnetic recording channel with media noise

Three models are considered for modelling media noise in magnetic recording channels: the microtrack model, the 2nd-order model, and the Nair–Moon model. Of these three models, the microtrack model is the one most detailed and most complex. The Nair–Moon model is the crudest and simplest. The 2nd-order model is a "derivative" of the microtrack model, and conceptually lies in between the microtrack and the Nair–Moon model. All of these models are well known in the literature. We describe the microtrack model in detail, and discuss what is referred to as *media noise* before presenting the other two models.

This Section is self contained, but it is assumed that the reader is familiar with the write and read process. Descriptions of the write and read process from a signal processing perspective can be found for instance in [10] and [14]. The underlying physics is explained in [15].

2.1 Microtrack model

In a first approximation, the magnetic recording channel is well modelled by a linear superposition of alternating Lorentzian pulses. Each Lorentzian pulse

$$h(t) = \lambda \frac{1}{1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2} \tag{1}$$

is the isolated step response to an abrupt change, i.e. a transition, of the magnetic flux from -1 to $+1$[1] (or from $+1$ to -1, where we have $-h(t)$). A normalizing factor λ is typically

[1]The values of -1 and $+1$ can also be thought of as left and right polarization of the magnetic material.

introduced such that the energy of the Lorentzian equals E_{L}, i.e. $\int_{-\infty}^{+\infty} |h(t)|^2 dt = E_{\mathrm{L}}$. We therefore have

$$h(t) = \sqrt{\frac{4E_{\mathrm{L}}}{\pi \mathrm{PW50}}} \frac{1}{1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2}, \tag{2}$$

where we set $E_{\mathrm{L}} = 1$ in what follows.

The Lorentzian pulse models the frequency-dispersive nature of the read-back signal, and it depends on a single parameter called the pulse width at 50% amplitude or PW50. The ratio PW50/T, where T is the data rate, is a measure of the normalized linear density in a hard-disk system. A small PW50/T causes less dispersion and therefore less ISI. It can be achieved, for a given rotation speed, either by a high-quality read head (i.e. small PW50) or alternatively by a low data rate (i.e. a large T).

The medium magnetization results from applying the head field during the write process. The magnetization changes as a function of the location x along the track, and does not change abruptly in response to the applied field. Thus, the transition is assigned a certain width. The widely used Williams–Comstock model for saturation recording [16] defines an *tanh*-like magnetization $m(x)$ according to

$$m(x) = M_r \tanh \frac{2x}{a\pi}, \tag{3}$$

where a denotes the so-called transition-width parameter a. For a position x far away from the transition, we experience the magnetization of the remanent state, i.e.

$$m(x \to \infty) = -m(x \to -\infty) = M_r. \tag{4}$$

The transition-width parameter a is a measure of the transition width. This can be seen from

$$\frac{d}{dx} m(x)|_{x=0} = \frac{2M_r}{\pi a}. \tag{5}$$

The average cross-track magnetization profile is a cumulative distribution function (cdf) that indicates the average polarization at location x for a single transition. It is obtained from $m(x)$ through shifting, i.e.

$$M(x) = \frac{1}{2}(1 + m(x)). \tag{6}$$

The jitter probability density function (pdf) is given by the derivative of the average cross-track magnetization profile. For an *tanh*-shaped average cross-track magnetization profile, the jitter pdf is [10]

$$p_{\mathrm{J}}(j) = \frac{M_r}{\pi a} \mathrm{sech}^2\left(\frac{2j}{\pi a}\right). \tag{7}$$

The characteristic function or Fourier transform of $p_{\mathrm{J}}(j)$ is then expressed in terms of hyperbolic functions as

$$P_{\mathrm{J}}(f) = \frac{a\pi^2(2\pi f)}{8} \mathrm{csch}\left[\frac{a\pi^2}{8}(2\pi f)\right] \mathrm{sech}\left[\frac{a\pi^2}{8}(2\pi f)\right].$$

To gain more insight from the following calculations, we approximate[2] the squared *sech*-pdf of the jitter by a Gaussian pdf as in [10] with variance

$$\sigma_{\mathrm{J}}^2 \;=\; \frac{\pi}{2}\cdot a^2 \tag{8}$$

and Fourier transform

$$P_{\mathrm{J}}(f) = \exp(-0.5(2\pi f \sigma_{\mathrm{J}})^2). \tag{9}$$

This corresponds to an average cross-track magnetization profile with error-function shape.

Media noise is data-dependent and results from the random microstructure of the grains in thin-film recording media. The microtrack model imitates the random zig-zag transition effects. It is specified by the parameters N, the number of microtracks, a, the transition width parameter, and L, the threshold below which two transitions erase each other. The random zig-zag form of a transition is captured by dividing the recording track into N equally-sized microtracks (in Fig. 1, there are four microtracks). An ideal transition exhibits an average cross-track magnetization profile with a step response shape. Such ideal transitions are written on each microtrack at a position shifted randomly from the ideal location of the overall transition. The noiseless output of the magnetic recording

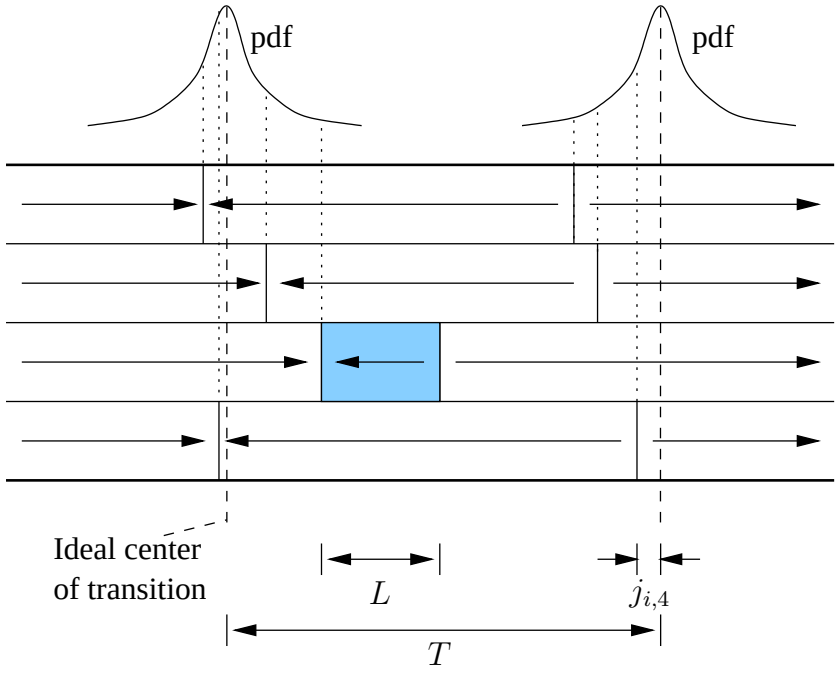

Figure 1: Microtrack channel as described in [17].

[2]This approximation is widely accepted [10].

channel to a single positive transition is then given by

$$v(t) = \frac{1}{N} \sum_{i=1}^{N} h(t - j_i). \tag{10}$$

The jitter process $\{j_i\}$ is modelled as a process that is independent and identically distributed (i.i.d.) according to $p_{\mathrm{J}}(.)$ that in our case is a Gaussian distribution.

The output $v(t)$ of our magnetic recording channel is therefore given by

$$v(t) = \frac{1}{N} \sum_{k=-\infty}^{+\infty} x_k \sum_{i=1}^{N} h(t - kT - j_{i,k}), \tag{11}$$

where $x_k = (u_k - u_{k-1})/\sqrt{2}$, with initial condition $u_{-1} = -1$. The u_k's are generated by a discrete memoryless source (DMS) and take on the two possible values ± 1 with equal probability, i.e. the input is a 0.5-Bernoulli process. Hence, $x_k \in \{-\sqrt{2}, 0, +\sqrt{2}\}$ and correlated. Moreover, $j_{i,k}$ is the jitter of the i-th microtrack at the k-th time step. Note that Eq. (11) is a time-varying convolution.

The noisy output $y(t)$ is given by the noiseless output $v(t)$ that is corrupted by AWGN, $n(t)$. The AWGN is determined by its one-sided power spectral density N_0 and represents electronics noise.

In summary, the behavior of our magnetic recording channel is specified by the five parameters PW50/T, N, a, L, and N_0, (see Fig. 2).

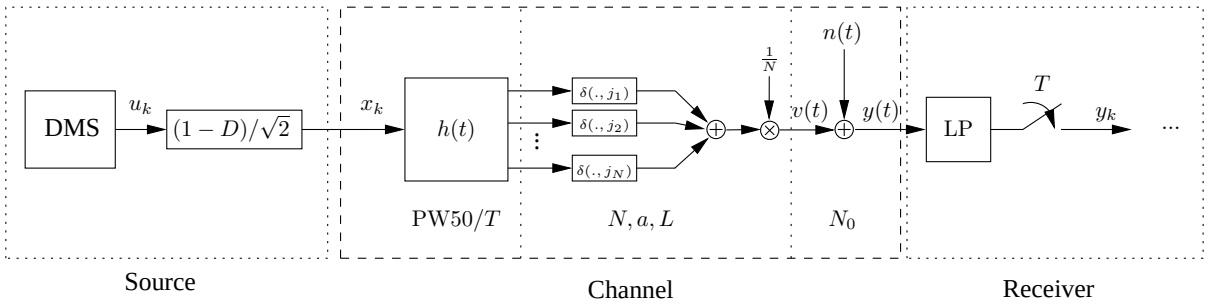

Figure 2: Model for the magnetic recording channel with five parameters.

2.2 Media noise: Ideal, smooth, and real transitions

The microtrack model promises to analyze write-head and media noise separately [10][3]. This separation is a consequence of the definition of *media* noise. We will adopt the

[3]The transition-width parameter, a, can be estimated from the head and media parameter using the Williams–Comstock approximation [10]. As a depends on head and medium, it is strictly speaking not possible to separate the influences of write head and medium completely. However, this separation is justified by noting that a is in first order directly proportional to the fly height.

definition of media noise, given by Caroselli in his thesis [10]: Media noise is for a given jitter distribution, the difference between a finite and an infinite number of microtracks.

For infinitely many microtracks, the noiseless output to a single transition at the input becomes

$$v_{\mathrm{N}\to\infty}(t) = p_{\mathrm{J}}(j) * h(t-j), \tag{12}$$

where $*$ denotes the convolution operator. Using Taylor expansion, we get

$$v_{\mathrm{N}\to\infty}(t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \mathrm{E}_{p_{\mathrm{J}}}[j^k] \frac{\partial^k}{\partial t^k} h(t). \tag{13}$$

On the other hand, applying Taylor expansion to Eq. (10), we obtain, for a finite number of microtracks

$$v(t) = h(t) - \sum_{i=1}^{N} \frac{j_i}{N} h'(t) + \sum_{i=1}^{N} \frac{j_i^2}{2N} h''(t) \pm \ldots. \tag{14}$$

The difference between a finite and infinite number of microtracks therefore equals a weighted sum of the derivatives of the Lorentzian. These weights are the differences between the ensemble moments and the estimated moments (by N samples) of the jitter pdf. Media noise is the difference between a finite and an infinite number of microtracks, i.e.

$$\begin{aligned} n_{\mathrm{m}}(t) &= v(t) - v_{\mathrm{N}\to\infty}(t) & (15)\\ &= -\left(\sum_{i=1}^{N} \frac{j_i}{N} - \mathrm{E}[j] \right) h'(t) + \frac{1}{2} \left(\sum_{i=1}^{N} \frac{j_i^2}{N} - \mathrm{E}[j^2] \right) h''(t) \pm \ldots. & (16) \end{aligned}$$

With increasing N, the estimated moments become more accurate, and the differences start vanishing. For real magnetic recording systems, N ranges from 10 to 50.

With this notion of media noise in mind, we can distinguish three different cross-track magnetization profiles and consequently three different *transitions* (see also Fig. 3):

1. Ideal transition: An ideal transition has width zero, i.e. $a = 0$. The jitter variance is zero (see Eq. (8)). There is no jitter, and all magnetic particles change their polarization at the same location.
2. Smooth transition: A smooth transition has an infinite number of microtracks, $N = \infty$, but a finite transition width a. It exhibits a *smooth* average cross-track magnetization profile that in our case is shaped like an error-function. The slope at the origin is determined by the inverse of the a-parameter, called transition-width parameter. The steepness of the slope is measure of the quality of the write head. A larger a signifies wider transitions.
3. Real transition: A real transition has a *finite* number of microtracks N and a finite transition width a. Owing to the granularity of the medium, it exhibits a *bumpy*, error-function-like cross-track magnetization profile. The thinner the granularity of the medium, the smoother the transition and the more microtracks are needed to model this transition.

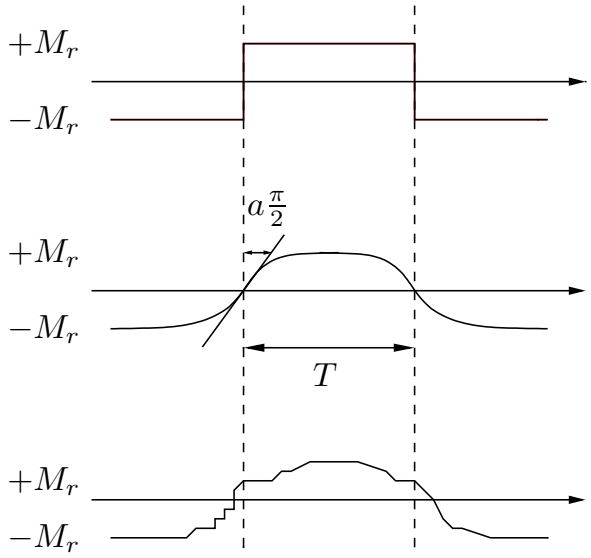

Figure 3: Average cross-track magnetization profiles for different transitions: ideal transition (top), smooth transition (middle), real transition (bottom). $+M_r$ is the positive and $-M_r$ the negative remanent state of the magnetization.

