On the random coding exponent of nonlinear gaussian channels

Kostis Xenoulis
Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
Athens, Greece 15784
Email: kxen@di.uoa.gr

Nicholas Kalouptsidis
Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
Athens, Greece 15784
Email: kalou@di.uoa.gr

***Abstract*—A random coding theorem for nonlinear additive gaussian channels is presented. Modeling the channel's nonlinear behavior as a causal, stationary Volterra system and under maximum likelihood decoding, an upper bound on the average error probability is obtained. The proposed bound is deduced by deploying exponential martingale inequalities. Cubic nonlinearities are used as example to illustrate the validity of the random coding exponent and the usefulness of the proposed technique in the treatment of nonlinear channels.**

I. Introduction

Various communication channels, including satellite and wireless, exhibit nonlinear behavior that degrades the quality of information transmission [1, Chap.14]. In satellite communications, power amplifiers usually operate near saturation for power efficiency and thus introduce nonlinear distortion. Likewise, in mobile cellular systems, power amplifiers in hand held terminal are forced to operate in nonlinear region to secure high power efficiency. The main obstacle in obtaining calculable, closed form upper bounds on the average error decoding probability of nonlinear channels stems mainly from the difficulty of defining the channel's output probability density function. Thus the general capacity formula in [2] cannot be easily employed. The aforementioned difficulty is strengthened by the fact that the nonlinear channel is not generally memoryless and thus moment generating functions appearing in Chernoff like bounds are hard to compute. Moreover, the calculation of the mutual information is hard, and no useful bounds can be obtained for the channel capacity.

Preliminary work regarding the random coding exponents of nonlinear gaussian channels is reported, among others, in [3], where the finite-state approach and the Perron-Frobenius theorem are utilized properly. The present work differs from previous by alleviating the aforementioned difficulties with the proper exploitation of the theory of martingale processes [4]. The latter are able to provide tight concentration inequalities for multivariate functions with memory and thus can lead to tight bounds on the average error decoding probability of nonlinear channels. Moreover, martingale inequalities are able to provide solutions even in non-stationary communication environments.

In what follows, we consider the transmission of an arbitrary set of messages $\mathcal{M}$ with cardinality $|\mathcal{M}|$ over a continuous output nonlinear additive gaussian noise channel. The nonlinear channel has the form [5]

$$\boldsymbol{y} = D\boldsymbol{u} + \boldsymbol{\nu} \tag{1}$$

where $\boldsymbol{y}$ is the channel output vector, $\boldsymbol{u}$ and $\boldsymbol{\nu}$ the channel input and noise vectors respectively, all of length N. The components of the noise vector $\boldsymbol{\nu}$ are i.i.d gaussian random variables with zero mean value and variance σ^2. D is a causal, stationary Volterra system with finite memory μ applied to the channel's input vectors. Thus,

$$D\boldsymbol{u} = (D\boldsymbol{u}(1), \dots, D\boldsymbol{u}(N)) \tag{2}$$

where

$$\begin{aligned} D\boldsymbol{u}(i) = h_0 + \sum_{j=1}^{\mu} \sum_{i_1=0}^{\mu} \dots \sum_{i_j=0}^{\mu} h_j(i_1, \dots, i_j) \times \\ \times u(i-i_1) \dots u(i-i_j) \end{aligned} \tag{3}$$

All Volterra kernels h_j are known and bounded. It is clear [6] that

$$\|D\boldsymbol{u}\|_\infty \leq g(\|\boldsymbol{u}\|_\infty) \tag{4}$$

where

$$g(x) = |h_0| + \sum_{j=1}^{\mu} \|h_j\| x^j, \quad x \geq 0 \tag{5}$$

and

$$\|h_j\| = \sum_{i_1=0}^{\mu} \dots \sum_{i_j=0}^{\mu} |h_j(i_1, \dots, i_j)| \tag{6}$$

