New Achievable Rates for Nonlinear Volterra Channels via Martingale Inequalities

Kostis Xenoulis   Nicholas Kalouptsidis
Department of Informatics & Telecommunications
National and Kapodistrian University of Athens
Athens 15784, Greece
E-mails: {kxen, kalou}@di.uoa.gr

Igal Sason
Department of Electrical Engineering
Technion-Israel Institute of Technology
Haifa 32000, Israel
E-mail: sason@ee.technion.ac.il

***Abstract*—This paper establishes new achievable rates for non-linear Volterra communication channels using refined versions of the Azuma-Hoeffding inequality. The characteristics of these rates are illuminated in special cases of interest that include time-invariant linear channels with memory, memoryless non-linear channels, and Volterra channel models.**

I. Introduction

Nonlinear effects are typically encountered in wireless communication systems and optical fibers, which degrade the quality of the information transmission. In satellite communication systems, the amplifiers located on board satellites typically operate at or near the saturation region in order to conserve energy. Saturation nonlinearities of amplifiers introduce non-linear distortion in the transmitted signals. Similarly, power amplifiers in mobile terminals are designed to operate in a non-linear region in order to obtain high power efficiency in mobile cellular communications. In optical fibers [1], dispersion noise and nonlinearities need to be overcome with advanced signal processing techniques. Nonlinear communication channels can be represented by Volterra models [2, Chapter 14]. Analytical foundations of the Volterra series operator were presented in [3]. The optical fiber channel has been alternately studied by Poisson type models in [4].

Random coding theorems (see, e.g, [5, Chapter 5.5] constitute critical performance measures of coded information transmission since they describe the exponential behavior of the decoding error probability. The main difficulty in the derivation of calculable closed-form upper bounds on the average decoding error probability when transmission takes place over non-linear channels stems mainly from the difficulty of obtaining the channel output probability density function. Previous work on the analysis of coded communication for Volterra channel models is presented, e.g., in [2, Chapter 14.5] and references therein.

Error probability bounds and achievable rates for Volterra channel models were recently developed in [6] using exponential martingale inequalities. Improved achievable rates for such nonlinear communication channels are derived in this paper, based on some refined versions of the Azuma-Hoeffding inequality from [7], [8] and [9] which provide improvements over the exponential martingale inequalities that were used in [6]. The resulting new expressions of achievable rates are exemplified in this work for linear and non-linear time-invariant channel models.

The paper is structured as follows. Section II provides new bounds on the error probability and achievable rates under ML decoding. These bounds are exemplified in Section III to time-invariant linear channels with memory and memoryless non-linear channels. These comparisons show a significant improvement over the results reported in [6]. Finally, Section IV concludes our discussion.

II. New Bounds on the Error Probability and Achievable Rates under ML Decoding

Under the random coding setup in [5, Section 5.5], consider an ensemble of block codes $\boldsymbol{C}$ where N is the block length, and R is the code rate in nats per channel use. The codewords of a codebook from this ensemble are assumed to be selected independently, and the symbols in each codeword are also assumed to be i.i.d. with an arbitrary probability assignment Q. Let $M = \lceil \exp(NR) \rceil$ be the cardinality of the considered codebook $\mathcal{C}$ from the ensemble $\boldsymbol{C}$. Assume that the relation between the transmitted codeword $\boldsymbol{u} = (u_1, \ldots, u_N) \in \mathcal{C}$ and the discrete-time channel output $\boldsymbol{y} = (y_1, \ldots, y_N)$ is given by the equality

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

where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ is an additive channel noise vector, and D denotes an operator describing the (linear or non-linear) communication channel without the additive noise. The Volterra operator D of order L and memory q is characterized by the equality (see [6, Eq. (4)])

$$[Du]_i = h_0 + \sum_{j=1}^{L} \sum_{i_1=0}^{q} \ldots \sum_{i_j=0}^{q} h_j(i_1, \ldots, i_j) u_{i-i_1} \cdots u_{i-i_j}$$

for $i \in \{1, \ldots, N\}$ and, in the above equation, $u_i \triangleq 0$ for $i \leq 0$. The codeword symbols $\{u_i\}$ are assumed to take values