2.3 Receive filter

For channels with AWGN and without data-dependent noise, a sampled matched filter provides a set of sufficient statistics for estimating the input sequence [18]. In the presence of data-dependent noise, this is not true in general.

In practice, an anti aliasing (low-pass) filter is used as receive filter before the sampler. The cut-off frequency of this low-pass filter is a trade-off between collecting most of the signal energy (large bandwidth) and maintaining a good SNR (low bandwidth). We adopt this technique and use a brick-wall-shaped low-pass filter together with a sampler at the channel output (Fig. 2). Clearly, this is suboptimal from a SNR point of view. However, as we are interested in information rates and not in receiver structures that might prefer high SNR, this is not of importance to us. Alternatively, we could also argue that the low-pass filter cannot be changed and is therefore part of the channel (in practice any read-back device has low-pass characteristic).

An alternative way to look at the microtrack channel is to relate it to a multipath environment known from the vast literature on mobile communications. This view is mainly motivated by the hope to exploit the "space" diversity provided by the microtrack model to find better receiver structures. In the mobile-communication setting, the N tracks correspond to N channels from N transmit antennas to N receive antennas. The microtrack channel can be characterized as frequency non-selective because the coherence time is zero (the impulse response of each microtrack is a Dirac, see also Fig. 2). The i.i.d. jitter process corresponds to a rapidly fading process, and the multiplicative model can be adopted [19]. Moreover, no channel state information is available to the receiver

(i.e. the jitter is unknown on each microtrack and for each time step) and the transmitter, except that the distribution (Gaussian) and the variance (a-parameter) are known.

Given that the actual jitter value is unknown to transmitter and receiver, we see no point in pursuing this idea further as there seems to be no way to exploit directly the diversity provided by the microtrack model.

2.4 2nd-order model

The 2nd-order model (2OM) is readily obtained from the microtrack model by using Taylor expansion and considering media noise up to the 2nd moment. The noisy output to a single positive transition, i.e. from -1 to $+1$, is given by

$$y(t) = p_{\mathrm{J}}(t) * h(t) + \sum_{i=1}^{N} \frac{j_i}{N} h^{'}(t) + \frac{1}{2}\left[\sum_{i=1}^{N} \frac{j_i^2}{N} - \mathrm{E}[j^2]\right] h^{''}(t) + n(t). \tag{17}$$

The distribution of the first media noise component is Gaussian, and that of the second one χ^2 (see Fig. 4). It can be shown that they are uncorrelated. The two derivatives are:

$$\frac{\partial}{\partial t} h(t) = -\lambda \frac{8t}{\mathrm{PW50}^2\left(1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2\right)^2}, \tag{18}$$

$$\frac{\partial^2}{\partial t^2} h(t) = \lambda\left[\frac{128t^2}{\mathrm{PW50}^4\left(1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2\right)^3} - \frac{8}{\mathrm{PW50}^3\left(1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2\right)^2}\right]. \tag{19}$$

The first one is an odd function and models the position jitter. The second one is an even function and models the width variation.

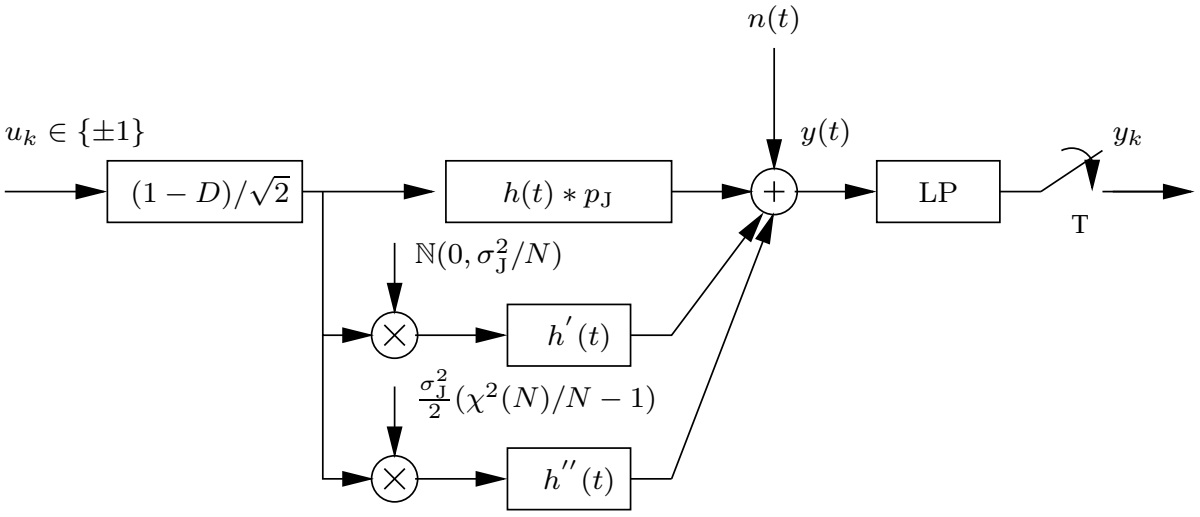

Figure 4: Block diagram of the 2nd-order model.

2.5 Nair-Moon model

Similarly to the 2OM, the Nair–Moon (NM) model [11] also uses two derivatives to mimic the position jitter and the pulse widening, but it uses two first-order derivatives of the Lorentzian: namely, the derivative with respect to time and the derivative with respect to PW50. The Lorentzian from Eq. (2) can be slightly rewritten to

$$h(t, \mathrm{PW50}) = c \cdot \frac{1}{\mathrm{PW50}} \frac{1}{1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2}, \tag{20}$$

where $c = \lambda \cdot \mathrm{PW50}$ is such that the norm equals E_L. The two derivatives then become

$$\frac{\partial}{\partial t} h(t, \mathrm{PW50}) = -c \frac{8t}{\mathrm{PW50}^3 \left(1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2\right)^2}, \tag{21}$$

$$\frac{\partial}{\partial \mathrm{PW50}} h(t, \mathrm{PW50}) = c \left[\frac{8t^2}{\mathrm{PW50}^4 \left(1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2\right)^2} - \frac{1}{\mathrm{PW50}^2 \left(1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2\right)^2} \right]. \tag{22}$$

The first is an odd function and models the position jitter. The second one is an even function and models the width variation.

Note that in contrast to the previous two models, the NM model assumes an ideal transition, i.e. $h(t)$ is not convoluted with a jitter distribution in the signal path (see Fig. 5). This results in higher signal energy at the output of the channel and consequently higher information rates, as will be shown in Sec. V. The data-dependent position and width jitter are modelled by two data-dependent additive white Gaussian noise sources that are characterized by their variances σ_J^2 and σ_W^2, respectively. Sometimes a filter matched to the Lorentzian is used as receive filter instead of the lowpass filter [20].

Whereas in the two previous models, the variance of the media noise was determined by the physical parameters (see Sect. III), the noise variances σ_J^2 and σ_W^2 can in principle

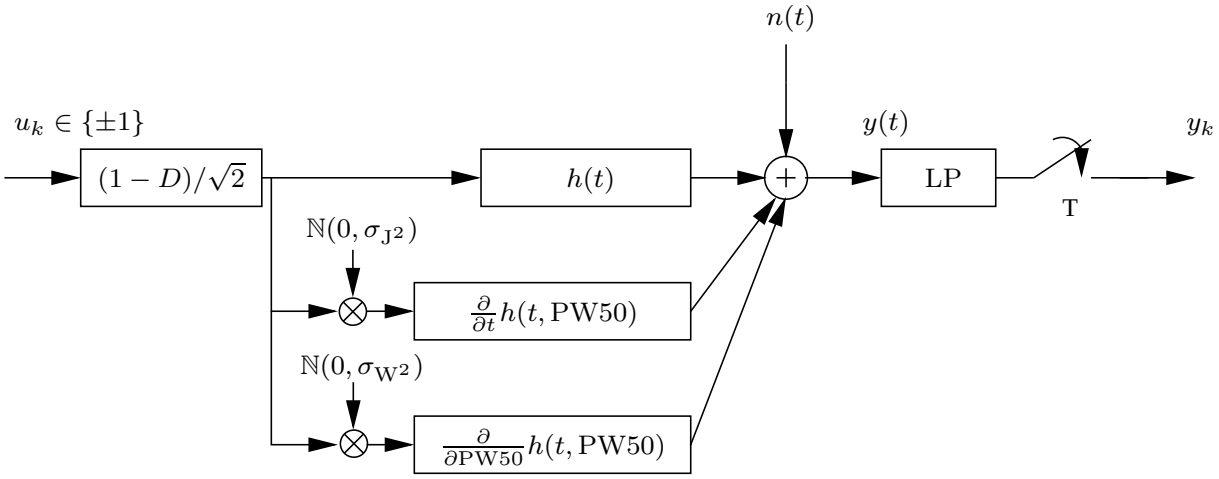

Figure 5: Block diagram of Nair-Moon model from [11].

take on any value. To assure a realistic media noise scenario in the NM model, these variances must be related to physical parameters. For the jitter variance we obtain

$$\sigma_{\mathrm{J}}^2 = a^2 \cdot \frac{\pi}{2}, \tag{23}$$

as for the 2OM and the microtrack model. At first glance, one might think that the width power can take on any value independent of the jitter power leading to a so-called jitter-to-width ratio (JWR) of

$$\mathrm{JWR} = \frac{\sigma_{\mathrm{J}}^2}{\sigma_{\mathrm{W}}^2}. \tag{24}$$

However, from the microtrack model [10], we know that the jitter power is inherently determined by the physical model, and is about three times the width power. We adopt this technique, and will set JWR = 3 for the rest of this report.

2.6 Comparison of the models

The microtrack model is the most accurate model. Its output results from a convolution of the input and the Lorentzian waveform with shifts that vary in time, i.e. a time-varying convolution. The microtrack model includes partial signal erasure. Neither of the other two models can do this. Moreover, other impairments caused by the medium can easily be incorporated into the microtrack model, for example nonlinear transition shifts, overwrite effects, and read-back nonlinearities [10].

The 2OM allows an efficient implementation as the time-varying convolution is avoided. Because of the Taylor expansion, the jitter is now multiplied by the Lorentzian and is no longer an argument of it. The first and second derivatives can therefore be computed at the beginning of a simulation and stored. The noise model still is data-dependent, but partial signal erasure and nonlinear transition shifts do not make sense because the jitter is no longer an argument of the Lorentzian.

In the NM model, data-dependent noise is modelled by two data-dependent AWGN noise sources. Partial signal erasure is not included in the Nair-Moon model. Contrary to the two previous models, *ideal* transitions here are assumed in the absence of media noise as transition response. For the two previous models the ratio between the position jitter power and width noise power is given inherently. Not so for the NM model. In order to obtain a physical meaningful model, the JWR cannot be arbitrary but must be determined from physical measurements or related to the microtrack model.

3 Average power spectral density

For all media noise models presented above, it is possible to compute the average power spectral density because their noiseless outputs are cyclostationary. While this is obvious for the NM model, it is not so clear for the microtrack model. Hence, we will present the

average power spectral density of the microtrack model next, and discuss it in detail. For the other two models, the average power spectral density is obtained by modifying the one from the microtrack model, as mentioned at the end of this section.

Media noise in the microtrack model is data-dependent noise. The output $v(t)$ of the microtrack model is a cyclostationary process, as shown in Appendix A. This allows us to compute an average power spectral density of the output, and consequently to state a SNR definition.

The power spectral density averaged over a period T is computed in a straight forward manner. It can be shown (see Appendix B) that the average power spectral density for the microtrack model is given by

$$\begin{aligned}\bar{\Phi}_V(f) &= \underbrace{\left(\frac{N-\varepsilon}{N}\right)^2 \frac{|H(f)|^2}{T} \cdot \Phi_X(f) \cdot |P_{\mathrm{J}}(f)|^2 +}_{S(f)} \\ &\quad \underbrace{\frac{N-\varepsilon}{N^2} \frac{|H(f)|^2}{T} r_X(0) \cdot \left[1 - |P_{\mathrm{J}}(f)|^2\right]}_{N_{\mathrm{m}}(f)}, \end{aligned} \tag{25}$$

where $H(f) = (\mathrm{PW50}/2) \cdot \pi \cdot \exp(-\mathrm{PW50}/2 \cdot |2\pi f|)$ is the spectrum of the Lorentzian, $\Phi_X(f)$ is the spectrum of the input sequence X, $P_{\mathrm{J}}(f)$ is the Fourier transform of the jitter probability, $r_X(0)$ is the average symbol energy of X, N is the number of microtracks, and ε is the average number of erased microtracks.

The average power spectral density consists of two terms: a signal term $S(f)$, called average signal power spectral density, and a signal-dependent noise term $N_{\mathrm{m}}(f)$, called average media noise power spectral density. The first models the pulse widening of the input signal, the second reflects the noise caused by the position uncertainty of the transitions in the output signal. The first term does not depend on the number of microtracks N but the second does. To obtain more insight into this formula, we will now consider special cases and relate them to results known from the literature (for ease of interpretation we set $\varepsilon = 0$).

3.1 Ideal transition: $P_{\mathrm{J}}(f) = 1$

For $P_{\mathrm{J}}(f) = 1$ we have

$$\bar{\Phi}_V(f) = \frac{|H(f)|^2}{T} \cdot \Phi_X(f). \tag{26}$$

This is equivalent to the case in which there is no jitter, only one track, and an ideal write head that causes an infinitely sharp transition.