II. Error Probability Analysis

Following the same setup as in [7, sec.5.5], let for each message m, a codeword $\boldsymbol{u}_m$ of length N be selected randomly from the ensemble of (N, R) block codes and transmitted with rate R over the nonlinear channel (1). All codewords $\boldsymbol{u}_m$, $m = 1, \dots, |\mathcal{M}| = \lceil \exp(NR) \rceil$ ($|\mathcal{M}|$ the cardinality of the set of messages $\mathcal{M}$) are chosen independently with the same probability Q from the ensemble. Under the assumption that maximum-likelihood decoding is performed at the output of the channel, an error occurs when, given the transmitted

978-1-4244-4536-3/09/$25.00 ©2009 IEEE

message m and the received vector $\boldsymbol{y}$, there exists another message $m' \neq m$ such that

$$\Pr(\boldsymbol{y}|\boldsymbol{u}_{m'}) \geq \Pr(\boldsymbol{y}|\boldsymbol{u}_m) \tag{7}$$

or equivalently,

$$\|\boldsymbol{y} - D\boldsymbol{u}_{m'}\|_2^2 \leq \|\boldsymbol{y} - D\boldsymbol{u}_m\|_2^2 \tag{8}$$

Since m is the transmitted message, $\boldsymbol{y} = D\boldsymbol{u}_m + \boldsymbol{\nu}$, so that the error criterion (8) is translated into

$$\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'} + \boldsymbol{\nu}\|_2^2 \leq \|\boldsymbol{\nu}\|_2^2 \tag{9}$$

Thus, in the ML decoding setup, given $m, \boldsymbol{u}_m, \boldsymbol{\nu}$, an error occurs if there exists another message m' such that the codeword $\boldsymbol{u}_{m'}$ is selected in such a way that (9) is satisfied. Consequently, the average ML error decoding probability

$$\overline{P}_{e,m} = \sum_{\boldsymbol{u}_m} Q(\boldsymbol{u}_m) \int_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{u}_m) \Pr(\text{error}|m, \boldsymbol{u}_m, \boldsymbol{y}) d\boldsymbol{y} \tag{10}$$

is equivalently expressed as

$$\overline{P}_{e,m} = \sum_{\boldsymbol{u}_m} Q(\boldsymbol{u}_m) \int_{\boldsymbol{\nu}} f(\boldsymbol{\nu}) \Pr(\text{error}|m, \boldsymbol{u}_m, \boldsymbol{\nu}) d\boldsymbol{\nu} \tag{11}$$

where $f(\boldsymbol{\nu})$ is the probability density of the noise vector $\boldsymbol{\nu}$. Moreover, if, for $\epsilon > 0$,

$$\boldsymbol{V} = \{\boldsymbol{\nu} : \|\boldsymbol{\nu}\|_2^2 \leq N(\sigma^2 + \epsilon)\}$$

and $\boldsymbol{V}'$ the corresponding complementary set, then

$$\begin{aligned}\overline{P}_{e,m} = &\sum_{\boldsymbol{u}_m} Q(\boldsymbol{u}_m) \int_{\boldsymbol{\nu} \in \boldsymbol{V}} f(\boldsymbol{\nu}) \Pr(\text{error}|m, \boldsymbol{u}_m, \boldsymbol{\nu}) d\boldsymbol{\nu} \\ &+ \sum_{\boldsymbol{u}_m} Q(\boldsymbol{u}_m) \int_{\boldsymbol{\nu} \in \boldsymbol{V}'} f(\boldsymbol{\nu}) \Pr(\text{error}|m, \boldsymbol{u}_m, \boldsymbol{\nu}) d\boldsymbol{\nu}\end{aligned} \tag{12}$$

or

$$\begin{aligned}\overline{P}_{e,m} \leq &\sum_{\boldsymbol{u}_m} Q(\boldsymbol{u}_m) \int_{\boldsymbol{\nu} \in \boldsymbol{V}} f(\boldsymbol{\nu}) \Pr(\text{error}|m, \boldsymbol{u}_m, \boldsymbol{\nu}) d\boldsymbol{\nu} \\ &+ \int_{\boldsymbol{\nu} \in \boldsymbol{V}'} f(\boldsymbol{\nu}) d\boldsymbol{\nu}\end{aligned} \tag{13}$$