This research has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)-Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

in a finite subset $\mathcal{R}$. Hence, there exits some $r > 0$ such that

$$\|\boldsymbol{u}\|_\infty \triangleq \max_{1\leq i\leq N} |u_i| < r, \ \forall \boldsymbol{u} \in \boldsymbol{C} \tag{2}$$

and consequently

$$\|D\boldsymbol{u}\|_\infty \leq g_D(\|\boldsymbol{u}\|_\infty) \leq g_D(r) \tag{3}$$

where

$$\begin{aligned} g_D(x) &\triangleq |h_0| + \sum_{j=1}^{L} \|h_j\| x^j, \quad x \geq 0 \\ \|h_j\| &\triangleq \sum_{i_1=0}^{q} \dots \sum_{i_j=0}^{q} |h_j(i_1, \dots, i_j)|. \end{aligned} \tag{4}$$

Following the notation in [6], let $\mathcal{D}_{\rm v}(Q)$ denote the average output variance (without the additive noise)

$$\mathcal{D}_{\rm v}(Q) \triangleq \frac{1}{N} \sum_{j=1}^{N} \mathrm{Var}\left(E[Du]_j\right). \tag{5}$$

Given two randomly selected codewords from the codebook $\boldsymbol{u} = (u_1, \dots, u_N)$, $\tilde{\boldsymbol{u}} = (\tilde{u}_1, \dots, \tilde{u}_N)$, let $\{\mathcal{F}_i\}_{i=1}^N$ be a filtration (i.e., $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_N$) where $\mathcal{F}_i \triangleq \sigma(U_1, \tilde{U}_1, \dots, U_i, \tilde{U}_i)$ is the minimal σ-algebra that is generated by $U_1, \tilde{U}_1, \dots, U_i, \tilde{U}_i$ (for $1 \leq i \leq N$), and $\mathcal{F}_0 \triangleq \{\emptyset, \Omega\}$.

The average decoding error probability $\overline{P}_{\rm e}$ under ML decoding, over the considered ensemble of codes, is upper bounded (for all possible transmitted codewords) by (see [6, Section II])

$$\overline{P}_{\rm e} \leq e^{NR - \left(\frac{\rho}{4\sigma_\nu^2}\right) N \mathcal{D}_{\rm v}(Q)} E\left[e^{\frac{\rho}{8\sigma_\nu^2} \sum_{i=1}^{N} Z_i}\right], \quad 0 \leq \rho \leq 1 \tag{6}$$

where

$$\begin{aligned} Z_i &\triangleq -E\left[\sum_{j=i}^{i+q} w_j^2 \big| \mathcal{F}_i\right] + E\left[\sum_{j=i}^{i+q} w_j^2 \big| \mathcal{F}_{i-1}\right] \\ w_j &\triangleq [Du]_j - [D\tilde{u}]_j. \end{aligned} \tag{7}$$

Remark 1: A typo has been identified in the denominator of the exponent in [6, eq. (10)], which should be $8\sigma_\nu^2$ instead of $4\sigma_\nu^2$.

The martingale difference sequence $\{Z_i\}_{i=0}^N$ with respect to the filtration $\mathcal{F}_i$ satisfies $E[Z_i|\mathcal{F}_{i-1}] = 0, \forall i \in [1, N]$. Let us assume (to be later justified with the calculation of the proper constants) that for a positive d and a non-negative sequence $\{\mu_l\}_{l=2}^m$, where $m \geq 2$ is an arbitrary even number,

$$\begin{aligned} &\max_{u_j, \tilde{u}_j, j \leq i} |Z_i| \leq d, \ \max_{u_j, \tilde{u}_j, j \leq i-1} \left| E\left[Z_i^l | \mathcal{F}_{i-1}\right] \right| \leq \mu_l \\ &\forall\, l \in \{2, \dots, m\}, \quad i \in \{1, \dots, N\} \end{aligned} \tag{8}$$

for an arbitrary $i \in \{1, \dots, N\}$, and set

$$\gamma_l \triangleq \frac{\mu_l}{d^l}, \quad \forall\, l = 2, \dots, m. \tag{9}$$

Parameter d denotes the maximum absolute value of the samples Z_i of the martingale difference sequence with respect to any realization of the filtration $\mathcal{F}_i$, while μ_l denotes the maximum absolute value of the condition mean value of every Z_i^l given $\mathcal{F}_{i-1}$, with respect to any realization of the filtration $\mathcal{F}_{i-1}$.