3.2 Smooth transition: $N = \infty$

For $N = \infty$ we have

$$\bar{\Phi}_V(f) = \frac{|H(f)|^2}{T} \cdot \Phi_X(f) \cdot |P_{\mathrm{J}}(f)|^2. \tag{27}$$

Because $N = \infty$ and the track width is finite, the granularity of the medium is zero — or stated otherwise, the medium is ideal and does not cause any media noise. The pulse widening is due to the non-ideal write head only. This is equivalent to the Fourier transform of Eq. (2) reported in [17].

3.3 Single-track channel: $N = 1$

For $N = 1$ we have

$$\bar{\Phi}_V(f) = \frac{|H(f)|^2}{T} \cdot \left[\Phi_X(f) \cdot |P_{\mathrm{J}}(f)|^2 + r_X(0) \left[1 - |P_{\mathrm{J}}(f)|^2 \right] \right]. \tag{28}$$

This is identical to the result in [21]. For a small jitter, the first-order Taylor expansion of the second term yields $r_X(0)(2\pi f \sigma_{\mathrm{J}})^2$. Assume now that the read-back device is a low-pass filter. To obtain an estimate of the shape of the average power spectral density after the low-pass filter, we multiply $\bar{\Phi}_V(f)$ by $1/(2\pi f)^2$ and see that the second term, i.e. the media noise term, becomes constant. It is therefore present in the entire spectrum of the read-back device as is band-limited white noise, and models the position uncertainty about the transitions. This noise can be reduced by increasing the number of microtracks, i.e. the resolution, which is equivalent to decreasing the granularity of the medium.

The signal and the media noise power spectral density are shown in Fig. 6 for random input X. In this case the input spectrum $\Phi_X(f)$ becomes

$$\Phi_X(f) = 1 - \cos(2\pi f) \qquad -1 \le f \le 1 \tag{29}$$

where f is the normalized frequency.

3.4 Including partial signal erasure

For the microtrack model, partial signal erasure is incorporated by assuming that on the average there are ε erased microtracks at each transition. Then we have only $N - \varepsilon$ "active" microtracks and the average power spectrum becomes as in Eq. (25)

$$\bar{\Phi}_V(f) = \left(\frac{N - \varepsilon}{N} \right)^2 \cdot \frac{|H(f)|^2}{T} \left[\Phi_X(f) \cdot |P_{\mathrm{J}}(f)|^2 + \frac{r_X(0)}{N - \varepsilon} \cdot \left[1 - |P_{\mathrm{J}}(f)|^2 \right] \right]. \tag{30}$$

Partial signal erasure reduces media noise on one hand but on the other hand reduces the signal power as well. The parameter ε is determined from

$$P_{\mathrm{erasure}} = \frac{\varepsilon}{N},$$

where P_{erasure} is related to T and L as shown in [10].

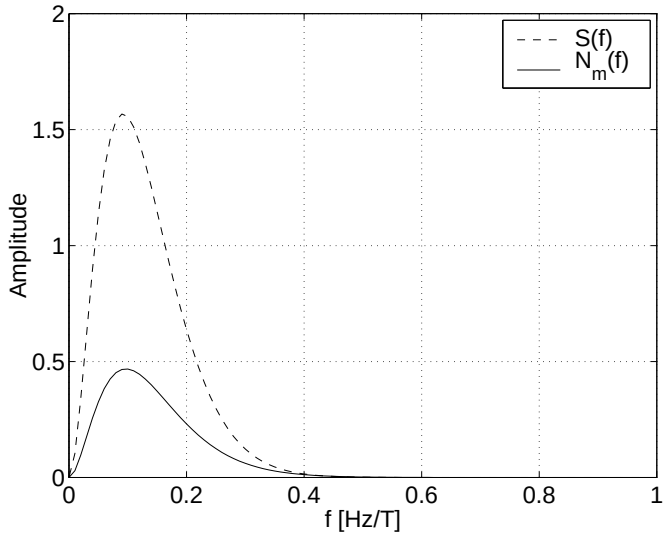

Figure 6: Average signal power spectral density $S(f)$ and average media noise power spectral density $N_{\mathrm{m}}(f)$ for the microtrack model at PW50/$T = 3.2$, $a = 0.3$, $r_X(0) = 1$, $\varepsilon = 0$, and $N = 1$. Note that for these parameters the two spectra have almost identical shape.

3.5 Signal-to-noise ratio (SNR)

SNR is the key performance measure in signal processing for assessing different algorithms. Definitions of SNR strongly depend on the underlying model. For our model, the definitions of $S(f)$ and $N_{\mathrm{m}}(f)$ are given by the average power spectral density of the output $v(t)$ as shown above, i.e.

$$S(f) \triangleq \left(\frac{N-\varepsilon}{N}\right)^2 \frac{|H(f)|^2}{T} \cdot \Phi_X(f) \cdot |P_{\mathrm{J}}(f)|^2 \tag{31}$$

$$N_{\mathrm{m}}(f) \triangleq \frac{N-\varepsilon}{N^2} \frac{|H(f)|^2}{T} \cdot r_X(0) \cdot \left[1 - |P_{\mathrm{J}}(f)|^2\right]. \tag{32}$$

Note that if a, PW50/T, T, $r_X(0)$, and ε are given, $N_{\mathrm{m}}(f)$ has one free parameter, namely N, which allows the media noise power spectral density to be scaled.

Because the optimum receive filter is unknown, we would like to have a definition of SNR independent of the receive filter. As signal energy, we take $\mathrm{E}[x_k^2] = E_s$, i.e. the input symbol energy. The media noise energy is measured *before* the receive filter. This is sufficient in defining a noise blend (see below) that is needed to compare different noise scenarios. With the help of the Parseval theorem, the media noise power (MNP) is defined as

$$\mathrm{MNP} \triangleq \int_{-\infty}^{+\infty} N_{\mathrm{m}}(f) df \tag{33}$$

$$= \frac{N-\varepsilon}{N^2} \int_{-\infty}^{+\infty} \frac{|H(f)|^2}{T} r_X(0) \cdot \left[1 - |P_{\mathrm{J}}(f)|^2\right] df. \tag{34}$$

Here it is important to note that the MNP is a function of the microtracks N because $N_{\mathrm{m}}(f)$ is a function of N. The SNR is now defined as

$$\mathrm{SNR} \triangleq \frac{E_s}{\frac{N_0}{2} + T \cdot \mathrm{MNP}}. \tag{35}$$

Note: T cancels out, and the SNR definition is independent of T.[4] The only difference of this definition to the one in [22] lies in how the MNP is computed. If the underlying model is the NM model, both MNPs are equal. The media noise factor (MF) indicates the amount of media noise that is part of the total noise, and is defined as

$$\mathrm{MF} \triangleq \frac{T \cdot \mathrm{MNP}}{\frac{N_0}{2} + T \cdot \mathrm{MNP}}. \tag{36}$$

MF $= 0.1$ means 10% media noise and 90% AWGN. The amount of media noise is controlled by adjusting N.

3.6 Average media noise power spectral density for 2OM

To obtain the average power spectral density $\bar{\Phi}_V(f)(.)$ for the 2OM only $N_{\mathrm{m}}(f)$ has to be replaced. Since the two derivatives (19) are independent, we obtain

$$N_{\mathrm{m}}(f) \;=\; r_X(0)\left[\sigma_J^2 \cdot \left|\mathcal{F}\{\frac{\partial}{\partial t}h(t)\}\right|^2 + \frac{\sigma_J^4}{2} \cdot \left|\mathcal{F}\{\frac{\partial^2}{\partial t^2}h(t)\}\right|^2\right], \tag{37}$$

where the Fourier transforms of the derivatives are given by

$$\mathcal{F}\{\frac{\partial}{\partial t}h(t)\} \;=\; -i\exp(-\mathrm{PW50}\pi|f|)\sqrt{\pi \mathrm{PW50} E_{\mathrm{L}}}2\pi|f|, \tag{38}$$

$$\mathcal{F}\{\frac{\partial^2}{\partial t^2}h(t)\} \;=\; -\exp(-\mathrm{PW50}\pi|f|)\sqrt{\pi \mathrm{PW50} E_{\mathrm{L}}}(2\pi f)^2. \tag{39}$$

Note that partial signal erasure is not included because the jitter is no longer an argument of the Lorentzian.

3.7 Average media noise power spectral density for NM model

For the NM model, the signal power spectral density and the media noise power spectral density have to be replaced. The signal power spectral density is obtained from the microtrack model by setting a to zero, i.e.

$$S(f) = \frac{|H(f)|^2}{T} \cdot \Phi_X(f). \tag{40}$$

[4] We always assume $T = 1$.

The media noise power spectral density becomes

$$N_{\mathrm{m}}(f) = r_X(0)\left[\sigma_J^2 \cdot \left|\mathcal{F}\{\frac{\partial}{\partial t}h(t,\mathrm{PW50})\}\right|^2 + \sigma_W^2 \cdot \left|\mathcal{F}\{\frac{\partial}{\partial \mathrm{PW50}}h(t,\mathrm{PW50})\}\right|^2\right], \qquad (41)$$

with Fourier transforms

$$\mathcal{F}\{\frac{\partial}{\partial t}h(t,\mathrm{PW50})\} = -i\exp(-\mathrm{PW50}\pi|f|)\sqrt{\pi \mathrm{PW50}E_{\mathrm{L}}}2\pi|f| \qquad (42)$$

$$\mathcal{F}\{\frac{\partial}{\partial \mathrm{PW50}}h(t,\mathrm{PW50})\} = -\exp(-\mathrm{PW50}\pi|f|)\sqrt{\pi \mathrm{PW50}E_{\mathrm{L}}}\pi|f|. \qquad (43)$$

Note that the NM model does not include partial signal erasure.

3.8 Spectra and MNPs of the different models

The media noise spectral densities for the three models are shown in Fig. 7 for $\mathrm{PW50}/T = 3.2, a = 0.3\ T, N = 1, r_X(0) = 1$, and $\varepsilon = 0$. The MNPs together with the abbreviations from Fig. 7 are given in Table 1. We see that the MNP of the microtrack model is less than the MNP of the 2OM and the NM model. The two derivatives in the latter two models are no longer arguments of the Lorentzian. Moreover, they are mutually independent and hence the triangle inequality applies.

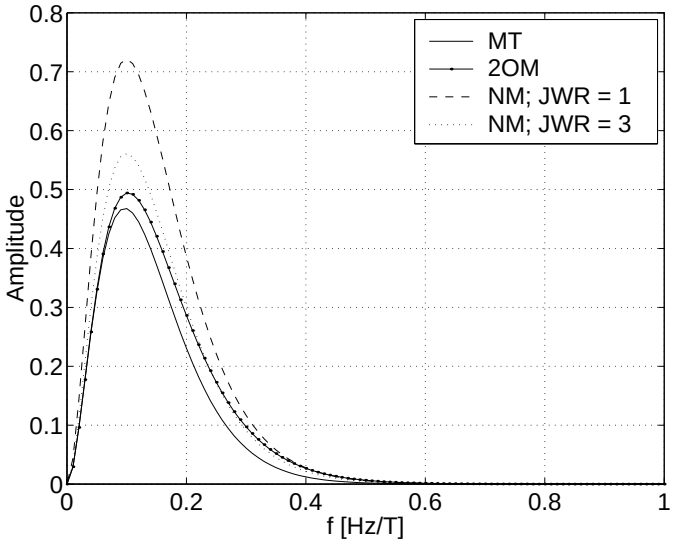

Figure 7: Noise power spectral densities for various media noise models at $\mathrm{PW50}/T = 3.2$, $a = 0.3$, $N = 1$, $r_X(0) = 1$, and $\varepsilon = 0$.

Table 1: MNP for various media noise models.

Model	Abbreviation	MNP
Microtrack	MT	0.1630
2nd Order Model	2OM	0.1904
Nair-Moon, JWR=1	NM, JWR=1	0.2638
Nair-Moon, JWR=3	NM, JWR=3	0.2051

4 Information Rates via the SLLB

The capacity of a linear binary-input discrete-time channel without media noise, e.g. a partial response channel, is an open problem in information theory [23]. Surprisingly enough, even the much simpler problem of computing the average mutual information (information rate) under the assumption that the input is i.i.d. and uniformly distributed over the input alphabet (uniform input information rate (UIIR)) has only been solved recently [24, 25]. The computation of the UIIR is accomplished by sampling both a long input sequence x and the corresponding output sequence y, followed by the computation of both $\log p(y|x)$ and $\log p(y)$ by means of a forward sum-product recursion on the joint source/channel trellis. It is therefore obvious to approximate otherwise intractable (non-finite state) channels such as the microtrack channel by a finite-state channel model, which can be done either by analysis (if possible) or simply by training a parameterized model of simulated or measured data [26]-[28].

In this report we pursue a different approach. First we compute information rates for i.i.d. power-limited Gaussian inputs in the classical way, and then translate this information rate into i.i.d. binary input (0.5-Bernoulli input process) by means of the SLLB that is a conjectured lower bound on the capacity of binary-input channels with memory. We will also generalize the SLLB to Markov input processes, which allows us to obtain non-uniform input information rates (NUIIRs).

