Achievable Rates for Nonlinear Volterra Channels

Kostis Xenoulis, *Student Member, IEEE*, and Nicholas Kalouptsidis, *Senior Member, IEEE*

***Abstract*—Random coding theorems and achievable rates for nonlinear additive noise channels are presented. Modeling the channel's nonlinear behavior as a causal, stationary Volterra system, upper bounds on the average error probability are obtained for maximum likelihood and weakly typical set decoding. The proposed bounds are deduced by treating correct decoding regions as subspaces of high concentration measure and deploying exponential martingale inequalities. Due to the union bound effect and the i.i.d. assumption imposed on the codewords components, the deduced exponents constitute only lower bounds on the true random coding exponents of nonlinear channels. Cubic and fourth-order nonlinearities are used as examples to illustrate the relation of the random coding exponents and achievable rates with respect to the channel's parameters.**

***Index Terms*—Achievable rates, cut-off rate, martingale inequalities, maximum likelihood/weakly typical set decoding, nonlinear channels, random coding exponents, Volterra systems.**

I. Introduction

VARIOUS communication channels, including wireless and optical fibers, exhibit nonlinear behavior that degrades the quality of information transmission. In satellite communication systems [1, Chap. 14], the amplifiers located on board satellites usually operate at or near the saturation region in order to conserve energy. Saturation nonlinearities of amplifiers introduce nonlinear distortions in the transmitted signals. In a similar way, power amplifiers in hand held mobile terminals are forced to operate in a nonlinear region to secure high power efficiency in mobile cellular communications [2, Chap. 12]. In optical fibers [3], dispersion noise and nonlinearities pose substantial channel impairments that need to be overcome with advanced signal processing techniques. Unlike short-reach multimode fibers where the transmitted signal will excite a number of propagating modes, longer reach applications cannot be modelled as linear due to the effects of chromatic dispersion (light propagates with a wavelength dependent velocity in fiber) and polarization-made dispersion arising from manufacturing defects, vibration or mechanical stresses in the fiber. The effect of direct (square law) detection on such optically linear impairments is to make them nonlinear in the electrical domain. Likewise, the single-mode fiber that permits only a single optical mode to propagate, is best described as a nonlinear channel with finite memory μ

$$y_i = f(u_i, u_{i-1}, \ldots, u_{i-\mu}) + \nu_i \tag{1}$$

Manuscript received June 26, 2009; revised April 01, 2010; accepted July 15, 2010. Date of current version February 18, 2011.

The authors are with the Department of Informatics and Telecommunications, University of Athens, Greece (e-mail: kxen@di.uoa.gr; kalou@di.uoa.gr).

Communicated by A. Lapidoth, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2104571

where ν_i comprises additive noise, u_i is the transmitted signal and y_i is the received sample. The function f could be a member of a dense subset of the set of continuous real valued functions on a compact subset of $\mathcal{R}^{\mu}$, such as the set of polynomial functions, neural networks and radial basis functions [4]. It is known [4], [5] that models of the form (1) are fairly general as they can approximate every time invariant causal stable system with fading memory. If f is a polynomial function, (1) becomes a Volterra model [6]. Volterra models have been widely used to model nonlinear channels [1], since they offer a reasonable balance between faithful representation of the physical system and analytic tractability.

Random coding theorems constitute critical performance measures of coded information transmission since they describe the exponential behavior of error probability. 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 formula for channel capacity given by [7] cannot easily be 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 for ensembles of codes. Previous work on the calculation of maximum likelihood (ML) error performance of nonlinear gaussian noise channels under Volterra series representations and for specific coded modulation schemes is reported among others in [1, Chap. 14], [8], [9], and [10].