For the special case where $l = 2$ it is noted that

$$E\left[Z_i^2 \big| \mathcal{F}_{i-1}\right] = \mathrm{Var}(Z_i | \mathcal{F}_{i-1}) \tag{10}$$

since $E[Z_i|\mathcal{F}_{i-1}] = 0$, and for ease of notation we set

$$\mu_2 \triangleq \sigma^2 \text{ with } \gamma_2 = \frac{\sigma^2}{d^2}. \tag{11}$$

Then one can upper bound the second term on the right hand side of (6) in the following two ways:

• First bounding technique ([9, Eq. (44)]): For an even $m \geq 2$

$$\begin{aligned} &E\left[\exp\left(\frac{\rho}{8\sigma_\nu^2} \sum_{i=1}^{N} Z_i\right)\right] \\ &\leq \left[1 + \sum_{l=2}^{m-1} \frac{\gamma_l}{l!} \left(\frac{\rho d}{8\sigma_\nu^2}\right)^l + \frac{\gamma_m \left(\frac{\rho d}{8\sigma_\nu^2}\right)^m \varphi_m\left(\frac{\rho d}{8\sigma_\nu^2}\right)}{m!}\right]^N \end{aligned} \tag{12}$$

where the function φ_m is defined in [9, Eq. (41)] as

$$\varphi_m(y) \triangleq \begin{cases} \frac{m!}{y^m}\left(e^y - \sum_{l=0}^{m-1} \frac{y^l}{l!}\right) & \text{if } y \neq 0 \\ 1 & \text{if } y = 0 \end{cases}. \tag{13}$$

In the special case where $m = 2$, the upper bound in (12) specializes to

$$E\left[\exp\left(\frac{\rho}{8\sigma_\nu^2} \sum_{i=1}^{N} Z_i\right)\right] \leq \left[1 + \gamma_2\left(e^{\frac{\rho d}{8\sigma_\nu^2}} - 1 - \frac{\rho d}{8\sigma_\nu^2}\right)\right]^N. \tag{14}$$

• Second bounding technique (Bennett's inequality) (see [7, Lemma 2.4.1] and [9, Eq. (18)]):

$$E\left[\exp\left(\frac{\rho}{8\sigma_\nu^2} \sum_{i=1}^{N} Z_i\right)\right] \leq \left(\frac{\gamma_2 e^{\frac{\rho d}{8\sigma_\nu^2}} + e^{-\gamma_2\left(\frac{\rho d}{8\sigma_\nu^2}\right)}}{1 + \gamma_2}\right)^N. \tag{15}$$

The parameters d and γ_2 of the martingale $\{Z_i, \mathcal{F}_i\}$ depend on the input probability distribution Q.

Combining the inequalities in (6) and (12) gives the following exponential upper bound on the conditional average error probability under ML decoding:

$$\begin{aligned} &\overline{P}_{\rm e} \leq \exp\Big(-N\big[R_1^{(m)}(\sigma_\nu^2) - R\big]\Big) \\ &R_1^{(m)}(\sigma_\nu^2) \\ &\triangleq \max_{Q} \max_{0 \leq \rho \leq 1} \left[\frac{\rho\, \mathcal{D}_{\rm v}(Q)}{4\sigma_\nu^2} - \ln\left(1 + \sum_{l=2}^{m-1} \frac{\gamma_l}{l!}\left(\frac{\rho d}{8\sigma_\nu^2}\right)^l \right.\right. \\ &\qquad \left.\left. + \frac{\gamma_m}{m!}\left(\frac{\rho d}{8\sigma_\nu^2}\right)^m \varphi_m\left(\frac{\rho d}{8\sigma_\nu^2}\right)\right)\right]. \end{aligned} \tag{16}$$

In the special case where $m = 2$, it gets the form

$$R_1^{(2)}(\sigma_\nu^2) = \max_Q \max_{0\le\rho\le1} \left[\frac{\rho\,\mathcal{D}_{\mathrm{v}}(Q)}{4\sigma_\nu^2} - \ln\left(1 + \gamma_2\left[\exp\left(\frac{\rho d}{8\sigma_\nu^2}\right) - 1 - \frac{\rho d}{8\sigma_\nu^2}\right]\right)\right]. \quad (17)$$