4.1 Shamai-Laroia bound

The UIIR[5] of a discrete-time channel with i.i.d. Gaussian inputs, per symbol energy constraint $\mathrm{E}[x_k^2] \leq E_s$, finite memory, real channel coefficients, and AWGN at the output, is given by [29]

$$I_{\mathrm{UIIR,G}} = \frac{1}{2\pi}\int_0^{\pi} \log_2(1+\frac{E_s}{N_0/2}|S_h(\Theta)|^2)d\Theta, \tag{44}$$

where $|S_h(\Theta)|^2$ is the discrete-time channel power spectrum at the receiver obtained by a sampled filter that is matched to the channel. $\frac{E_s}{N_0/2}$ is the SNR. The capacity of the binary-input AWGN channel without memory for a given SNR is denoted by $C_{\text{Bin w/o mem}}(\mathrm{SNR})$;

[5] In the literature, UIIR is also termed *symmetric* information rate [25].

it can be easily computed [19]. The Shamai-Laroia lower bound $C_{\mathrm{SLLB}}(.)$ is then given by [9]

$$C_{\mathrm{SLLB}}(\frac{E_s}{N_0/2}) = C_{\text{Bin w/o mem}}(2^{I_{\mathrm{UIIR,G}}} - 1) \quad (45)$$

$$\leq C_{\text{Bin with mem}}(\frac{E_s}{N_0/2}). \quad (46)$$

To obtain a better understanding of the SLLB, consider a discrete-time channel with real coefficients $h_0, h_1, \ldots, h_L$ after a whitened matched filter. The output y_k at time k is given by

$$\begin{aligned} y_k &= \sum_{i=0}^{L} x_{k-i} h_i + n_k \\ &= h_0 x_k + \sum_{i=1}^{L} x_{k-i} h_i + n_k, \end{aligned}$$

where the second term in the last line is the ISI term and a length-L approximation of the whitening filter was employed (for more details see [13]). Because the input to our channel is binary, the noise consists of an ISI term, and the AWGN term, n_k, and is structured (a mixture of different distributions). The basic idea of the SLLB is now to replace the structured noise term by an AWGN noise term $\tilde{n}_k$ of the *same* power, i.e.

$$y_k = h_0 x_k + \sum_{i=1}^{L} x_{k-i} h_i + n_k \quad (47)$$

$$= h_0 x_k + \tilde{n}_k. \quad (48)$$

The problem of computing the information rate of a binary-input discrete-time channel with memory is now reduced to computing the information rate of a binary input channel with AWGN without memory.

The algorithmic description for computing UIIR via the SLLB (the points enumerated correspond to the numbers in Fig. 8) is as follows

1. Determine the desired $\mathrm{SNR_b}$ at which you want to know the UIIR of the binary-input channel with memory.
2. Compute the rate R_{g} of the same channel at the same $\mathrm{SNR_b}$ but with i.i.d. Gaussian input.
3. Compute the necessary $\mathrm{SNR_g}$ to achieve the same rate R_{g} for a Gaussian input channel without memory.
4. Compute R_{b} of the binary input channel without memory at $\mathrm{SNR_g}$.
5. Assign this rate R_{b} to the UIIR of the i.i.d. binary-input channel with memory at $\mathrm{SNR_b}$.

The general idea behind the SLLB is to consider only the second moment of the structured noise and to replace its distribution by a Gaussian. In the next subsection, we will briefly address some theoretical problems resulting therefrom.

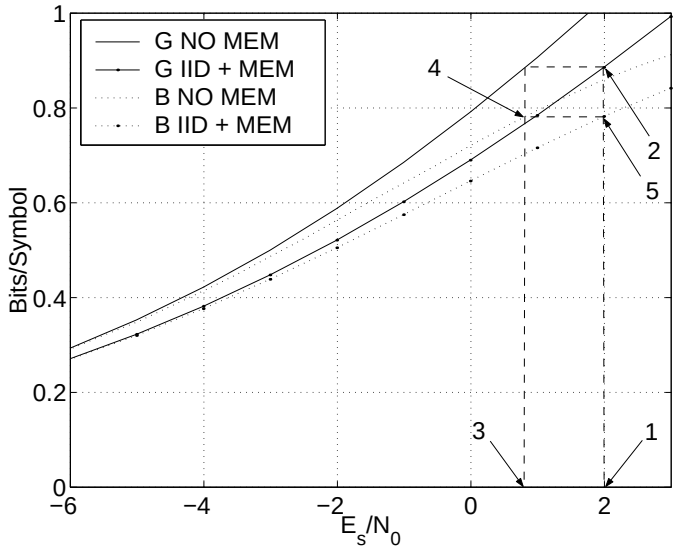

Figure 8: Conceptual sketch of the algorithmic procedure to compute UIIRs via the SLLB. "G" means Gaussian input and "B" binary input.

4.2 Lindeberg condition

In the Shamai–Laroia bound a mixture of ISI terms and AWGN is replaced by a new AWGN component with the same power as the mixture. Intuitively one would think that for a large number of ISI terms this approximation is good because of the central limit theorem (CLT). However, this is wrong. Because the sum of the squared channel coefficients is normalized to one, the coefficients of the channel impulse response tend to zero if their number goes to infinity. Therefore the so-called Lindeberg condition (see [30]) is not fulfilled, i.e. the variances of the ISI terms and the AWGN can be very different.

When we have a sum of independent random variables with different distributions, the Lindeberg condition (for an elegant proof see [30]) asserts that the shapes of the individual distributions are negligible compared with the total shape such that the CLT holds.

Lindeberg condition [30]: *Assume that $x_1, x_2, \ldots$ are mutually independent random variables distributed according to the pdfs $p_1, p_2, \ldots$ with zero mean and variance σ_i^2. Now define*

$$s_n^2 \triangleq \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2. \tag{49}$$

If we assume that for each $t > 0$

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|y| \geq t s_n} y^2 p_k dy \to 0, \tag{50}$$

then the distribution of the normalized sum

$$S = \frac{X_1 + X_2 + \cdots + X_n}{s_n} \tag{51}$$

tends to the normal distribution with zero mean and unit variance.

The Lindeberg condition guarantees that the individual variances σ_i^2 are small compared with the their sum $s^2 = \sum_i \sigma_i^2$, in the sense that for a given ϵ and sufficiently large n

$$\frac{\sigma_i}{s_i} < \epsilon \quad k = 1, \ldots, n. \tag{52}$$

As example consider the case where X_k are uniformly distributed between $-a_k$ and $+a_k$ (the variance is $1/3a_k^2$). Hence the Lindeberg condition is satisfied if the a_k remain bounded and $a_1^2 + \cdots + a_n^2 \to \infty$. In this case the sum in Eq. (52) vanishes identically for all n sufficiently large. On the other hand, if $\sum_k a_k^2 \leq \infty$, then s_n remains bounded and Eq. (52) is not fulfilled. In this case the CLT does not hold.

The general approach in the literature for proving the CLT is to utilize a scaled sum, S, and to show by expansion of the moment-generating functions that the difference between the moments of S and the normal distribution vanishes for n large. As scaling factor, one often uses the square root of the summed variances, s_n^2, although depending on the particular sequence S, other scaling factors are more effective (see [30, 31]). In our case, we see that the Lindeberg condition is not fulfilled for *all* possible values of the channel coefficients. However, this does not imply that for large n the structured noise distribution never converges to a Gaussian distribution. Consider for instance the case that all ISI terms $h_1, \ldots, h_L$ are equal. For large L, the ISI terms are well approximated by a Gaussian distribution. On the other hand, if for instance one term, say h_0, is considerably larger than the other ones, the resulting distribution will be dominated by the distribution of this h_0, which is not a Gaussian.

Irrespective of whether the structured noise term is well approximated by a Gaussian distribution, the central question is whether the Gaussian distribution is a worst-case distribution for the power constraint $\sum_{\ell=1}^{L} |h_\ell|^2 + |n_k|^2$ (h_0 is not an ISI term). If this is the case, then the SLLB will always deliver a lower bound on the UIIR and therefore a lower bound on the capacity. If not, it might be that the SLLB even delivers a rate above capacity. The SLLB is therefore a *conjectured* lower bound on capacity.

Related to this question is the work of Shamai and Verdú in [32]. Therein, it is shown that for a given noise power AWGN is not worst-case noise on a binary input channel without memory. This is in sharp contrast to i.i.d. Gaussian inputs, where it is well known that AWGN is worst-case noise. In the binary case, the worst-case noise distribution is in general a mixture of two lattice probability mass functions. The authors also show that for the binary-input channel without memory the difference in capacity for Gaussian and worst-case noise can amount up to 0.118 Bits/Symbol.

The difference of the SLLB and the true UIIR depends on the closeness of the Gaussian distribution to the structured distribution (the closeness can be measured by the Berry–Esseén Theorem [30]). The empirical results in [24] may serve as illustrative example.

Therein, the difference between the SLLB and the exact UIIR (computed via sum-product algorithm) is shown for two different channels. This difference is only pronounced at high SNR where the structured noise exhibits a different distribution than the Gaussian. The difference is bigger for the DICODE channel[6], i.e. $h(D) = (1-D)/\sqrt{2}$, than for a channel with memory six whose channel coefficients are Gaussian like.

4.3 Non-Uniform Input Information Rates (NUIIR) via the generalized SLLB (g-SLLB)

Here, we generalize the SLLB to inputs that are generated by a parameterized Markov process. For a given memory of the Markov process, the transition probabilities are free parameters that can be optimized in order to shape the (parameterized) input spectrum to the channel. Note that in contrast to the water-filling procedure for Gaussian input, we impose here a parameterized spectrum on the input. By shaping the parameterized input spectrum to the channel, a higher information rate is achieved for Gaussian inputs that is then translated via the SLLB to binary inputs resulting in the NUIIR. This procedure is termed g-SLLB where the "g" stands for "generalized" because it generalizes the SLLB, in which only i.i.d. inputs were considered.

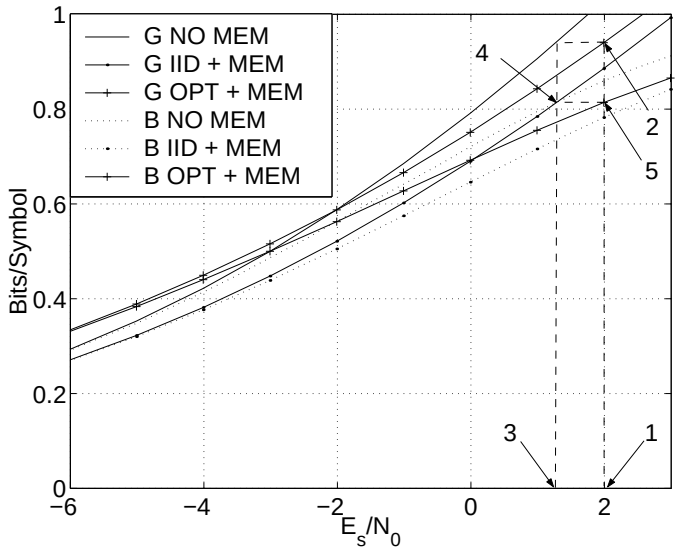

Figure 9: Conceptual sketch of the algorithmic procedure to compute NUIIRs via the g-SLLB. "G" means Gaussian input and "B" binary input.

[6]Note: The distribution of ISI and AWGN of the DICODE channel is very different form a Gaussian distribution.

The algorithmic description for computing NUIIRs via the g-SLLB (the points enumerated correspond to the numbers in Fig. 9) is as follows:

1. Determine the desired $\mathrm{SNR}_{\mathrm{b}}^{\mathrm{n}}$ at which you want to know the NUIIR of the binary-input channel with memory.
2. Compute the non-i.i.d., i.e. optimized rate, $R_{\mathrm{g}}^{\mathrm{n}}$, of the same channel at the same $\mathrm{SNR}_{\mathrm{b}}^{\mathrm{n}}$ but with the parameterized input spectrum that has been optimized (the input is still Gaussian but *correlated*).
3. Compute the necessary $\mathrm{SNR}_{\mathrm{g}}^{\mathrm{n}}$ to achieve the same rate $R_{\mathrm{g}}^{\mathrm{n}}$ for a Gaussian input channel without memory.
4. Compute the rate $R_{\mathrm{b}}^{\mathrm{n}}$ of the binary input channel without memory at $\mathrm{SNR}_{\mathrm{g}}^{\mathrm{n}}$.
5. Assign this rate $R_{\mathrm{b}}^{\mathrm{n}}$ to the NUIIR of the *optimized* binary-input channel with memory at $\mathrm{SNR}_{\mathrm{b}}^{\mathrm{n}}$.

All NUIIRs provided in this report were obtained by optimization for a memory-one Markov source (this is essentially the correlation box in Fig. 2). There is a vast literature on computing power spectral densities of Markov sources. For a memory-one source the spectrum can be found in [33].

For the g-SLLB, Fig. 10 shows the same behavior as reported in [24], namely that at low SNR a transmission above the capacity of power limited Gaussian inputs without memory is possible. Note also that at very low SNR the UIIR and the NUIIR for binary inputs tend in both cases to the rates achieved with Gaussian inputs.

In Fig. 11 the rates obtained by the SLLB for the DICODE channel are related to those from [24]. We conclude that we obtain i.i.d. and optimized rates that are very close the exact ones.

A related approach has been investigated in [34, 35], with the difference that the transformation to binary inputs via the SLLB was substituted by a constrained water-filling procedure. The constraints imposed on the input spectrum approximate the spectrum of binary signals. An exact spectrum for binary inputs requires infinitely many constraints.

An alternative approach would be to approximate the optimum unconstrained input spectrum (from water-filling) by a Markov process [36]. We did not pursue this further.