Computationally feasible random coding results for nonlinear channels are not readily linked to the coding theorem for finite state channels [11, Sec. 5.9], since the channel state must be known at the receiver. Furthermore, the approach in [12], where the Perron–Frobenius theorem is utilized, provides closed form expressions only for the generalized symmetric cut-off rate. In order to overcome these restrictions, the present work approaches the specific coding theorem problem in a different and more general manner (nonsymmetric inputs), by utilizing the theory of martingale processes [13]. Erroneous decoding regions are interpreted as subspaces of low concentration measure and as such are bounded by exponential martingale inequalities. The latter provide tight concentration inequalities for multivariate functions with memory. Moreover, the proposed approach allows the treatment of nonlinear channels when the corresponding transition probability law is unknown or when suboptimum decoding algorithms are performed. According to [14], [15], and [16], there exist cases where ML decoding is performed with respect to the wrong channel law, due to imprecise channel measurement, or the respective decoding implementation is too complex. To handle such cases, the current work employs a weakly typical set decoding procedure with respect to an arbitrary input-output dependence measure, different from the information density $\log p(\boldsymbol{u},\boldsymbol{y})/p(\boldsymbol{u})p(\boldsymbol{y})$ used in strong typical set decoding [17].

Under the random coding setup presented in [11, Sec. 5.5], we consider the transmission of an arbitrary set of messages $\mathcal{M}$ with cardinality $|\mathcal{M}|$ over a nonlinear additive noise channel (1), where f is polynomial. Specifically, let for each message $m \in \mathcal{M}$, $1 \leq m \leq |\mathcal{M}|$, where $|\mathcal{M}| = \lceil \exp(NR) \rceil$, an N-length codeword $\boldsymbol{u}$ be selected randomly from the ensemble of (N,R) block codes $\boldsymbol{C}$ and transmitted over the nonlinear channel

$$\boldsymbol{y} = D\boldsymbol{u} + \boldsymbol{\nu}. \tag{2}$$

The channel output vector is $\boldsymbol{y} = (y_1,\ldots,y_N)$ whereas $\boldsymbol{u} = (u_1,\ldots,u_N)$ and $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_N)$ are the channel input and noise vectors, respectively, and $[Du]_i = f(u_i,\ldots,u_{i-\mu})$, $1 \leq i \leq N$. All codewords are chosen independently with the same probability Q from the ensemble $\boldsymbol{C}$ while the components of all codewords are independent and identically distributed (i.i.d.). Codeword components u_i take values in a finite subset of $\mathcal{R}$. Therefore, there is $r > 0$ such that

$$\|\boldsymbol{u}\|_\infty = \max_{1\leq i\leq N} |u_i| < r,\ \forall \boldsymbol{u} \in \boldsymbol{C}. \tag{3}$$

Since the static function f in (1) is continuous there exists $M > 0$ such that

$$\forall \boldsymbol{u} : \|\boldsymbol{u}\|_\infty < r \Rightarrow \|D\boldsymbol{u}\|_\infty < M.$$

In particular, if f is polynomial, then

$$[Du]_i = h_0 + \sum_{j=1}^{d}\sum_{i_1=0}^{\mu}\cdots\sum_{i_j=0}^{\mu} h_j(i_1,\ldots,i_j)u_{i-i_1}\cdots u_{i-i_j} \tag{4}$$

and

$$\|D\boldsymbol{u}\|_\infty \leq g_D(\|\boldsymbol{u}\|_\infty) \leq g_D(r) \tag{5}$$

where

$$g_D(x) = |h_0| + \sum_{j=1}^{d} \|h_j\| x^j, \quad x \geq 0 \quad \text{with}$$
$$\|h_j\| = \sum_{i_1=0}^{\mu}\cdots\sum_{i_j=0}^{\mu} |h_j(i_1,\ldots,i_j)|. \tag{6}$$

In the sequel, we assume input causality i.e., $u_i = 0$ for all $i \leq 0$. Section II treats additive gaussian noise nonlinear channels and provides exponential upper bounds on the average ML error decoding probability. Section III introduces weakly typical set decoding for nonlinear channels with unknown transition probability law. The selected typical dependence measure leads to random coding exponents that depend on the channel input-output correlation [18], in contrary to the classical memoryless approach, where the random exponents are related to the channel's mutual information [11, eq. (4.2.2)]. Cubic and fourth order nonlinear channels are treated in Section IV, where the effects of the code and channel characteristics on the behavior of the random coding exponents are demonstrated. Section V indicates directions for further research.