Similarly, the second bounding technique in (15) gives that

$$\overline{P}_{\mathrm{e}} \le \exp\left(-N\left[R_2(\sigma_\nu^2) - R\right]\right)$$

$$R_2(\sigma_\nu^2) \triangleq \max_Q \max_{0\le\rho\le1} \left[\frac{\rho}{4\sigma_\nu^2}\cdot\mathcal{D}_{\mathrm{v}}(Q) - \ln\left(\frac{\gamma_2 \exp\left(\frac{\rho d}{8\sigma_\nu^2}\right) + \exp\left(-\frac{\rho\gamma_2 d}{8\sigma_\nu^2}\right)}{1+\gamma_2}\right)\right]. \quad (18)$$

Remark 2: When there is no information about the conditional variance of the martingale $\{Z_i, \mathcal{F}_i\}_{i=0}^N$ or the calculation is too complex, then by setting $\gamma_2 = 1$ in both (17) and (18), one obtains, respectively, the following achievable rates:

$$R_1'(\sigma_\nu^2) = \max_Q \max_{0\le\rho\le1} \left[\frac{\rho\mathcal{D}_{\mathrm{v}}(Q)}{4\sigma_\nu^2} - \ln\left(e^{\frac{\rho d}{8\sigma_\nu^2}} - \frac{\rho d}{8\sigma_\nu^2}\right)\right] \quad (19)$$

and

$$R_2'(\sigma_\nu^2) = \max_Q \max_{0\le\rho\le1} \left[\frac{\rho\,\mathcal{D}_{\mathrm{v}}(Q)}{4\sigma_\nu^2} - \ln\cosh\left(\frac{\rho d}{8\sigma_\nu^2}\right)\right]. \quad (20)$$

Remark 3: According to [9, Proposition 2], the achievable rate $R_2(\sigma_\nu^2)$ in (18) is larger than $R_1^{(2)}(\sigma_\nu^2)$ in (17).

Remark 4: Following the discussion in [9, Section III.C.1], a closed form expression for the maximal value of ρ in (17) leads to the bound in [9, Corollary 4] (see the proof in [9, Appendix C]).

Remark 5: Both $R_1(\sigma_\nu^2)$ and $R_2(\sigma_\nu^2)$ are continuous functions of Q and ρ. Their maxima are attained since Q and ρ vary over compact sets.

III. Achievable Rates for Random Coding over Some Channel Models

The achievable rates developed above are exemplified in this section for certain special cases of interest.

A. Binary-Input AWGN Channel

For the transmission of information through the channel $y = u + \nu$, $\nu \sim N(0, \sigma_\nu^2)$, the parameters $\mathcal{D}_{\mathrm{v}}(Q)$, d and γ_2 defined in (5), (8), (9) and (11) can be calculated analytically. For the channel input $u \in \{-A, A\}$ where $Q(u = A) = \alpha$, one gets by setting $f(\alpha) \triangleq \alpha(1-\alpha)$

$$Z_i = -\left(u_i - \tilde{u}_i\right)^2 + 8A^2 f(\alpha)$$

and

$$\begin{aligned} d &= 4A^2(1 - 2f(\alpha)), \\ \mathcal{D}_{\mathrm{v}}(Q) &= 4A^2 f(\alpha), \\ \gamma_2 &= \frac{2f(\alpha)}{1-2f(\alpha)}. \end{aligned} \quad (21)$$

Furthermore, letting $\mathrm{SNR} \triangleq A^2/\sigma_\nu^2$, the achievable rates in (17) and (18) are equivalently expressed as follows:

$$R_1^{(2)}(\mathrm{SNR}) = \max_{0\le\alpha\le1} \max_{0\le\rho\le1} \Bigg[\rho f(\alpha)\,\mathrm{SNR} - \ln\Bigg(1 + \frac{2f(\alpha)}{1-2f(\alpha)}\Bigg[e^{\frac{\rho(1-2f(\alpha))\,\mathrm{SNR}}{2}} - 1 - \frac{\rho\,(1-2f(\alpha))\;\mathrm{SNR}}{2}\Bigg]\Bigg)\Bigg] \quad (22)$$

$$R_2(\mathrm{SNR}) = \max_{0\le\alpha\le1} \max_{0\le\rho\le1} \Bigg[\rho\mathrm{SNR} f(\alpha) - \ln\left(2f(\alpha)\,e^{\frac{\rho(1-2f(\alpha))\,\mathrm{SNR}}{2}} + (1-2f(\alpha))e^{-\rho\mathrm{SNR}f(\alpha)}\right)\Bigg]. \quad (23)$$