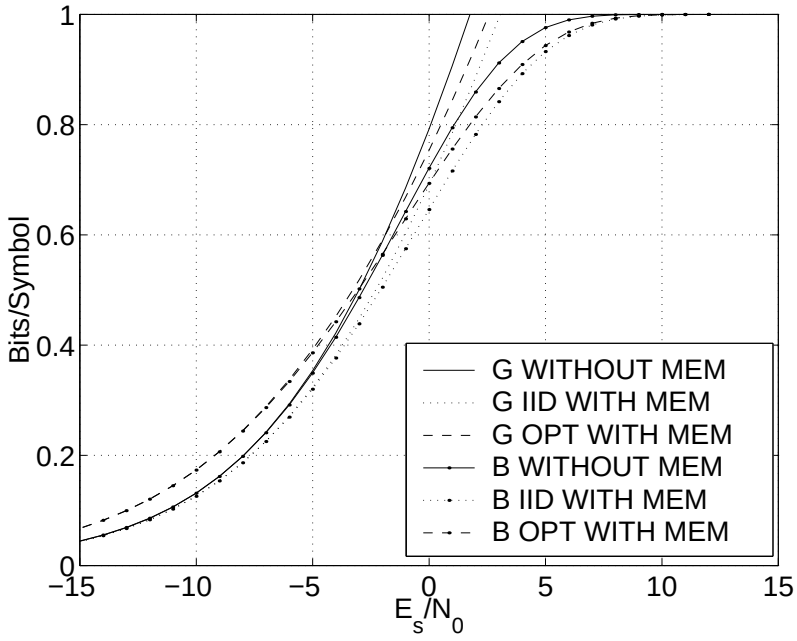

Figure 10: UIIR and NUIIR for DICODE channel.

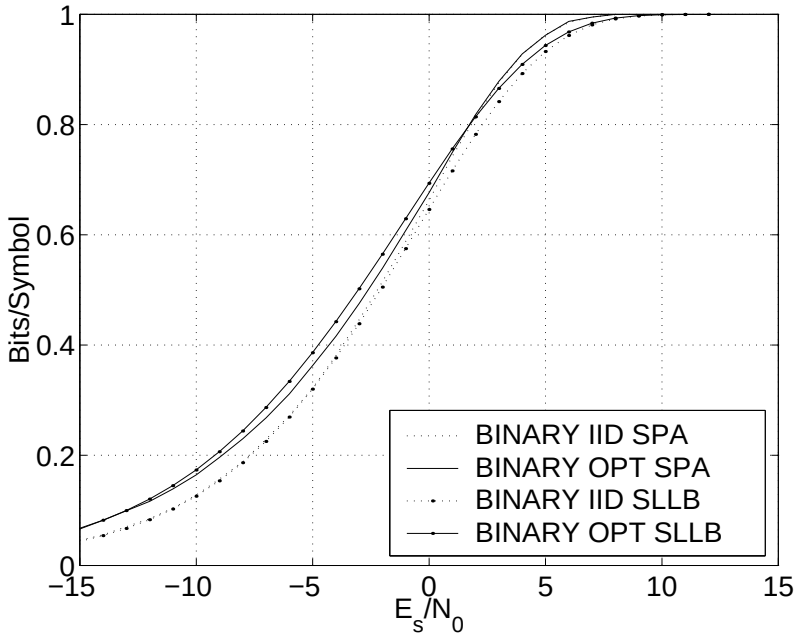

Figure 11: Comparison of SLLB and SPA method from [24] for the DICODE channel.

5 Numerical results

In this sections, numerical results are provided in order to study the influence of the channel parameters $N, a, L, \mathrm{PW50}/T, N_0$ (with $T = 1$, $r(0) = 1$) qualitatively and quantitatively. The input sequence X was assumed to be random resulting into a spectrum given by Eq. (29). Moreover, the media noise factor MF is varied to examine the effect of media noise for a fixed noise power and to answer the question whether media noise is preferable to AWGN from an information-theoretic perspective. To this end, all three models for media noise are used.

5.1 Information rates without media noise

In this subsection, E_s/N_0 is chosen as SNR where $E_s = 1$ is the energy of the input signal. The number of microtracks is assumed to be infinite, i.e. there is no media noise.

In Fig. 12, UIIRs for the case of no memory, i.e. binary phase-shift keying (BPSK), and a Lorentzian channel with densities $\mathrm{PW50}/T = 2.0, 2.6, 3.2$, and 3.8 are shown. The Lorentzian is normalized to $E_\mathrm{L} = 1$. The jitter variance is kept small, i.e. $a = 0.1\ T$. This setting allows us to study the loss in SNR due to increased $\mathrm{PW50}/T$. At 0.9 Bits/Symbol, a loss of 4 dB in SNR can be observed when increasing $\mathrm{PW50}/T$ from 2.0 to 3.8.

In Fig. 13, the UIIRs are computed for $\mathrm{PW50}/T$=3.2 and varying jitter variance σ_J^2 with $a = 0.1, 0.3$, and $0.5\ T$. Again there is no media noise, and one can observe the loss in SNR due to wider transitions. At 0.9 Bits/Symbol a loss of 2 dB is experienced when the transition width is increased from 0.1 to $0.5\ T$. For $a = 0.5\ T$ essentially, the transitions are much wider, resulting in alternating output pulses that overlap more. This overlapping reduces amplitude and energy of the output signal.

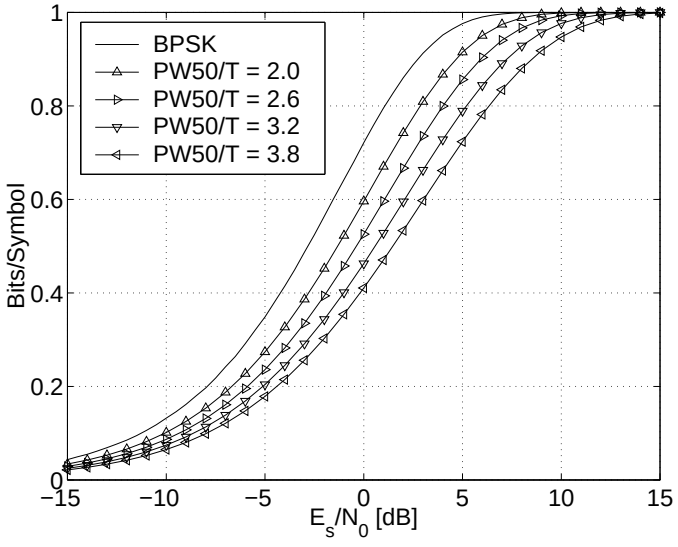

Figure 12: UIIR vs. E_s/N_0 [dB] for different $\mathrm{PW50}/T$.

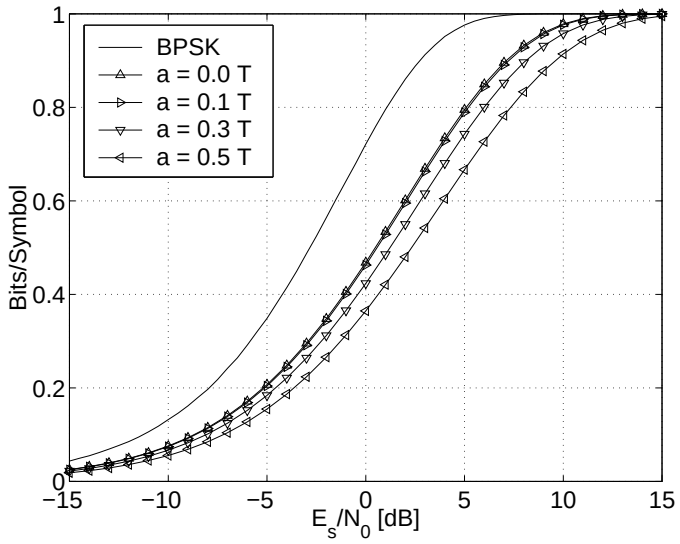

Figure 13: UIIR vs. E_s/N_0 [dB] for PW50/$T = 3.2$ and various values of a.

5.2 Information rates with media noise and varying MF

Again PW50/T=3.2 and the jitter variance is determined by $a = 0.3\ T$. E_s/N_0 is chosen as SNR that consists only of electronics noise. In Figures 14 and 15, the amount of media noise varies along the curves from 0 to 1. For a very high SNR, i.e. a high E_s/N_0, there is almost no electronics noise, but media noise that was not taken into account in the SNR. This allows us to study the limiting effect of media noise.

The number of microtracks is 100, 10, and 1 (see Fig. 14). As $N = 100$, there is virtually no difference to the information rate of Fig. 13, where for the same a and the same PW50/T an infinite number of microtracks was assumed. For $N = 1$, the jitter is strong enough to preclude transmission above 0.9 Bits/Symbol, even in the absence of electronics noise (see Fig. 14).

Finally, the influence of the erasure probability is shown in Fig. 15. As the number of microtracks was set to $N = 100$, $E = 10$ corresponds to an erasure probability of 10%. In [10], it was shown that in the microtrack model with $a = 0.3\ T$ an erasure probability of 10% translates into an erasure threshold $L = 0.75\ a$. At 0.9 Bits/Symbol, this results in a SNR loss of about 1 dB.

We conclude that partial signal erasure results in a loss of SNR and consequently in a shift of the UIIR curves only. In what follows, we will set the erasure probability to zero. By appropriately shifting it is possible to obtain the resulting curves for any amount of erasure probability. The qualitative behavior, i.e. the shape, of the curves will remain unchanged.

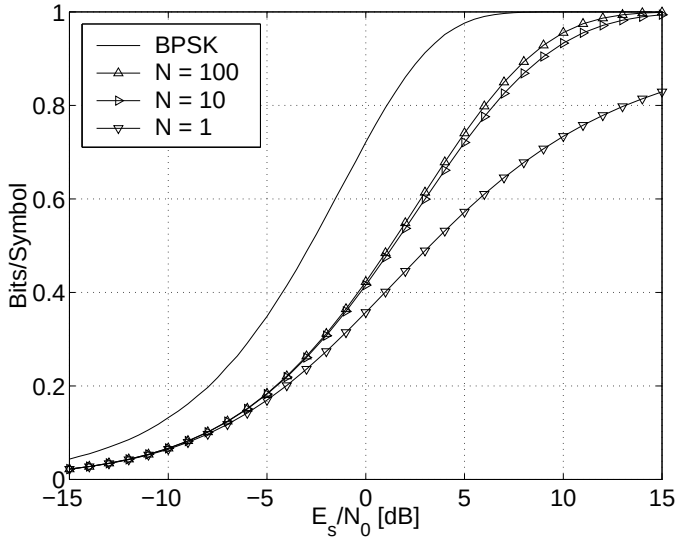

Figure 14: UIIR vs. E_s/N_0 [dB] for PW50/$T = 3.2$, $a = 0.3\ T$, and different numbers of microtracks N.

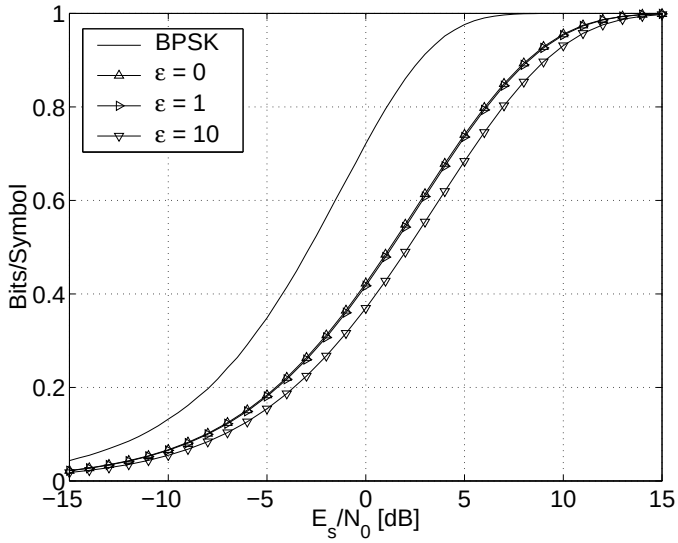

Figure 15: UIIR vs. E_s/N_0 [dB] for PW50/$T = 3.2$, $a = 0.3\ T$, $N = 100$, and $E = 0, 1, 10$.

5.3 Information rates with fixed SNR blend

In practice, we are given a certain operating point, i.e. a particular PW50/T, N, a, L, and N_0. For our computation, we can artificially generate operating points in the following way. For a given a and a given PW50/T, one first computes the MNP at $N = 1$. After having fixed the MF, we solve Eq. (36) for N_0. Various SNR points are then obtained by varying N. In a computer-simulation environment, N must be an integer number. For our computation however, we can assume N to be a real number. In the remainder of this report, the SNR definition of Eq. (35) is used, i.e.

$$\mathrm{SNR} = \frac{E_s}{\frac{N_0}{2} + T \cdot \mathrm{MNP}}. \tag{53}$$

In this subsection, only the microtrack model was used as the underlying media noise model.

5.3.1 Varying PW50/T

Figures 16 and 17 show how the UIIR varies at SNR = 10 [dB] for different noise blends with $a = 0.1\ T$ and $a = 0.3\ T$, respectively. Instead of using the UIIR as y-axis, the so-called lineal capacity is show, i.e. the UIIR is multiplied by PW50/T.

We notice that a higher MF, i.e. more media noise, leads in some cases to a higher lineal capacity, in particular at high normalized linear densities PW50/T (see also Fig. 23). Note that no saturation is visible for the lineal capacity, even at PW50/$T = 3.8$.