II. Nonlinear Gaussian Noise Channels

A. Pairwise Error Probability Analysis

Consider a specific (N,R) block code $\mathcal{C}$ and suppose that ML decoding is performed at the output of channel (2). Suppose also that the noise vector $\boldsymbol{\nu}$ is i.i.d gaussian with zero mean and variance σ_ν^2. An ML error occurs if, given the transmitted message m and the received vector $\boldsymbol{y}$, another message $m' \neq m$ exists such that

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

or equivalently

$$\|\boldsymbol{y} - D\boldsymbol{u}_{m'}\|_2^2 \leq \|\boldsymbol{y} - D\boldsymbol{u}_m\|_2^2. \tag{7}$$

Since m is the transmitted message, $\boldsymbol{y} = D\boldsymbol{u}_m + \boldsymbol{\nu}$, so that the error criterion (7) is translated into

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

If $A_{m,m'}$ denotes the event (8), then the ML error decoding probability for the specific code $\mathcal{C}$, given that the message m is transmitted, satisfies

$$P_e(m|\mathcal{C}) = \Pr\left(\bigcup_{m'\neq m} A_{m,m'}\right)$$

and is upper bounded as in Gallager [11, eq. (5.6.2)]

$$P_e(m|\mathcal{C}) \leq \left(\sum_{m'\neq m} \Pr\left(A_{m,m'}\right)\right)^{\rho} \quad \text{for } 0 \leq \rho \leq 1. \tag{9}$$

We now recall the variant of Hölder's inequality

$$\left(\sum_i a_i\right)^{\rho} \leq \sum_i a_i^{\rho} \text{ for } 0 \leq \rho \leq 1 \text{ and } a_i \geq 0$$

and the bound presented in [2, pp. 139–140]

$$\Pr(A_{m,m'}) \leq \exp\left(-\frac{\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2}{4\sigma_\nu^2}\right).$$

Using these bounds, (9) becomes

$$P_e(m|\mathcal{C}) \leq \sum_{m'\neq m} \exp\left(-\rho\frac{\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2}{4\sigma_\nu^2}\right). \tag{10}$$

The above upper bound applies to a specific block code $\mathcal{C}$. Following the random coding setup of Gallager [11, Chap. 5], we average both terms of (10) over the ensemble $\boldsymbol{C}$ of (N,R) block codes, described in Section I. The average ML error decoding probability $\bar{P}_{e,m}$, given the transmitted message m, satisfies

$$\bar{P}_{e,m} \le E\left[\sum_{m' \neq m} \exp\left(-\rho \frac{\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2}{4\sigma_\nu^2}\right)\right]. \quad (11)$$

Since the codewords are generated independently, (11) is equivalently expressed as

$$\bar{P}_{e,m} \le \sum_{\boldsymbol{u}_m} Q(\boldsymbol{u}_m) \sum_{m' \neq m} \sum_{\boldsymbol{u}_{m'}} Q(\boldsymbol{u}_{m'}) \cdot \exp\left(-\rho \frac{\|D\boldsymbol{u}_m - D\boldsymbol{u}_{m'}\|_2^2}{4\sigma_\nu^2}\right). \quad (12)$$

In a similar manner to [11, eq. (5.6.9)], $\boldsymbol{u}_m$, $\boldsymbol{u}_{m'}$ are dummy variables of summation in (12), so that subscripts m, m' can be dropped and the bound is independent of m, m'. Since there are $|\mathcal{M}| - 1$ choices of $m' \neq m$, the upper bound in (10) satisfies through (11) and (12)

$$\bar{P}_{e,m} \le (|\mathcal{M}| - 1) E\left[\exp\left(-\rho \frac{\|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2}{4\sigma_\nu^2}\right)\right]. \quad (13)$$