The simplicity of the specific communication channel allows one to calculate analytically the normalized higher-order conditional moments γ_l in (9). It gives that for every integer $l \ge 2$

$$\gamma_l = 2f(\alpha)\left|\left(\frac{2f(\alpha)}{1-2f(\alpha)}\right)^{l-1} + (-1)^l\right|. \quad (24)$$

Under the above analysis, one can calculate the achievable rate provided by (16) for several values of m and compare it with the one provided by (17) where $m = 2$. More specifically, the achievable rate (16) is expressed in terms of SNR, and it gets the form:

$$R_1^{(m)}(\mathrm{SNR}) = \max_{0\le\alpha\le1} \max_{0\le\rho\le1} \Bigg[\rho\,\mathrm{SNR}\left(1-2f(\alpha)\right) - \ln\Bigg(1 + \sum_{l=2}^{m-1} \frac{\gamma_l}{l!}\left(\frac{\rho\,(1-2f(\alpha))\;\mathrm{SNR}}{2}\right)^l + \frac{\gamma_m}{m!}\left(\frac{\rho\,(1-2f(\alpha))\;\mathrm{SNR}}{2}\right)^m \varphi_m\left(\frac{\rho\,\mathrm{SNR}\left(1-2f(\alpha)\right)}{2}\right)\Bigg)\Bigg]. \quad (25)$$

We next proceed to numerically compare the achievable rates $R_2(\mathrm{SNR})$ in (23), $R_1^{(6)}(\mathrm{SNR})$ in (25) and $R_1^2(\mathrm{SNR})$ in (22), with the achievable rate incorrectly provided in [6, Eq.(84)] and stated here correctly as:

$$R_l(\mathrm{SNR}) = \max_{0\le\alpha\le1} \max_{0\le\rho\le1} 2f(\alpha)\left(\frac{\rho\,\mathrm{SNR}}{2} - \frac{e^{\frac{\rho(1-2f(\alpha))\,\mathrm{SNR}}{2}} - 1 - \frac{\rho(1-2f(\alpha))\,\mathrm{SNR}}{2}}{1-2f(\alpha)}\right). \quad (26)$$

Figure 1 illustrates that all achievable rates provided in this analysis are larger than that provided by (26). Note also that $R_1^{(m)}(\mathrm{SNR})$ converges fast as m grows. Furthermore, the achievable rate $R_1^{(6)}(\mathrm{SNR})$ approximates very well the achievable rate $R_2(\mathrm{SNR})$ provided in (23), and both of them are very close to the capacity of the binary-input additive white Gaussian noise channel for large values of the SNR. The rate loss in the low SNR regime is mainly due to the use of the union bound that led to the establishment of (6). Furthermore, the achievable rate $R_1^{(6)}(\mathrm{SNR})$ almost coincides numerically with $R_2(\mathrm{SNR})$. We proceed now to examine more complex

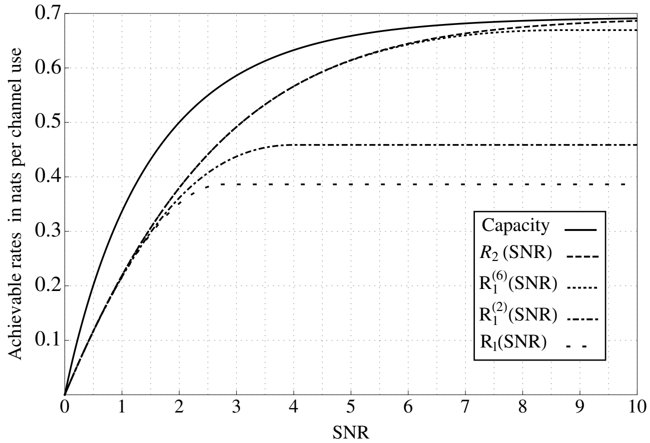

Fig. 1. Comparison among the maximal achievable rate (capacity) and the achievable rates $R_2(\mathrm{SNR})$ (23), $R_1^{(6)}(\mathrm{SNR})$ (25), $R_1^{(2)}(\mathrm{SNR})$ (22), and the (corrected) old bound $R_l(\mathrm{SNR})$ (26) from [6] for the binary-input AWGN channel in terms of $\mathrm{SNR} \triangleq A^2/\sigma_\nu^2$. Rates are expressed in nats per channel use.

cases of Volterra channel models where analytic calculation of the parameters $\mathcal{D}_{\mathrm{v}}(Q)$ and γ_2 is possible.