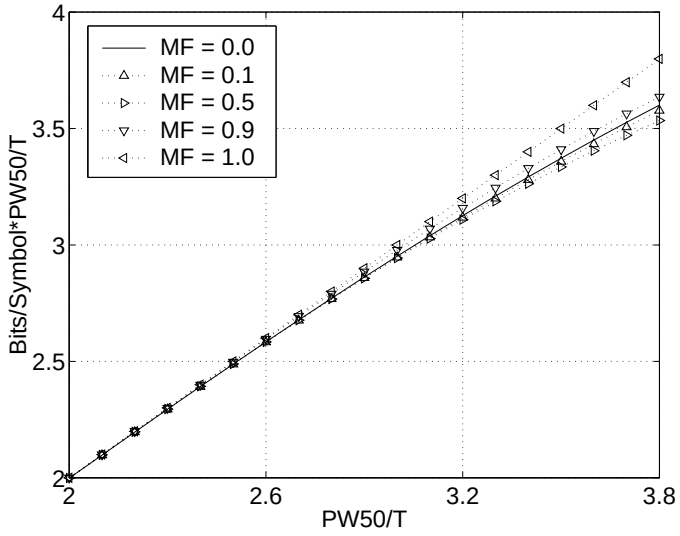

Figure 16: Lineal capacity vs. PW50/T at SNR = 10 [dB] and $a = 0.1\ T$.

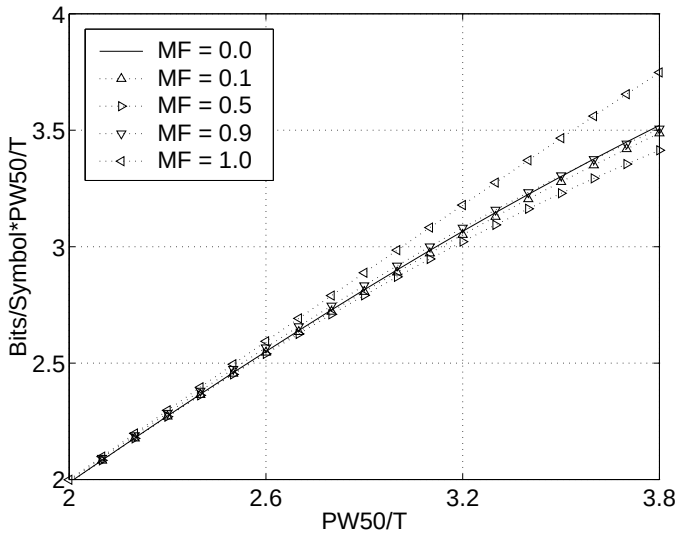

Figure 17: Lineal capacity vs. PW50/T at SNR = 10 [dB] and $a = 0.3\ T$.

5.3.2 Varying a

Again PW50/$T = 3.2$ and the jitter variance is varied from 0.0 to 0.5 T. SNR is fixed to 10 [dB]. Figures 18 and 19 show UIIRs at the normalized linear densities PW50/$T = 2.0$ and PW50/$T = 3.2$, respectively.

Again, we notice that media noise is in some cases preferable to AWGN from an information-theoretic perspective.

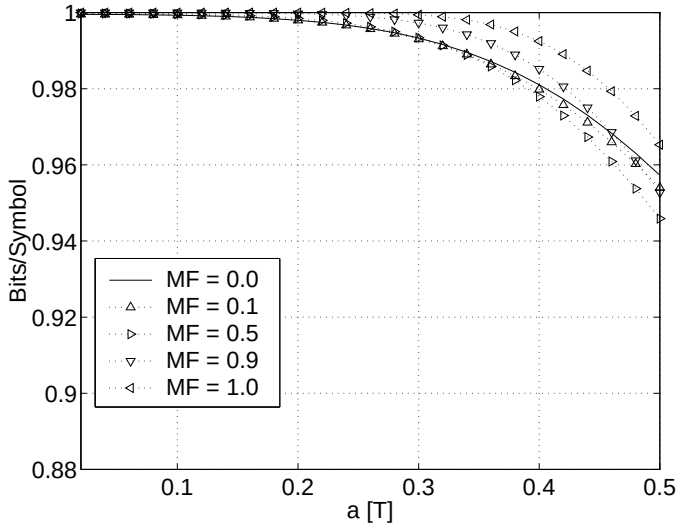

Figure 18: UIIR vs. a for PW50/$T = 2.0$ at SNR = 10 [dB].

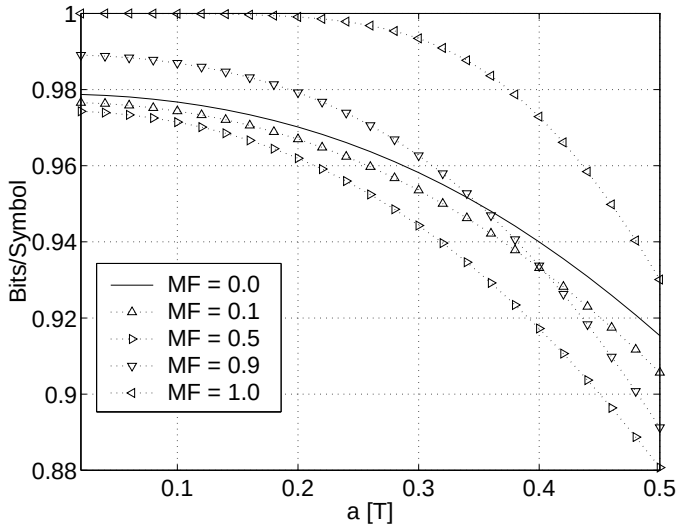

Figure 19: UIIR vs. a for PW50/$T = 3.2$ at SNR = 10 [dB].

5.3.3 Best-case and worst-case MF

In Fig. 20, the best case $\mathrm{MF}_{\mathrm{BC}}$, i.e. the SNR blend that leads to the highest UIIR, is shown for different SNR and varying a. The worst case blend, i.e. $\mathrm{MF}_{\mathrm{WC}}$, is shown in Fig. 21.

We notice that

- the best blend is dependent on the SNR and a;
- at low SNR and high a, AWGN is preferable to media noise. This means that for wide transitions media noise is worse;
- for high SNR and low a, the more media noise the better. This means that for small transitions media noise is preferable.

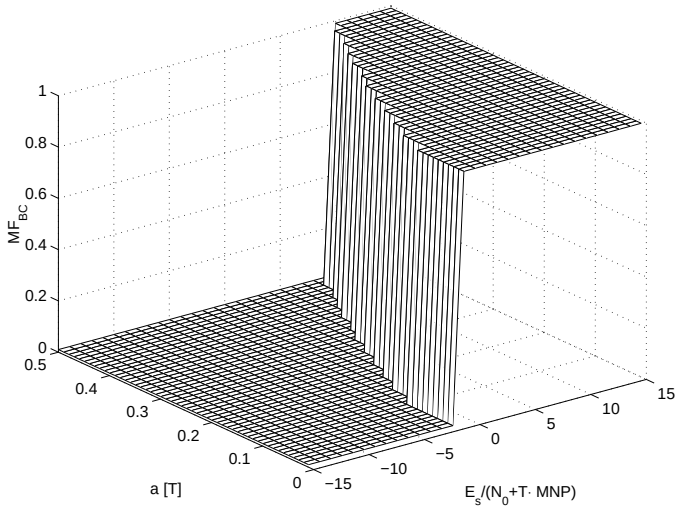

Figure 20: Best-case blend for varying SNR and a at PW50/$T = 3.2$.

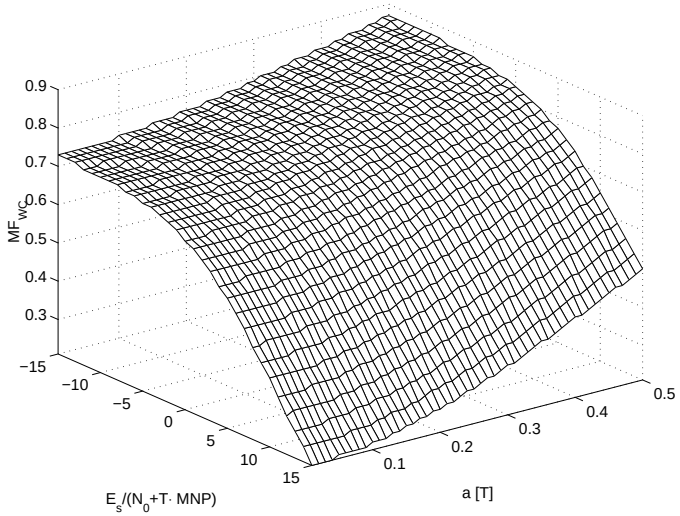

Figure 21: Worst-case blend for varying SNR and a at PW50/$T = 3.2$.

5.4 Comparison of media noise models

In Figures 22 and 23, UIIRs vs. SNR are shown for the microtrack model and various SNR blends at PW50/T = 2.0 and PW50/T = 3.2, respectively. We notice that from a certain SNR point on, media noise is better than AWGN. This SNR point depends on the MF. This is more pronounced at PW50/T = 3.2 than at PW50/T = 2.0.

The UIIRs vs. SNR for the 2OM and NM model at PW50/T = 3.2 and a = 0.3 T are shown in Fig. 24 and Fig. 25 respectively. We observe that the 2OM models leads generally to lower and the NM to higher UIIRs. This becomes more visible from Fig. 26 and Fig. 27 where the UIIRs are shown for MF = 0.0 and MF = 1.0 respectively.

We observe the following:

- For all models media noise is in some cases preferable to AWGN, in particular at high rates;
- The UIIRs for NM model are constantly higher than for the other two models because the signal path is not attenuated with a (see also Eq. 40). This difference can best be seen in Fig. 26, where the UIIRs of the three models for MF = 0.0 are depicted.
- The NM model prefers media noise the most, whereas the 2OM model prefers AWGN the most. The microtrack model is in between. This is evident from Fig. 27, where the UIIRs of the three models for MF = 1.0 are shown.

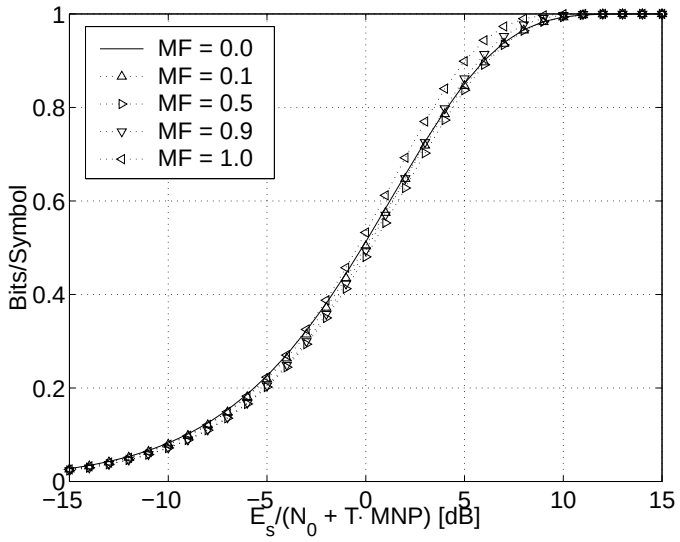

Figure 22: UIIR vs. SNR for various SNR blends at PW50/T = 2.0 and a = 0.3 T for the microtrack model.

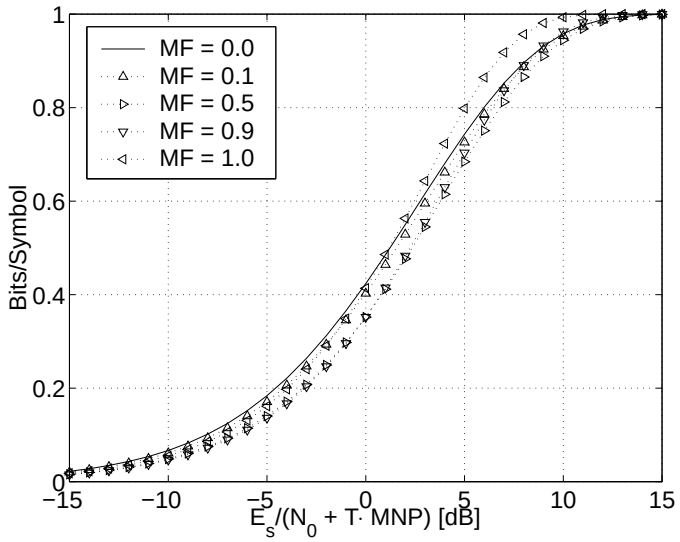

Figure 23: UIIR vs. SNR for various SNR blends at PW50/T = 3.2 and a = 0.3 T for the microtrack model.

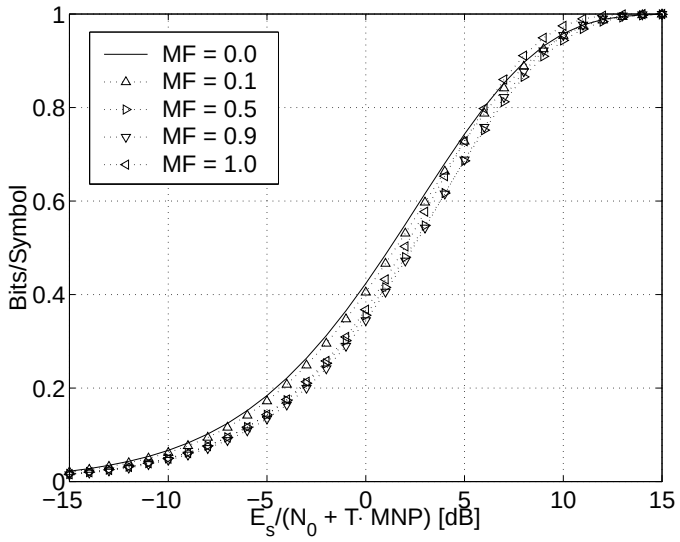

Figure 24: UIIR vs. SNR for various SNR blends at PW50/T = 3.2 and a = 0.3 T for the 2nd Order model.