Suppose that u_j, u'_j are the jth components of the corresponding random vectors $\boldsymbol{u}$, $\boldsymbol{u}'$. Let

$$\begin{aligned} &\emptyset = F_0 \subset F_1 \subset \cdots F_N \quad \text{where} \\ &F_i = \{u_1, \ldots, u_i, u'_1, \ldots, u'_i\} \end{aligned} \quad (14)$$

and

$$\begin{aligned} \Delta_i &= X_i - X_{i-1}, \quad 1 \le i \le N \\ X_i &= E\left[\|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2 | F_i\right] \\ X_0 &= E\left[\|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2\right] \\ X_N &= \|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2. \end{aligned} \quad (15)$$

We refer to the sequence $\{\Delta_i\}_{i=1}^N$ as the martingale difference sequence [19] of the random variable X_N with respect to the joint filter $\{F_i\}_{i=0}^N$ in (14). The mean values appearing in (15) are with respect to all codewords the random variables $\boldsymbol{u}$, $\boldsymbol{u}'$ can be assigned to. Under the previous setup, we note that

$$\begin{aligned} \sum_{i=1}^N \Delta_i &= X_N - X_0 \\ &= \|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2 - E\left[\|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2\right] \end{aligned} \quad (16)$$

and thus, (13) is equivalently expressed as

$$\bar{P}_{e,m} \le (|\mathcal{M}| - 1) \exp\left(-\frac{\rho}{4\sigma_\nu^2} E\left[\|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2\right]\right) \cdot E\left[\exp\left(-\frac{\rho}{4\sigma_\nu^2} \sum_{i=1}^N \Delta_i\right)\right]. \quad (17)$$

Due to the random coding setup and the independence of the ensemble's codewords, it holds

$$E\left[\|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2\right] = 2\left(\sum_{j=1}^N E\left[([Du]_j)^2\right] - E\left[[Du]_j\right]^2\right). \quad (18)$$

Moreover, under the assumption that the components u_j, $1 \le j \le N$ of all codewords are i.i.d., it holds for $\mu + 1 \le j \le N$

$$E\left[([Du]_j)^2\right] - E\left[[Du]_j\right]^2 = \mathcal{D}_v(Q). \quad (19)$$

Furthermore, due to the causality assumption stated after (6), the variances $E\left[([Du]_j)^2\right] - E\left[[Du]_j\right]^2$, $1 \le j \le \mu$ are different from $\mathcal{D}_v(Q)$. For large block lengths N, the previous μ terms are negligible, and thus, from (18) and (19)

$$\begin{aligned} &E\left[\|D\boldsymbol{u} - D\boldsymbol{u}'\|_2^2\right] \\ &= 2\left(\sum_{j=1}^N E\left[([Du]_j)^2\right] - E\left[[Du]_j\right]^2\right) \\ &= 2N\mathcal{D}_v(Q). \end{aligned} \quad (20)$$

Finally, combining (17) and (20), we obtain

$$\bar{P}_{e,m} \le \exp\left(NR - \frac{\rho}{2\sigma_\nu^2} N\mathcal{D}_v(Q)\right) \cdot E\left[\exp\left(-\frac{\rho}{4\sigma_\nu^2} \sum_{i=1}^N \Delta_i\right)\right]. \quad (21)$$

B. Random Coding Theorem

The development of exponential upper bounds for the mean value on the right-hand side of (21) requires bounds on the conditional deviations $dev^+(\Delta_i)$ and conditional variances $var(\Delta_i | F_{i-1})$, where according to [19, pp. 24]

$$\begin{aligned} dev^+(\Delta_i) &= \max_{u_i, u'_i} \Delta_i \\ var(\Delta_i | F_{i-1}) &= E\left[(X_i - X_{i-1})^2 | F_{i-1}\right]. \end{aligned} \quad (22)$$

Appropriate bounds are derived in the lemma that follows.