Additionally, due to the union bound,

$$\begin{aligned}&\Pr(\text{error}|m, \boldsymbol{u}_m, \boldsymbol{\nu}) \leq \\ &\Pr\left(\bigcup_{m' \neq m} \left\{ \|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'} + \boldsymbol{\nu}\|_2^2 \leq \|\boldsymbol{\nu}\|_2^2 \right\} \middle| m, \boldsymbol{u}_m, \boldsymbol{\nu} \right) \\ &\leq \sum_{m' \neq m} \Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'} - \boldsymbol{\nu}\|_2^2 \leq \|\boldsymbol{\nu}\|_2^2 \ |m, \boldsymbol{u}_m, \boldsymbol{\nu} \right)\end{aligned} \tag{14}$$

where utilizing the triangle inequality

$$\begin{aligned}&\Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'} - \boldsymbol{\nu}\|_2^2 \leq \|\boldsymbol{\nu}\|_2^2 \ |m, \boldsymbol{u}_m, \boldsymbol{\nu} \right) = \\ &\Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'} - \boldsymbol{\nu}\|_2 \leq \|\boldsymbol{\nu}\|_2 \ |m, \boldsymbol{u}_m, \boldsymbol{\nu} \right) \leq \\ &\Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2 - \|\boldsymbol{\nu}\|_2 \leq \|\boldsymbol{\nu}\|_2 \ |m, \boldsymbol{u}_m, \boldsymbol{\nu} \right) = \\ &\Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \leq 4\|\boldsymbol{\nu}\|_2^2 \ |m, \boldsymbol{u}_m, \boldsymbol{\nu} \right)\end{aligned} \tag{15}$$

Combining (13), (14) and (15), the average ML error decoding probability is upper bounded by

$$\begin{aligned}\overline{P}_{e,m} \leq &\sum_{m' \neq m} \sum_{\boldsymbol{u}_m} Q(\boldsymbol{u}_m) \int_{\boldsymbol{\nu} \in \boldsymbol{V}} f(\boldsymbol{\nu}) \times \\ &\Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \leq 4\|\boldsymbol{\nu}\|_2^2 \middle| m, \boldsymbol{u}_m, \boldsymbol{\nu} \right) d\boldsymbol{\nu} \\ &+ \Pr(\boldsymbol{\nu} \in \boldsymbol{V}')\end{aligned} \tag{16}$$

But when $\boldsymbol{\nu} \in \boldsymbol{V}$, it holds $\|\boldsymbol{\nu}\|_2^2 \leq N(\sigma^2 + \epsilon)$, so that

$$\begin{aligned}&\Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \leq 4\|\boldsymbol{\nu}\|_2^2 \middle| m, \boldsymbol{u}_m, \boldsymbol{\nu} \right) \\ &\leq \Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \leq 4N(\sigma^2 + \epsilon) \middle| m, \boldsymbol{u}_m, \boldsymbol{\nu} \right)\end{aligned} \tag{17}$$

Therefore, from (16) and (17)

$$\begin{aligned}\overline{P}_{e,m} \leq &\sum_{m' \neq m} \Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \leq 4N(\sigma^2 + \epsilon) \right) \\ &+ \Pr(\boldsymbol{\nu} \in \boldsymbol{V}')\end{aligned} \tag{18}$$

The second summand in the right hand side of (18) equals

$$\Pr(\boldsymbol{\nu} \in \boldsymbol{V}') = \Pr\left(\|\boldsymbol{\nu}\|_2^2 > N(\sigma^2 + \epsilon) \right) \tag{19}$$

Since $\|\boldsymbol{\nu}\|_2^2$ is a sum of i.i.d random variables, then according to the Chernoff bounding technique [8, Th. 3.2], there exists a positive constant κ such that

$$\Pr\left(\|\boldsymbol{\nu}\|_2^2 > N(\sigma^2 + \epsilon) \right) \leq \exp(-N\kappa) \tag{20}$$