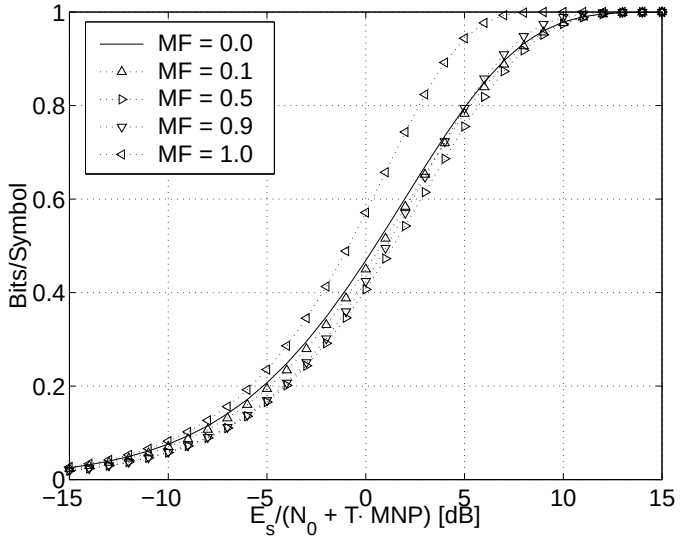

Figure 25: UIIR vs. SNR for various SNR blends at PW50/T = 3.2 and a = 0.3 T for the Nair-Moon model.

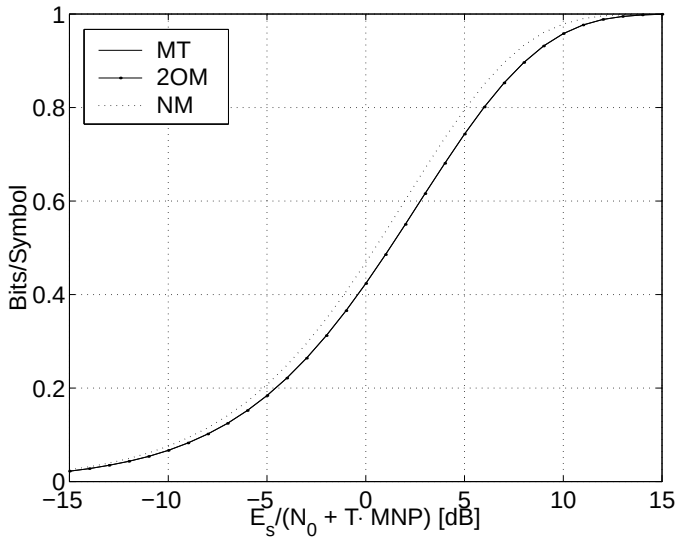

Figure 26: UIIRs vs. SNR for the three different models at PW50/T = 3.2, a = 0.3 T, and MF = 0.0.

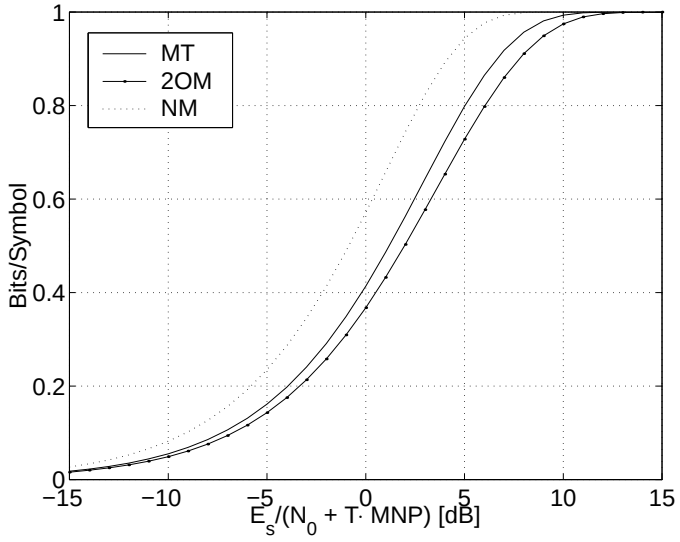

Figure 27: UIIRs vs. SNR for the three different models at PW50/T = 3.2, a = 0.3 T, and MF = 1.0.

5.5 NUIIRs for different MF scenarios

Figure 28 shows UIIRs and NUIIRs for AWGN and media noise-dominated noise scenarios on the microtrack model. We conclude that shaping the input to the channel is only

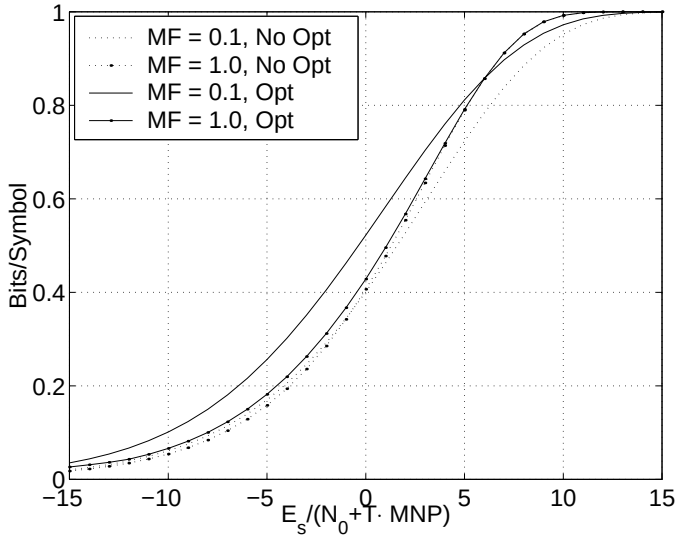

Figure 28: UIIRs and NUIIRs vs. SNR for two blends at PW50/$T = 3.2$, $a = 0.3\ T$.

meaningful in a noise scenario dominated by AWGN. If media noise dominates (i.e. for MF $= 1.0$ and high SNR), this optimization does not lead to a higher UIIR because the noise is already shaped to the channel, i.e. the ratio of signal to noise spectral density is flat. However, if AWGN predominates (i.e. MF$= 0.1$) the ratio of signal to noise spectral density is not flat and we can shape the input, resulting in a substantial gain of the output SNR. This gain in SNR results in a higher information rate, which at a medium-to-low SNR outperforms the UIIRs obtained from the media-noise-dominated noise scenario.

6 Conclusions

The compound magnetic recording channel was modelled as a media noise model followed by a Lorentzian and an AWGN channel. The behavior of this channel is determined by the five parameters PW50/T, N, a, L, and N_0. Three different media noise models were investigated: the microtrack model, the 2nd-order model, and the Nair–Moon model. By noting that all media noise models deliver a cyclostationary output, the average power spectral density at the output of the magnetic recording channel was computed. With help of this average power spectral density, the influence of the five parameters on the achievable information rate was studied by means of the SLLB. Information rates for 0.5-Bernoulli and optimized memory-one Markov processes were computed. We found that

- media noise is in some cases preferable to AWGN noise. This applies in particular for high rates, 0.5-Bernoulli inputs, and a large amount of media noise.
- optimized information rates were also obtained, by considering the spectrum of Markov sources with memory one. When AWGN was the dominating noise source, considerable gain could be observed compared to the i.i.d. information rate. However, when media noise was the dominating noise source, the improvement was negligible. Moreover, the optimized information rate with AWGN outperformed the optimized information rates with media noise at low to medium rates.
- partial signal erasure is detrimental because it reduces the signal power.

Acknowledgements

It is a pleasure to acknowledge many stimulating discussions with Dr. E. Eleftheriou of IBM's Zurich Research Laboratory.

A Proof of cyclostationarity

In order to prove that the noiseless output of the microtrack channel (11) is cyclostationary, we have to show that (i) the mean and (ii) the autocorrelation function are periodic in T (see [37]), i.e. that

$$\begin{aligned} \text{(i)} \quad & \mathrm{E}[y(t)] && = \mathrm{E}[y(t+T)], && (54) \\ \text{(ii)} \quad & \mathrm{E}[y(t)\cdot y(t+\tau)] && = \mathrm{E}[y(t+T)\cdot y(t+\tau+T)], && (55) \end{aligned}$$

where the expectation goes over the data sequence X and the jitter sequence J for the mean and over both the data sequence X at time t and $\tilde{X}$ at time $t+T$ as well as both jitter sequences J at time t and $\tilde{J}$ at time $t+T$ for the autocorrelation function.

Because the waveform is deterministic and periodic in T, we only have to show that the data and jitter distributions are jointly cyclostationary of order two [37]. Noting that the jitter depends on the current data x_k and is on each microtrack i distributed independently of the other microtracks, the joint distribution of data and jitter can be written as

$$p_{XJ} \;=\; \prod_{k=-\infty}^{+\infty} \left[\prod_{i=1}^{N} p_{J_{i,k}|X_k} \right] p_{X_k|X_{k-1}}. \tag{56}$$

Data and jitter form a hidden Markov process. If the hidden process (the data process) is stationary, so is the observable process (jitter process). Therefore condition (i) and (ii) are fulfilled.

B Derivation of the average power spectral density of the microtrack model

Following [19, 21], we can derive the average power spectral density for the microtrack model in a straightforward way.

The noiseless signal at the output of the microtrack model is given by

$$v(t) = \frac{1}{N} \sum_{k=-\infty}^{+\infty} x_k \sum_{i=1}^{N} h(t - kT - j_{i,k}), \tag{57}$$

where $x_k = (u_k - u_{k-1})/\sqrt{2}$. The u_k's are i.i.d. and $\in \{-1, +1\}$ with initial condition $u_{-1} = -1$. Hence, $x_k \in \{-\sqrt{2}, 0, +\sqrt{2}\}$ and correlated. Furthermore, $h(t) = \lambda \cdot 1/(1 + (2t/\mathrm{PW50})^2)$, i.e. $h(.)$ is the Lorentzian pulse, and $j_{i,k}$ is the jitter of the i-th microtrack at the k-th time step.

Now, the autocorrelation function of $v(.)$ is

$$\begin{aligned}
R_v(t+\tau,t) &= \mathrm{E}[v(t+\tau)\cdot v(t)] \\
&= \mathrm{E}\Big[\sum_{k=-\infty}^{+\infty}\sum_{i_1=1}^{N}\frac{x_k}{N}h(t+\tau-kT-j_{i_1,k})\cdot\sum_{\ell=-\infty}^{+\infty}\sum_{i_2=1}^{N}\frac{x_\ell}{N}h(t-\ell T-j_{i_2,\ell})\Big] \\
&= \sum_{k=-\infty}^{+\infty}\sum_{\ell=-\infty}^{+\infty}\mathrm{E}\Big[\sum_{i_1=1}^{N}\frac{x_k}{N}h(t+\tau-kT-j_{i_1,k})\cdot\sum_{i_2=1}^{N}\frac{x_\ell}{N}h(t-\ell T-j_{i_2,\ell})\Big] \\
&= \sum_{k=-\infty}^{+\infty}\sum_{\ell=-\infty}^{+\infty}\frac{1}{N^2}\mathrm{E}\Big[x_k x_\ell\sum_{i_1=1}^{N}h(t+\tau-kT-j_{i_1,k})\cdot\sum_{i_2=1}^{N}h(t-\ell T-j_{i_2,\ell})\Big].
\end{aligned}$$

Because of data-dependency we have to go through all possible cases for x_k and x_ℓ to obtain

$$\begin{aligned}
R_v(t+\tau,t) &= \sum_{k=-\infty}^{+\infty}\sum_{\ell=-\infty}^{+\infty}\frac{1}{N^2}r(k-\ell)\mathrm{E}\Big[\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}h(t+\tau-kT-j_{i_1,k})\cdot h(t-\ell T-j_{i_2,\ell})\Big] \\
&= \sum_{k=-\infty}^{+\infty}\sum_{\ell=-\infty}^{+\infty}\frac{1}{N^2}r(k-\ell)\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\mathrm{E}\Big[h(t+\tau-kT-j_{i_1,k})\cdot h(t-\ell T-j_{i_2,\ell})\Big]
\end{aligned}$$

Substituting $m=k-\ell$ results in

$$\begin{aligned}
R_v(t+\tau,t) &= \sum_{m=-\infty}^{+\infty}r(m)\frac{1}{N^2}\sum_{k=-\infty}^{+\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} \\
&\quad \mathrm{E}\Big[h(t+\tau-kT-j_{i_1,k})\cdot h(t-kT+mT-j_{i_2,k-m})\Big] \\
&= \sum_{m=-\infty}^{+\infty}r(m)\frac{1}{N^2}\sum_{k=-\infty}^{+\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \\
&\quad h(t+\tau-kT-j_{i_1,k})\cdot h(t-kT+mT-j_{i_2,k-m})p_{j_{i_1,k}}p_{j_{i_2,k-m}}dj_{i_1,k}dj_{i_2,k-m}
\end{aligned}$$

where in the last line we used the property that the jitters are independent from one transition to the next. Now with random binary input

$$r(m)=\begin{cases}1 & \text{if } m=0\\ -0.5 & \text{if } m=\pm 1\\ 0 & \text{otherwise,}\end{cases} \tag{58}$$

this yields

$$
\begin{aligned}
R_v(t+\tau,t) &= \frac{r(0)}{N^2}\sum_{k=-\infty}^{+\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \\
&\quad h(t+\tau-kT-j_{i_1,k})\cdot h(t-kT-j_{i_2,k-1})p_{j_{i_1,k}}p_{j_{i_2,k}}dj_{i_1,k}dj_{i_2,k} \\
&\quad +\frac{r(1)}{N^2}\sum_{k=-\infty}^{+\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \\
&\quad h(t+\tau-kT-j_{i_1,k})\cdot h(t-(k-1)T-j_{i_2,k-1})p_{j_{i_1,k}}p_{j_{i_2,k-1}}dj_{i_1,k}dj_{i_2,k-1} \\
&\quad +\frac{r(-1)}{N^2}\sum_{k=-\infty}^{+\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \\
&\quad h(t+\tau-kT-j_{i_1,k})\cdot h(t-(k+1)T-j_{i_2,k+1})p_{j_{i_1,k+1}}p_{j_{i_2,k}}dj_{i_1,k}dj_{i_2,k+1}.
\end{aligned}
$$

The first term is called the *self-jitter*, the other two are symmetric and termed *adjacent-jitter*. Because $R_v(t+\tau,t) = R_v(t+\tau+T,t+T)$, the stochastic process is periodic, i.e. it is wide sense cyclostationary (see also Appendix A).