Lemma 1: Under the assumptions that the components of all codewords $\boldsymbol{u}_m$, $1 \le m \le \mathcal{M}$ are mutually independent and r is chosen as in (3), the martingale differences Δ_i (15) satisfy

$$\begin{aligned} dev^+(-\Delta_i) &\le 4(\mu+1) g_D(r)^2 \\ var(-\Delta_i | F_{i-1}) &\le 16(\mu+1)^2 g_D(r)^4. \end{aligned} \quad (23)$$

Proof: Let

$$d_j = [Du]_j - [Du']_j, \quad 1 \le j \le N. \quad (24)$$

Then, X_i in (15) equals

$$\begin{aligned} X_i &= E\left[\sum_{j=1}^{N} d_j^2 \middle| F_i\right] \\ &= E\left[\sum_{j=1}^{i-1} d_j^2 \middle| F_i\right] + E\left[\sum_{j=i}^{i+\mu} d_j^2 \middle| F_i\right] \\ &\quad + E\left[\sum_{j=i+\mu+1}^{N} d_j^2 \middle| F_i\right]. \end{aligned} \tag{25}$$

Under the lemma assumptions, all samples d_j, for $j \geq i + 1 + \mu$ and $j \leq i-1$ are independent from the samples u_i, u_i'. Consequently

$$X_i = E\left[\sum_{j=1}^{i-1} d_j^2 \middle| F_{i-1}\right] + E\left[\sum_{j=i}^{i+\mu} d_j^2 \middle| F_i\right] + E\left[\sum_{j=i+\mu+1}^{N} d_j^2 \middle| F_{i-1}\right] \tag{26}$$

and

$$X_{i-1} = E\left[\sum_{j=1}^{i-1} d_j^2 \middle| F_{i-1}\right] + E\left[\sum_{j=i}^{i+\mu} d_j^2 \middle| F_{i-1}\right] + E\left[\sum_{j=i+\mu+1}^{N} d_j^2 \middle| F_{i-1}\right]. \tag{27}$$

Therefore, due to (26), (27), and (15)

$$\Delta_i = E\left[\sum_{j=i}^{i+\mu} d_j^2 \middle| F_i\right] - E\left[\sum_{j=i}^{i+\mu} d_j^2 \middle| F_{i-1}\right]. \tag{28}$$

Using (5), we have

$$|[Du]_j| \leq \|D\boldsymbol{u}\|_\infty \leq g_D(r) \tag{29}$$

and thus

$$\begin{aligned} d_j^2 &= ([Du]_j - [Du']_j)^2 \leq \|D\boldsymbol{u} - D\boldsymbol{u}'\|_\infty^2 \\ &\leq (\|D\boldsymbol{u}\|_\infty + \|D\boldsymbol{u}'\|_\infty)^2 = 4g_D(r)^2. \end{aligned} \tag{30}$$

The second summand on the right-hand side of (28) is nonpositive. Therefore, from (28)–(30), it holds

$$-\Delta_i \leq |\Delta_i| < 4(\mu+1)g_D(r)^2. \tag{31}$$

The proof is established if we insert (31) in (22). ∎

The bounds provided by Lemma 1 lead to random coding upper bounds on the average ML error decoding probability. Tighter bounds can be obtained analytically for Volterra systems D of short memory, as shown in Section IV.

Theorem 1: Consider the transmission of an arbitrary set of messages $\mathcal{M}$ over a nonlinear Volterra additive gaussian noise channel (2) with memory μ. The components of the noise vector are i.i.d. random variables with 0 mean value and variance σ_ν^2. For each message m, $1 \leq m \leq |\mathcal{M}|$, an N-length codeword $\boldsymbol{u}_m$ is selected from the ensemble $\boldsymbol{C}$ of (N,R) block codes with probability Q, independently from all other codewords and is transmitted over the channel. If ML decoding is performed at the receiver, each codeword is chosen with i.i.d components and the assumptions of Lemma 1 are valid, then for $\mathcal{D}_v(Q)$ given by (19) the average error decoding probability $\bar{P}_e$ is upper bounded as

$$\bar{P}_e \leq e^{-N\left(E_c(Q,D,\sigma_\nu^2)-R\right)} \tag{32}$$

where $E_c(Q,D,\sigma_\nu^2)$ is given by (33), at the bottom of the page, and

$$\kappa = 4(\mu+1)g_D(r)^2. \tag{34}$$

Proof: From the arguments of [19, Lemma 3.16] and surrounding it and [19, Lemma 2.8], if $\{Z_i\}_{i=1}^N$ is a martingale difference sequence with maximum conditional deviation and maximum sum of conditional variances satisfying, respectively

$$\begin{aligned} \max_i dev^+(Z_i) &\leq b \\ \max_{F_N} \sum_{i=1}^{N} var(Z_i|F_{i-1}) &\leq V \end{aligned} \tag{35}$$

then for every $h > 0$, it holds

$$E\left[\exp\left(h\sum_{i=1}^{N} Z_i\right)\right] \leq \exp\left(\frac{\exp(hb)-1-hb}{b^2}V\right). \tag{36}$$