Let $\{Y_i\}_{i=1}^N$ denote the martingale difference sequence of the random variable $\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2$ with respect to the joined filter $\{F_i\}_{i=0}^N$ (Doob's martingale [9]), where

$$\emptyset = F_0 \subseteq F_1 \subseteq \ldots \subseteq F_N$$

$$F_i = \{\boldsymbol{u}_m(1), \ldots, \boldsymbol{u}_m(i), \boldsymbol{u}_{m'}(1), \ldots, \boldsymbol{u}_{m'}(i)\} \tag{21}$$

and

$$\begin{aligned}X_i &= E\left[\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \,\middle|\, F_i \right] \\ X_0 &= E\left[\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \right] \\ X_N &= \|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \\ Y_i &= X_i - X_{i-1}\end{aligned} \tag{22}$$

$\boldsymbol{u}_m(j), \boldsymbol{u}_{m'}(j)$ denote the j-th components of the corresponding vectors $\boldsymbol{u}_m, \boldsymbol{u}_{m'}$ and causality is assumed. The mean values appearing in (20) are with respect to the different codewords $\boldsymbol{u}_m, \boldsymbol{u}_{m'}$ the messages m, m' respectively can be encoded onto. Noting that

$$\sum_{i=1}^N Y_i = \|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 - E\left[\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \right] \tag{23}$$

the probability appearing in the sum in the right hand side of (18) equals

$$\begin{aligned}&\Pr\left(\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \leq 4N(\sigma^2 + \epsilon) \right) = \\ &\Pr\left(\sum_{i=1}^N Y_i \leq -\left(E\left[\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2 \right] - 4N(\sigma^2 + \epsilon) \right) \right)\end{aligned} \tag{24}$$

Moreover, let

$$\mathcal{K}_D(Q,N) = E[\|D\boldsymbol{u}\|_2^2] - \|E\left[D\boldsymbol{u}\right]\|_2^2 \tag{25}$$

Then, due to the random coding setup and the independency of the ensemble's codewords, it holds

$$-E_{\boldsymbol{u}_m,\boldsymbol{u}_{m'}}\left[\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2\right] = -2\mathcal{K}_D(Q,N) \tag{26}$$

Finally, combining (18)-(20) and (24)-(26)

$$\begin{aligned}&\overline{P}_{e,m} \le \exp(-N\kappa)+\\ &\sum_{m'\neq m} \Pr\left(\sum_{i=1}^{N} Y_i \le -2\left(\mathcal{K}_D(Q,N) - 2N(\sigma^2+\epsilon)\right)\right)\end{aligned} \tag{27}$$

Lemma 1: Assume that the components of all codewords $\boldsymbol{u}_m$, $1 \le m \le \mathcal{M}$ are independent and identically distributed random variables and let $r > 0$ such that

$$\max_{1\le m\le \mathcal{M}} \|\boldsymbol{u}_m\|_\infty < r \tag{28}$$

Then, the absolute value $|Y_i|$ of the martingale differences Y_i (20) is upper bounded as

$$|Y_i| \le 4(\mu+1)g(r)^2, \quad i = 1,\dots,N \tag{29}$$

Proof: Let

$$\boldsymbol{d}_{m,m'} = D\boldsymbol{u}_m - D\boldsymbol{u}_{m'} \tag{30}$$

Then, Y_i equals

$$Y_i = E\left[\|\boldsymbol{d}_{m,m'}\|_2^2 \;|F_i\right] - E\left[\|\boldsymbol{d}_{m,m'}\|_2^2 \;|F_{i-1}\right] \tag{31}$$

All samples $\boldsymbol{d}_{m,m'}(j)$, $j \ge i+1+\mu$ are independent from the random variables generating F_i due to the i.i.d. assumption on the input. Likewise, all samples $\boldsymbol{d}_{m,m'}(j)$, $j \le i-1$ do not depend on the i-th component of the filter F_i due to causality. Thus, the first summand in the right hand side of (31) equals

$$\begin{aligned}&E\left[\sum_{j=1}^{N}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_i\right] = E\left[\sum_{j=1}^{i-1}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_i\right] +\\ &E\left[\sum_{j=i}^{i+\mu}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_i\right] + E\left[\sum_{j=i+\mu+1}^{N}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_i\right]\\ =&E\left[\sum_{j=1}^{i-1}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_{i-1}\right] + E\left[\sum_{j=1}^{i+\mu}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_i\right] +\\ &E\left[\sum_{j=i+\mu+1}^{N}(\boldsymbol{d}_{m,m'}(j))^2\right]\end{aligned} \tag{32}$$