B. Time-Invariant Linear Channels with Memory

Consider a time-invariant linear channel model with finite memory q

$$[Du]_j = \sum_{k=0}^{q} h_k u_{j-k}. \tag{27}$$

Suppose that the input to the channel is a binary i.i.d. sequence with values from the alphabet $\{-A, A\}$, with $Q(u_i = A) = \alpha$ for all $i \in [0, N]$. Then setting $\boldsymbol{h} \triangleq [h_0\ h_1 \ldots h_q]$

$$\mathcal{D}_{\mathrm{v}}(Q) = 2\|\boldsymbol{h}\|_2^2 \left(E[u^2] - E[u]^2\right) = 4\|\boldsymbol{h}\|_2^2 A^2 f(\alpha) \tag{28}$$

where it is reminded that $f(\alpha) \triangleq \alpha(1-\alpha)$. Furthermore, the martingale-difference sequence $\{Z_i\}_{i=1}^N$ is proved to satisfy

$$\begin{aligned} Z_i = &- \|\boldsymbol{h}\|_2^2((u_i - \tilde{u}_i)^2 - E[(u-\tilde{u})^2]) - \\ &2(u_i - \tilde{u}_i)\sum_{l=1}^{q} \hat{h}_l (u_{i-l} - \tilde{u}_{i-l}), \quad \hat{h}_l \triangleq \sum_{k=0}^{q-l} h_k h_{k+l}. \end{aligned} \tag{29}$$

It then follows that

$$d = 4A^2 \left((1 - 2f(\alpha))\, \|\boldsymbol{h}\|_2^2 + 2\sum_{l=1}^{q} |\hat{h}_l| \right). \tag{30}$$

Moreover, the conditional variance $\mathrm{Var}(Z_i|\mathcal{F}_{i-1})$ satisfies

$$\begin{aligned} \mathrm{Var}(Z_i|\mathcal{F}_{i-1}) = &\|\boldsymbol{h}\|_2^4 \left(E\left[(u-\tilde{u})^4\right] - \left(E\left[(u-\tilde{u})^2\right]\right)^2 \right) + \\ &4E\left[(u-\tilde{u})^2\right] \sum_{k=1}^{q}\sum_{l=1}^{q} \hat{h}_k \hat{h}_l (u_{i-k} - \tilde{u}_{i-k})(u_{i-l} - \tilde{u}_{i-l}) \end{aligned} \tag{31}$$

so that

$$\gamma_2 \le 2f(\alpha) \frac{(1-2f(\alpha))\|\boldsymbol{h}\|_2^4 + 4\sum_{k=1}^{q}\sum_{l=1}^{q} |\hat{h}_k \hat{h}_l|}{\left((1-2f(\alpha))\|\boldsymbol{h}\|_2^2 + 2\sum_{l=1}^{q} |\hat{h}_l|\right)^2}. \tag{32}$$

C. Memoryless Nonlinear Channels

Consider the transmission of information over a memoryless nonlinear channel with the following polynomial characteristic of degree L

$$[Du]_j = \sum_{k=0}^{L} h_k u_j^k. \tag{33}$$

The average output covariance $\mathcal{D}_{\mathrm{v}}(Q)$ is given by

$$\mathcal{D}_{\mathrm{v}}(Q) = \sum_{k=1}^{L}\sum_{l=1}^{L} h_k h_l \left(E\left[u^{k+l}\right] - E\left[u^k\right] E\left[u^l\right] \right). \tag{34}$$

Due to the memoryless nature of the specific system, the martingale-difference sequence $\{Z_i\}_{i=1}^N$ satisfies

$$Z_i = -\left(\sum_{k=1}^{L} h_k (u_i^k - \tilde{u}_i^k) \right)^2 + E\left[\left(\sum_{k=1}^{L} h_k (u^k - \tilde{u}^k) \right)^2 \right] \tag{35}$$

while the parameter μ_2 defined in (8) is given by

$$\begin{aligned} \mu_2 = &E\left[\left(\sum_{k=1}^{L} h_k (u^k - \tilde{u}^k) \right)^4 \right] - \\ &\left(E\left[\left(\sum_{k=1}^{L} h_k (u^k - \tilde{u}^k) \right)^2 \right] \right)^2. \end{aligned} \tag{36}$$