The basic idea in establishing (36) goes as follows: we express the left-hand side of (36) as

$$E\left[\exp\left(h\sum_{i=1}^{N-1} Z_i\right) E_{u_N,u_N'}\left[\exp\left(hZ_N\right)|F_{N-1}\right]\right] \tag{37}$$

and we bound the inner mean value of (37) as in the proof of Theorem 3.15 in [19, Lemma 2.8]

$$\begin{aligned} &E_{u_N,u_N'}\left[\exp\left(hZ_N\right)|F_{N-1}\right] \\ &\quad \leq \exp\left(\frac{\exp\left(hb\right)-1-hb}{b^2} var(Z_N|F_{N-1})\right). \end{aligned} \tag{38}$$

We apply the above result setting $Z_i = -\Delta_i$. Due to Lemma 1, we have $V = N\kappa^2$ and $b = \kappa$. Consequently, the mean value on the right-hand side of (21) is upper bounded due to (36) as

$$\begin{aligned} &E\left[\exp\left(-\frac{\rho}{4\sigma_\nu^2}\sum_{i=1}^{N}\Delta_i\right)\right] \\ &\quad \leq \exp\left(\frac{\exp\left(\frac{\rho}{4\sigma_\nu^2}\kappa\right)-1-\frac{\rho}{4\sigma_\nu^2}\kappa}{\kappa^2}N\kappa^2\right). \end{aligned} \tag{39}$$

$$E_c(Q,D,\sigma_\nu^2) = \begin{cases} \frac{1}{2\sigma_\nu^2}\mathcal{D}_v(Q) - \left(\exp\left(\frac{\kappa}{4\sigma_\nu^2}\right) - 1 - \frac{\kappa}{4\sigma_\nu^2}\right), & \mathcal{D}_v(Q) > \frac{\kappa}{2}\left(\exp\left(\frac{\kappa}{4\sigma_\nu^2}\right)-1\right) \\ \frac{1}{\kappa}\left(-2\mathcal{D}_v(Q) + (\kappa + 2\mathcal{D}_v(Q))\ln\left(1 + \frac{2\mathcal{D}_v(Q)}{\kappa}\right)\right), & \text{otherwise} \end{cases} \tag{33}$$

Combining (21) and (39), we have

$$\bar{P}_{e,m} \leq \exp\left(-N\left[\frac{\rho\mathcal{D}_v(Q)}{2\sigma_\nu^2} - \left(\exp\left(\frac{\rho\kappa}{4\sigma_\nu^2}\right) - 1 - \frac{\rho\kappa}{4\sigma_\nu^2}\right) - R\right]\right). \quad (40)$$

For fixed Q and $0 \leq \rho \leq 1$, let

$$\xi(\rho) = \frac{\rho}{2\sigma_\nu^2}\mathcal{D}_v(Q) - \left(\exp\left(\frac{\rho}{4\sigma_\nu^2}\kappa\right) - 1 - \frac{\rho}{4\sigma_\nu^2}\kappa\right).$$

The first and second derivatives of $\xi(\rho)$ are, respectively

$$\xi'(\rho) = \frac{\mathcal{D}_v(Q)}{2\sigma_\nu^2} - \kappa\frac{\exp\left(\frac{\rho\kappa}{4\sigma_\nu^2}\right) - 1}{4\sigma_\nu^2}$$

$$\xi''(\rho) = -\kappa^2\frac{\exp\left(\frac{\rho\kappa}{4\sigma_\nu^2}\right)}{(4\sigma_\nu^2)^2}.$$