In a similar manner, the second summand in the right hand side of (31) satisfies

$$\begin{aligned}&E\left[\sum_{j=1}^{N}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_{i-1}\right] = E\left[\sum_{j=1}^{i-1}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_{i-1}\right] +\\ &E\left[\sum_{j=i}^{i+\mu-1}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_{i-1}\right] + E\left[\sum_{j=i+\mu}^{N}(\boldsymbol{d}_{m,m'}(j))^2\right]\end{aligned} \tag{33}$$

Combining (31)-(33)

$$\begin{aligned}Y_i =&E\left[\sum_{j=i}^{i+\mu}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_i\right] -\\ &E\left[\sum_{j=i}^{i+\mu-1}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_{i-1}\right] - E\left[(\boldsymbol{d}_{m,m'}(i+\mu))^2\right]\end{aligned} \tag{34}$$

Noting that the sample $\boldsymbol{d}_{m,m'}(i+\mu)$ is independent of the filter F_{i-1}

$$E\left[(\boldsymbol{d}_{m,m'}(i+\mu))^2\right] = E\left[(\boldsymbol{d}_{m,m'}(i+\mu))^2 \;|F_{i-1}\right] \tag{35}$$

so that from (34) and (35) it holds

$$Y_i = E\left[\sum_{j=i}^{i+\mu}(\boldsymbol{d}_{m,m'}(j))^2\Big|F_i\right] - E\left[\sum_{j=i}^{i+\mu}(\boldsymbol{d}_{m,m'}(j))^2\Big|F_{i-1}\right] \tag{36}$$

Under the assumption that every codeword $\boldsymbol{u}_m$, $1 \le m \le |\mathcal{M}|$ satisfies (28),

$$|D\boldsymbol{u}_m(j)| \le \|D\boldsymbol{u}_m\|_\infty \le g(\|\boldsymbol{u}_m\|_\infty) \le g(r) \tag{37}$$

so that

$$\begin{aligned}(\boldsymbol{d}_{m,m'}(j))^2 =&(D\boldsymbol{u}_m(j) - D\boldsymbol{u}_{m'}(j))^2 \le \|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_\infty^2\\ \le&(\|D\boldsymbol{u}_m\|_\infty + \|D\boldsymbol{u}_{m'}\|_\infty)^2 = 4g(r)^2\end{aligned} \tag{38}$$

Consequently from (36) and (38)

$$|Y_i| \le E\left[\sum_{j=i}^{i+\mu}(\boldsymbol{d}_{m,m'}(j))^2 \;\Big|F_i\right] < 4(\mu+1)g(r)^2 \tag{39}$$

■

The bound of lemma 1 is rather loose. Tighter bounds can be obtained analytically for Volterra systems D of short memory, as shown in section III. Nevertheless, the bound in (29) applies to general finite memory Volterra systems D and thus leads to a random coding upper bound on the average ML error decoding probability.

Theorem 1: Under the assumptions stated above, let ϵ, κ be positive constants and suppose that

$$\mathcal{K}_D(Q,N) \ge 2N(\sigma^2+\epsilon) \tag{40}$$

Then,

$$\overline{P}_{e,m} \le e^{-N\left(E_c(Q,N,D,\sigma^2)-R\right)} + e^{-N\kappa} \tag{41}$$

where

$$E_c(Q,N,D,\sigma^2) = \frac{1}{8}\left(\frac{\frac{1}{N}\mathcal{K}_D(Q,N) - 2\left(\sigma^2+\epsilon\right)}{2\left(\mu+1\right)g(r)^2}\right)^2 \tag{42}$$