For the spectrum, we therefore need the average autocorrelation function (i.e. the process must be a wide sense stationary before we can apply the Wiener–Khintchine relation). The average autocorrelation function of $v(.)$ is by definition

$$R_v(\tau) = \frac{1}{T}\int_0^T R_v(t+\tau,t),$$

and with

$$\int_0^T \sum_{k=-\infty}^{+\infty} = \int_{-\infty}^{+\infty}$$

the self-jitter terms becomes

$$\frac{r(0)}{TN^2}\int_{-\infty}^{+\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(t+\tau-kT-j_{i_1,k})\cdot h(t-kT-j_{i_2,k})p_{j_{i_1,k}}p_{j_{i_2,k}}dj_{i_1,k}dj_{i_2,k}dt. \quad (59)$$

The double sum over i_1 and i_2 leads to N^2 summands. N of those result from the autocorrelation within one time step of the N microtracks, i.e. $i_1 = i_2$. This leads to

$$
\begin{aligned}
&\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(t+j-kT-j_{i_1,k})h(t-kT-j_{i_1,k})p_{j_{i_1,k}}p_{j_{i_1,k}}dj_{i_1,k}dj_{i_1,k}dt \quad (60)\\
= &\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(t+j-kT-j_{i_1,k})h(t-kT-j_{i_1,k})p_{j_{i_1,k}}dj_{i_1,k}dt, \quad (61)
\end{aligned}
$$

where the second integral over $j_{i_1,k}$ can be omitted as both integrals are definite. The other $N(N-1)$ terms are crosscorrelations, i.e. $i_1 \neq i_2$, within one time step and have

the following form

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(t+\tau-kT-j_{i_1,k})\cdot h(t-kT-j_{i_1,k})p_{j_{i_1,k}}p_{j_{i_2,k}}dj_{i_1,k}dj_{i_2,k}dt \quad (62)$$

$$=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} h(t+\tau-kT-j_{i_1,k})p_{j_{i_1,k}}dj_{i_1,k}\cdot\int_{-\infty}^{+\infty} h(t-kT-j_{i_2,k})p_{j_{i_2,k}}dj_{i_2,k}\right]dt \quad (63)$$

Let us now introduce the following new terms:

$$R_h(\tau) \triangleq \int_{-\infty}^{+\infty} h(t+\tau)h(t)dt \quad (64)$$

$$\hat{h}(t) \triangleq \int_{-\infty}^{+\infty} h(t-j)p_j dj \quad (65)$$

$$\hat{R}_h(\tau) \triangleq \int_{-\infty}^{+\infty} \hat{h}(t+\tau)\hat{h}(t)dt. \quad (66)$$

The self-jitter term can be written as

$$\frac{r(0)}{TN^2}\left[\sum_{i=1}^{N}\int_{-\infty}^{\infty} R_h(\tau)p_{j_i}dj_i + N(N-1)\int_{-\infty}^{+\infty}\hat{h}(t+\tau)\hat{h}(t)dt\right]. \quad (67)$$

With the new definitions, we obtain

$$\frac{r(0)}{TN^2}\left[N\cdot R_h(\tau)+(N^2-N)\cdot\hat{R}_h(\tau)\right].$$

From the adjacent-jitter term, only crosscorrelations are obtained because the time steps are different. Hence, this leads to

$$\frac{1}{T}\sum_{m\neq 0} r(m)\frac{1}{N^2}\left[N^2\cdot\hat{R}_h(\tau-mT)\right].$$

Therefore the average autocorrelation function of $v(.)$ becomes

$$\bar{R}_v(\tau)=\frac{r(0)}{N^2}\cdot\frac{1}{T}\left[N\cdot R_h(\tau)+(N^2-N)\hat{R}_h(\tau)\right]+\frac{1}{T}\sum_{m\neq 0} r(m)\hat{R}_h(\tau-mT). \quad (68)$$

For the spectrum of the average autocorrelation function, let us introduce the following terms:

$$\begin{aligned}\mathcal{F}\{\hat{h}(t)\} &\triangleq H(f)\cdot P_{\mathrm{J}}(f)\\ \mathcal{F}\{R_h(\tau)\} &\triangleq |H(f)|^2\\ \mathcal{F}\{\hat{R}_h(\tau)\} &\triangleq |H(f)|^2\cdot|P_{\mathrm{J}}(f)|^2.\end{aligned}$$

Furthermore, let us note that

$$\mathcal{F}\Big\{ \sum_m r(m) \cdot \hat{R}_h(\tau - mT) \Big\} = \Phi_X(f) \cdot |H(f)|^2 \cdot |P_{\mathrm{J}}(f)|^2.$$

Then the spectrum of the average autocorrelation function of $v(.)$ becomes

$$\bar{\Phi}_V(f) = \frac{|H(f)|^2}{T} \left[\Phi_X(f) \cdot |P_{\mathrm{J}}(f)|^2 + \frac{r_X(0)}{N} \left[1 - |P_{\mathrm{J}}(f)|^2 \right] \right]. \tag{69}$$

The formula in Eq. (25) is obtained by considering only $N - \varepsilon$ microtracks.

C Useful formulas for the Lorentzian pulse

C.1 Time- and frequency-domain representation

The Lorentzian is given in the time domain by

$$h(t) = \sqrt{\frac{4E_{\mathrm{L}}}{\pi \mathrm{PW50}}} \frac{1}{1 + \left(\frac{2t}{\mathrm{PW50}}\right)^2} \tag{70}$$

such that $\int_{-\infty}^{+\infty} |h(t)^2| dt = E_{\mathrm{L}}$. Its Fourier transform is

$$H(f) = \sqrt{\mathrm{PW50} \pi E_{\mathrm{L}}} e^{-\pi \mathrm{PW50} |f|}. \tag{71}$$

C.2 Energy of the n-th derivative of the Lorentzian

By inspection, we found that

$$\int_{-\infty}^{+\infty} |\frac{\partial^n h(t)}{\partial t}|^2 dt = \left[\prod_{i=1}^{2n} i \right] \frac{E_{\mathrm{L}}}{\mathrm{PW50}^{2n}}. \tag{72}$$

References

[1] R. Wood, "The Feasibility of Magnetic Recording at 1 Terabit per Square Inch," *IEEE Trans. Magnetics*, vol. 36, no. 1, pp. 36–42, Jan. 2000.

[2] A. Taratorin, D. Cheng, and E. Marinero, "Media Noise, Nonlinear Distortions, and Thermal Stability in High Density Recording," *IEEE Trans. Magnetics*, vol. 36, no. 1, pp. 80–85, Jan. 2000.

[3] H. N. Bertram and M. Williams, "SNR and Density Limit Estimates: A Comparison of Longitudinal and Perpendicular Recording," *IEEE Trans. Magnetics*, vol. 36, no. 1, pp. 4–9, Jan. 2000.

[4] A. Kavčić and A. Parapoutian, "A Signal-Dependent Autoregressive Channel Model," *in Proceedings of IEEE INTERMAG Conference*, Kyongju, Korea, May 1999.

[5] S. M. Yuan and H. N. Bertram, "Statistical Data Analysis of Magnetic Recording Noise Mechanism," *IEEE Trans. Magnetics*, vol. 28, pp. 84–92, Jan. 1992.

[6] R. Wood, "Detection and Capacity Limits in Magnetic Media Noise," *IEEE Trans. Magnetics*, vol. 34, no.4, pp. 1848–1850, July 1998.

[7] C. A. French and J. K. Wolf, "Bounds on Capacity of a Peak Power Constrained Gaussian Channel," *IEEE Trans. Magnetics*, vol. 24, no. 5, pp. 2247–2262, Sept. 1998.

[8] T. R. Oenning and J. Moon, "The Effect of Jitter Noise on Binary Input Intersymbol Interference Channel Capacity," *in Proc. of IEEE Int. Conference on Communications*, vol. 8, pp. 2416–2420, Helsinki, Finland, June 11 - 14, 2001.

[9] Sh. Shamai and R. Laroia, "The Intersymbol Interference Channel: Lower Bounds on Capacity and Channel precoding Loss," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1388–1404, Sept. 1996.

[10] J. P. Caroselli Jr., "Modeling, Analysis, and Mitigation of Medium Noise in Thin Film Magnetic Recording Channels," *Ph.D. Thesis*, University of California, San Diego, 1998.

[11] S. K. Nair, H. Shafiee, and J. Moon, "Modeling and Simulation of Advanced Read Channels," *IEEE Trans. Magnetics*, vol. 29, no. 6, pp. 4056–4058, Nov. 1993.

[12] D. Hoesli and E. Svensson, "Low-Density Parity-Check Codes for Magnetic Recording," Diploma Thesis no. 7103, Signal and Information Processing Lab., ETH Zurich, Switzerland, March, 2000.

[13] A. Dholakia, E. Eleftheriou, and T. Mittelholzer, "On Iterative Decoding for Magnetic Recording Channels," *in Proc. 2nd Int. Symp. on Turbo Codes & Related Topics*, pp. 219–225, Brest, France, September 4 - 7, 2000.

[14] G. Stromberg, "Signal Space Detection with Application to Magnetic Recording," *Ph.D. Thesis*, University of Dortmund, Sept. 2000.

[15] H. N. Bertram, "The Theory of Magnetic Recording," *Cambridge University Press*, April 1994.

[16] M. L. Williams and R. L. Comstock, "An Analytical Model for the Write Process in Digital Magnetic Recording," *Proc. AIP Conf.*, vol. 5, p. 738, 1971.

[17] J. S. Goldberg and J. K. Wolf, "Implementation and Analysis of Nonlinear Effects in the Microtrack Model," *IEEE Trans. Magnetics*, vol. 35, no. 5, pp. 2256–2258, Sept. 1999.

[18] G. D. Forney Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interfernce," *IEEE Trans. Inform. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[19] J. G. Proakis, *Digital Communications*, New York: McGraw-Hill, 1989.

[20] T. R. Oenning, "Channel Capacity and Coding for Magnetic Recording Channels with Media Noise," *Ph.D. Thesis*, University of Minnesota, Sept. 2000.

[21] C. P. M. J. Baggen, "An Information Theoretic Approach to Timing Jitter," *Ph.D. Thesis*, University of California at San Diego, 1993.

[22] J. Moon, "Signal-to-Noise Ratio Definition for Magnetic Recording Channels with Transition Noise," *IEEE Trans. Magnetics*, vol. 35, no. 5, part 2, pp. 3881–3883, Sept. 2000.

[23] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for Digital Recorders," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2260–2299, Oct. 1998.

[24] D. Arnold and H.-A. Loeliger, "On the Information Rate of Binary-Input Channels with Memory," *in Proc. of IEEE Int. Conf. on Communications*, vol. 9, pp. 2692–2695, Helsinki, Finland, June 11 - 14, 2001.

[25] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the Achievable Information Rates of Finite State ISI Channels," *Proc. GLOBECOM*, 2001.

[26] S. Yang and A. Kavčić, "On the Capacity of Data-Dependent Autoregressive Noise Channels," to be presented at INTERMAG 2002, Amsterdam, The Netherlands, April 28 - May 2, 2002.

[27] D. Arnold and H.-A. Loeliger, "Computing Information Rates of the Microtrack Channel," to be presented at INTERMAG 2002, Amsterdam, The Netherlands, April 28 - May 2, 2002.

[28] D. Arnold and H.-A. Loeliger, "On Finite-State Information Rates from Channel Simulations," to be presented at ISIT 2002, Lausanne, Switzerland, June 30 - July 5, 2002.

[29] Sh. Shamai, L. H. Ozarow, and A. D. Wyner, "Information Rates of a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs," *IEEE Trans. Inform. Theory*, vol. 37, no. 6, Nov. 1991.

[30] W. Feller, "An Introduction to Probability Theory and Its Application," vol. II, New York: John Wiley & Sons Inc., 1966.

[31] M. Peligrad and S. Utev, "Central Limit Theorem for Linear Processes," *Annals of Probability*, vol. 25, no. 1, pp. 443–456, Jan. 1997.

[32] Sh. Shamai and S. Verdú, "Worst-Case Power-Constrained Noise for Binary-Input Channels," *IEEE Trans. Inform. Theory*, vol. 38, no. 5, pp. 1494–1511, Sept. 1992.

[33] G. Bilardi, R. Padovani, and G. L. Periobon, "Spectral Anaysis of Functions of Markov Chains with Applications," *IEEE Trans. Commun.*, vol. 31, no. 7, pp. 853–861, July 1983.

[34] Sh. Shamai and I. Bar-David, "Upper Bounds on Capacity for a Constrained Gaussian Channel," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1079–1084, Sept. 1989.

[35] Z.-N. Wu, S. Lin, and J. M. Cioffi, "Numerical Results on Capacity Bounds for Magnetic Recording," *Proc. GLOBECOM 98*, pp. 3385–3390.

[36] J. Justensen, "Information Rates and Power Spectra of Digital Codes," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 457–472, May 1982.

[37] W. A. Gardner and L. E. Franks, "Characterization of Cyclostationary Random Signal Processes," *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 4–14, Jan. 1975.