The parameter d in (8) becomes

$$\begin{aligned} d = &\Lambda E\left[\left(\sum_{k=1}^{L} h_k (u^k - \tilde{u}^k) \right)^2 \right] \\ \text{with } \Lambda \triangleq &\max\left\{ 1, \frac{\max_{u_i, \tilde{u}_i} \left(\sum_{k=1}^{L} h_k (u_i^k - \tilde{u}_i^k) \right)^2}{E\left[\left(\sum_{k=1}^{L} h_k (u^k - \tilde{u}^k) \right)^2 \right]} - 1 \right\}. \end{aligned} \tag{37}$$

Hence, from (36), (37) and the definition of γ_2 in (11)

$$\gamma_2 = \frac{1}{\Lambda^2} \frac{E\left[\left(\sum_{k=1}^{L} h_k (u^k - \tilde{u}^k) \right)^4 \right]}{\left(E\left[\left(\sum_{k=1}^{L} h_k (u^k - \tilde{u}^k) \right)^2 \right] \right)^2} - \frac{1}{\Lambda^2}. \tag{38}$$

D. Examples of Third-Order Volterra Channels

Under the same setup of the previous subsection regarding the channel input characteristics, we consider next the transmission of information over the Volterra system D_1 of order $L = 3$ and memory $q = 2$, whose kernels are depicted in Table I. Such system models are used in the base-band

TABLE I
KERNELS OF THE 3RD ORDER VOLTERRA SYSTEM D_1 WITH MEMORY 2

kernel	$h_1(0)$	$h_1(1)$	$h_1(2)$	$h_2(0,0)$	$h_2(1,1)$	$h_2(0,1)$
value	1.0	0.5	-0.8	1.0	-0.3	0.6

kernel	$h_3(0,0,0)$	$h_3(1,1,1)$	$h_3(0,0,1)$	$h_3(0,1,1)$
value	1.0	-0.5	1.2	0.8

kernel	$h_3(0,1,2)$
value	0.6

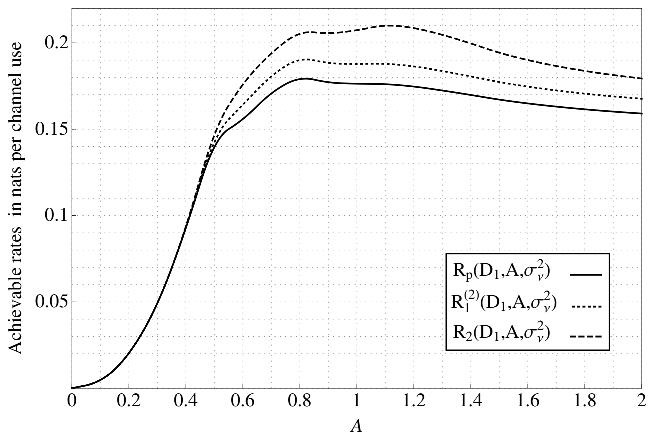

Fig. 2. Comparison of the achievable rates $R_1^{(2)}(D_1,A,\sigma_\nu^2)$ (17) and $R_2(D_1,A,\sigma_\nu^2)$ (18) with the old bound $R_p(D_1,A,\sigma_\nu^2)$ of [6, Fig.2] for the nonlinear channel with kernels depicted in Table I and noise variance $\sigma_\nu^2 = 1$. Rates are expressed in nats per channel use.

representation of nonlinear narrow-band communication channels. Due to complexity of the channel model, the calculation of the achievable rates provided by (17), (18) requires the numerical calculation of the parameters d and σ^2 and thus of γ_2 for the martingale $\{Z_i, \mathcal{F}_i\}_{i=0}^N$. In order to achieve this goal, we have to calculate $|Z_i - Z_{i-1}|$ and $\mathrm{Var}(Z_i|\mathcal{F}_{i-1})$ for all possible combinations of the input samples which contribute to the aforementioned expressions. Thus, the analytic calculation of d and γ_l increases as the system's memory q increases. Numerical results are provided in Figure 2 for the case where $\sigma_\nu^2 = 1$. The new achievable rates $R_1^{(2)}(D_1,A,\sigma_\nu^2)$ and $R_2(D_1,A,\sigma_\nu^2)$ of (17) and (18), which depend on the channel input parameter A, are compared to the achievable rate provided in [6, Fig.2] and shown to be larger than the latter.