Proof: The second summand in the right hand side of (27) becomes a martingale concentration inequality if criterion (40) is satisfied. Specifically, according to Hoeffding's inequality

[9, th. 3.10], if $\{Y_i\}_{i=1}^N$ is a bounded martingale difference sequence, $|Y_i| \leq c_i$, then for every $t \geq 0$

$$\Pr\left(\sum_{i=1}^{N} Y_i \leq -t\right) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{N} c_i^2}\right) \quad (43)$$

Thus, combining (27), (43) and lemma 1

$$\overline{P}_{e,m} \leq \sum_{m' \neq m} \exp\left(-\frac{1}{2}\frac{\left(2\left(\mathcal{K}_D(Q,N) - 2\hat{V}\right)\right)^2}{N\left(4\left(\mu+1\right)g(r)^2\right)^2}\right) + e^{-N\kappa}$$
$$\leq \exp\left(NR - \frac{N}{8}\frac{\left(\frac{1}{N}\mathcal{K}_D(Q,N) - 2(\sigma^2+\epsilon)\right)^2}{\left(\left(\mu+1\right)g(r)^2\right)^2}\right) + e^{-N\kappa} \quad (44)$$

■

Tighter upper and lower bounds on the martingale differences $|Y_i|$,$1 \leq i \leq N$ in relation with tight exponential martingale inequalities [9] can lead to tighter lower bounds on the capacity of nonlinear gaussian channels and are currently studied.

III. Application

Under the assumptions of theorem 1, we consider the transmission of information over the nonlinear channel

$$y_i = \alpha + \beta u_i + \gamma u_i u_{i-1} u_{i-2} + \nu_i, \quad i = 1, \ldots, N \quad (45)$$

with $E[\nu_i] = 0$ and $E\left[\nu_i^2\right] = \sigma^2$. Then,

$$g(x) = |\alpha| + |\beta|x + |\gamma|x^3, \quad x \geq 0 \quad (46)$$

Let the components of each codeword u_i, $1 \leq i \leq N$ be chosen independently and equiprobably from the set $\{-A, A\}$ with probability $\{1/2, 1/2\}$. Then $\|u\|_\infty = A$. The martingale difference sequence $\{Y_i\}_{i=1}^N$ (22) becomes using (36)

$$Y_i = E\left[\sum_{j=i}^{i+2} \boldsymbol{d}_{\boldsymbol{u},\tilde{\boldsymbol{u}}}(j) \,\Big|F_i\right] - E\left[\sum_{j=i}^{i+2} \boldsymbol{d}_{\boldsymbol{u},\tilde{\boldsymbol{u}}}(j) \,\Big|F_{i-1}\right] \quad (47)$$

where

$$\boldsymbol{d}_{\boldsymbol{u},\tilde{\boldsymbol{u}}}(j) = \left(\beta u_j + \gamma u_j u_{j-1} u_{j-2} - \beta\tilde{u}_j - \gamma\tilde{u}_j\tilde{u}_{j-1}\tilde{u}_{j-2}\right)^2 \quad (48)$$

$u_j, \tilde{u}_j$, $j = 1, \ldots, N$ are the components of two different codewords $\boldsymbol{u}, \tilde{\boldsymbol{u}}$. Regarding the mean value in (26), due to the causality assumption, $u_j = \tilde{u}_j = 0$ for $j < 1$, so that,

$$E\left[\left(\boldsymbol{d}_{\boldsymbol{u},\tilde{\boldsymbol{u}}}(j)\right)^2\right] = 2\beta^2 E\left[u^2\right], \quad j \leq 2 \quad (49)$$

Moreover, for $j \geq 3$, due to the independence of the codewords' components

$$E\left[\left(\boldsymbol{d}_{\boldsymbol{u},\tilde{\boldsymbol{u}}}(j)\right)^2\right] = \Big(2\beta^2 E\left[u^2\right] + 2\gamma^2 E\left[u^2\right]^3 -$$
$$2\beta^2 E[u]^2 + 4\beta\gamma E\left[u^2\right] E[u]^2 - 4\beta\gamma E[u]^4 - 2\gamma^2 E[u]^6\Big) \quad (50)$$

Noting that $E[u] = 0$ and $E\left[u^2\right] = A^2$, and combining (49) and (50), the mean value in (23) equals

$$2\mathcal{K}_D(Q,N) = 2\left(2\beta^2 A^2\right) + 2(N-2)\left(\beta^2 A^2 + \gamma^2 A^6\right) \quad (51)$$

Consequently the criterion (40) is translated into

$$\beta^2 A^2 + \frac{N-2}{N}\gamma^2 A^6 > 2\left(\sigma^2 + \epsilon\right) \quad (52)$$

For large values of code length N, the term $(N-2)/N$ approaches 1 and does not affect condition (52), necessary for the application of Hoeffding's inequality. In the simulations discussed below, it is assumed that the aforementioned term does not contribute to the random coding exponent. Closed formed expressions for the upper and lower bounds of the martingale differences (47) are difficult to obtain, but can easily be calculated for specific values of β, γ and A. For the specific example, the martingale differences Y_i, $1 \leq i \leq N$ can take 16 different patterns. The minimum and maximum values of the corresponding patterns are depicted in table I for various values of A and are shown to be symmetric around 0. Moreover, the looseness of the bound of lemma 1 for the specific communication channel is made evident through the same table. Thus, employing the bound $|Y_i| \leq \mathcal{C}(A, \beta, \gamma)$, for $1 \leq i \leq N$, in the proof of theorem 1, instead of the bound derived in lemma 1, we obtain using (51) and (52)

$$\overline{P}_{e,m} \leq \exp\left(-N\left(E_c'(A, \beta, \gamma, \sigma^2) - R\right)\right) + e^{-N\kappa} \quad (53)$$

where

$$E_c'(A, \beta, \gamma, \sigma^2) = 2\left(\frac{\beta^2 A^2 + \gamma^2 A^6 - 2(\sigma^2 + \epsilon)}{\mathcal{C}(A, \beta, \gamma)}\right)^2 \quad (54)$$

Note that the constant term α in (45) does not participate neither in the random coding exponent $E_c'(A, \beta, \gamma, \sigma^2)$ nor in the error decoding probability upper bound (54). The interaction among the noise variance σ^2, the Volterra coefficients (β, γ) and the characteristic of the constellation A, as well as their effect on the random coding exponent $E_c'(A, \beta, \gamma, \sigma^2)$, are presented in figs. 1,2. Fig.1 shows the minimum values of A such that (52) is satisfied, for $\sigma^2 = 1$ and various values for (β, γ). As A increases, the exponent $E_c'(A, \beta, \gamma, 1)$ converges to the constant value 0.5. In fig.2, the Volterra coefficients are kept fixed $(\beta, \gamma) = (0.1, 0.5)$, while the noise variance σ^2 and A vary. It is noted that the minimum values of the characteristic A, such that criterion (52) is satisfied, is monotonic with respect to σ^2. Like in fig.1, as A increases, the exponent $E_c'(A, 0.1, 0.5, \sigma^2)$ converges to 0.5 for all values of σ^2. Moreover, in both figs. 1,2, the exponents are noted to behave abnormally for low values of A. Concluding, for the aforementioned information transmission setups, if the transmission rate R satisfies $R < 0.5$, then according to (53) and (54), $E_c'(A, \beta, \gamma, \sigma^2) - R$ is strictly positive and thus $\overline{P}_{e,m}$ approaches zero as the blocklength N increases.

IV. Conclusion

An upper bound on the average maximum likelihood error decoding probability for nonlinear additive gaussian channels

TABLE I

THE MAXIMUM AND MINIMUM VALUES OF THE MARTINGALE DIFFERENCE SEQUENCE $\{Y_i\}_{i=1}^N$ IN COMPARISON WITH THE BOUND OF LEMMA 1, FOR VOLTERRA COEFFICIENTS $(\alpha, \beta, \gamma) = (0, 0.1, 0.5)$ AND VARIOUS VALUES OF A.