E. General Volterra Channel Models

From the previous analysis it follows that in the case of more complex Volterra systems, it is preferable to select $\gamma_2 = 1$. Then the achievable rate $R_2'(\sigma_\nu^2)$ (20) admits a closed-form solution according to the analysis. The upper bound (18) on the average error decoding probability $\overline{P}_{\mathrm{e}}$ is expressed for $0 \le \rho \le 1$ as

$$\overline{P}_{\mathrm{e}} \le \exp\left(-N\left[\frac{\rho\mathcal{D}_{\mathrm{v}}(Q)}{4\sigma_\nu^2} - \ln\cosh\left(\frac{\rho d}{8\sigma_\nu^2}\right) - R\right]\right). \quad (39)$$

Calculation shows that

$$\rho_{\mathrm{opt}} = \begin{cases} \frac{8\sigma_\nu^2}{d}\tanh^{-1}\left(\frac{2\mathcal{D}_{\mathrm{v}}(Q)}{d}\right) & ,\mathcal{D}_{\mathrm{v}}(Q) \le \frac{d}{2}\tanh\left(\frac{d}{8\sigma_\nu^2}\right) \\ 1 & ,\mathcal{D}_{\mathrm{v}}(Q) > \frac{d}{2}\tanh\left(\frac{d}{8\sigma_\nu^2}\right) \end{cases}. \quad (40)$$

Concluding the analysis, the upper bound in (39) is expressed equivalently as

$$\overline{P}_{\mathrm{e}} \le \exp\left(-N\left[E_c(Q,D,\sigma_\nu^2) - R\right]\right) \quad (41)$$

where

$$E_c(Q,D,\sigma_\nu^2) = \begin{cases} \frac{2\mathcal{D}_{\mathrm{v}}(Q)}{d}\tanh^{-1}\left(\frac{2\mathcal{D}_{\mathrm{v}}(Q)}{d}\right) \\ \quad +\frac{1}{2}\ln\left[1-\left(\frac{2\mathcal{D}_{\mathrm{v}}(Q)}{d}\right)^2\right] & ,\ \mathcal{D}_{\mathrm{v}}(Q) \le \frac{d}{2}\tanh\left(\frac{d}{8\sigma_\nu^2}\right) \\ \frac{\mathcal{D}_{\mathrm{v}}(Q)}{4\sigma_\nu^2} - \ln\cosh\left(\frac{d}{8\sigma_\nu^2}\right) & ,\ \text{otherwise} \end{cases} \quad (42)$$

and $E_c(Q,D,\sigma_\nu^2) \ge 0$ in both cases.

IV. SUMMARY

New achievable rates for random coding are derived in this work for non-linear communication channels. Exponential martingale inequalities are properly utilized for this purpose. The achievable rates obtained in this work improve previous results obtained for the same channel models in [6]. This improvement is facilitated by the use of some refined versions of the Azuma-Hoeffding inequality that were introduced in [8], [10], and more recently in [9]. The analysis presented here can easily be extended to the case where the noise is a complex Gaussian or circularly complex Gaussian, and the parameters of the Volterra system are complex valued as it is in the Volterra representation of bandpass nonlinear channels [2, Section 14.3].

REFERENCES

[1] A. C. Singer, N. R. Shanbhag, and Hyeon–Min Bae, "Electronic dispersion compensation," *IEEE Signal Processing Magazime*, vol. 25, no. 6, pp. 110–130, 2008.
[2] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*, Kluwer Academic/ Plenum Publishers, 1999.
[3] S. Boyd, L. O. Chua and C. A. Desoer, "Analytical foundations of Volterra series," *IMA Journal of Mathematical Control & Information*, vol. 1, pp. 243–282, 1984.
[4] M. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. on Information Theory*, vol. 26, no. 6, pp. 710–715, 1980.
[5] R. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, New York, 1968.
[6] K. Xenoulis and N. Kalouptsidis, "Achievable rates for nonlinear Volterra channels," *IEEE Trans. on Information Theory*, vol. 57, no. 3, pp. 1237–1248, March 2011.
[7] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, second edition, 1997.
[8] C. McDiarmid, "On the method of bounded differences," *Surveys in Combinatorics*, vol. 141, pp. 148–188, Cambridge University Press, Cambridge, 1989.
[9] I. Sason, "On refined versions of the Azuma-Hoeffding inequality with applications in information theory," last updated on February 2012. [Online]. Available at http://arxiv.org/abs/1111.1977.
[10] C. McDiarmid, "Concentration," *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248, Springer, 1998.