A	$\max\{Y_i\}_{i=1}^N$	$\min\{Y_i\}_{i=1}^N$	$4(\mu+1)g(A)^2$
2	38.48	-38.48	211.68
3	397.08	-397.08	2285.28
4	2150.72	-2150.72	12597.1
5	8063	-8063	47628
6	23847.1	-23847.1	141528
7	59785.9	-59785.9	355834
8	132712	-132712	791355
9	268347	-268347	1.60221×10^6
10	504002	-504002	3.01201×10^6

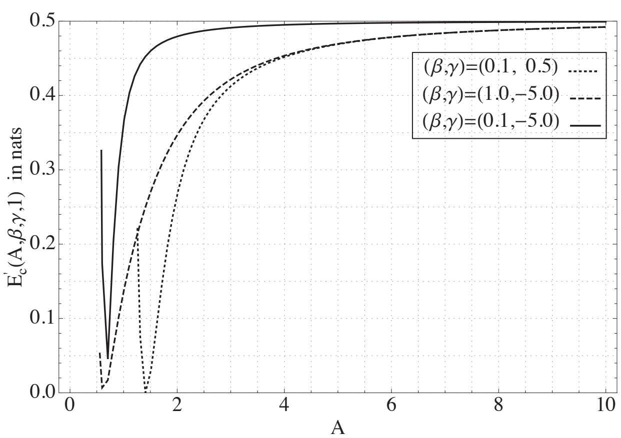

Fig. 1. Random coding exponent $E_c'(A, \beta, \gamma, \sigma^2)$ for noise variance $\sigma^2 = 1$ and various values of A and Volterra coefficients (β, γ).

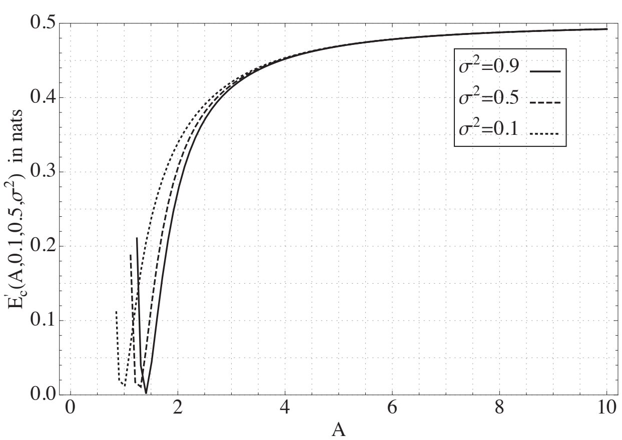

Fig. 2. Random coding exponent $E_c'(A, \beta, \gamma, \sigma^2)$ with Volterra coefficients $(\beta, \gamma) = (0.1, 0.5)$ and various values of A and noise variance σ^2.

is presented. Modeling the channel's nonlinear behavior as a stationary, causal Volterra system, a random coding exponent and thus a lower bound on the channel's capacity are deduced, utilizing concentration inequalities. In special cubic nonlinearities, upper bounds are derived, leading to useful remarks about the channel's behavior. The proposed technique covers linear intersymbol interference channels. Due to the union bound effect, the derived results provide lower bounds to the actual coding exponents. Future work on the random coding exponents of nonlinear channels includes among others, possible ways, in relation with martingale theory, to leverage the union bound bound effect, the establishment of tight lower bounds on the channel capacity, the revocation of the input i.i.d. assumption and the treatment of non-stationary Volterra kernels.

REFERENCES

[1] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Springer, 1999.

[2] S. Verdu and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[3] S. Shamai and S. Raghavan, "On the generalized symmetric cutoff rate for finite-state channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1333 – 1346, 1995.

[4] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*. New York:Academic Press, 1980.

[5] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Krieger Publishing Company, 2006.

[6] S. Boyd, "Volterra series: Engineering fundamentals," Ph.D. dissertation, Berkeley, 1985.

[7] R. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, New York, 1968.

[8] D. Forney Jr, "Capacity of awgn channels," M.I.T. Lecture Notes, 2005.

[9] C. McDiarmid, "Concentration," *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248, 1998.