Since $\xi''(\rho) < 0$, $\xi(\rho)$ is strictly concave, and thus, has a unique maximum, at

$$\rho_{opt} = \frac{4\sigma_\nu^2\ln\left(1 + \frac{2\mathcal{D}_v(Q)}{\kappa}\right)}{\kappa}.$$

If $\rho_{opt} > 1$, then $\xi(\rho)$ is increasing in $\rho \in [0,1]$ and its maximum is achieved at $\rho = 1$ in (40). If $\rho_{opt} \leq 1$, then $\xi(\rho_{opt})$ defines the exponent $E_c(Q, D, \sigma_\nu^2)$ in (40). Finally, for both cases of ρ_{opt}, $E_c(Q, D, \sigma_\nu^2)$ is non-negative, since $\xi(0) = 0$ and $\xi(\rho)$ is increasing for $\rho \leq \rho_{opt}$. ■

Remark: The output variance $\mathcal{D}_v(Q)$ appearing (33) depends on the system parameters and the input statistics. The corresponding expression is complicated in the general case. Consider, for instance, a polynomial Volterra system (4) of order $d = 2$ where the input samples u_i take values from the set $\{-A, A\}$ with probability $Q(u_i = A) = \pi$. Then

$$\begin{aligned}\mathcal{D}_v(Q) = &\gamma_2\sum_{i=0}^{\mu}h_1(i)^2 + 2\gamma_3\sum_{i=0}^{\mu}h_1(i)h_2(i,i)\\ &+ 4\gamma_2\gamma_1\sum_{i=0}^{\mu}\sum_{j=0}^{\mu}h_1(i)h_2(i,j)\\ &+ \gamma_4\sum_{i=0}^{\mu}h_2(i,i)^2\\ &+ 2\gamma_2^2\sum_{i=0}^{\mu}\sum_{j=0}^{\mu}h_2(i,j)^2\\ &+ 4\gamma_1\gamma_3\sum_{i=0}^{\mu}\sum_{j=0}^{\mu}h_2(i,j)h_2(i,i)\\ &+ 4\gamma_1^2\gamma_2\sum_{i=0}^{\mu}\sum_{j=0}^{\mu}\sum_{k=0}^{\mu}h_2(i,j)h_2(i,k)\end{aligned}$$

where

$$\begin{aligned}\gamma_1 &= A(2\pi - 1)\\ \gamma_2 &= 4A^2\pi(1-\pi)\\ \gamma_3 &= 8A^3\pi(1-\pi)(1-2\pi)\\ \gamma_4 &= A^4\left(5A^2 + A(2\pi - 1)\right)\\ &\quad - 2A^4\left(6 - 32\pi + 80\pi^2 - 96\pi^3 + 48\pi^4\right).\end{aligned}$$

Corollary 1: All rates below $\max_Q E_c(Q, D, \sigma_\nu^2)$ (33) are achievable for transmission of information over a nonlinear additive gaussian noise channel under ML decoding.

Discussion: In order to illustrate the meaning of Theorem 1, we consider the discrete memoryless channel $y = u + \nu$ with discrete input and continuous output and compare the respective achievable rate of Theorem 1 with the corresponding channel capacity $\max_Q I(U;Y)$ [11, eq. (4.2.3)]. For this channel and for low values of signal to noise ratio, numerical simulations demonstrate a very good match. This is discussed in Section IV. A more refined analysis of the random coding problem for nonlinear channels under the martingale setup and its corresponding exponential inequalities is expected to yield tighter achievable rates. This is supported by the recent work [20] where sub-martingales and optional stopping theorems are utilized properly in the reproof of Burnashev's reliability function [21].

III. Weakly Typical Set Decoding

In this section, decoding rules for nonlinear channels are interpreted as concentration measures and martingale theory is utilized. The analysis can be applied to cases where the channel's transition probability law is generally unknown or a suboptimum decoding algorithm is adopted. Weakly typical set decoding is first considered under the assumption that a suitable bounded function, concentrated around its mean, exists. Then the nonlinear model (2) is undertaken using a correlation measure of the channel output, where the decoding function is not necessarily bounded.

A. Bounded Concentrated Functions and Error Probability Analysis

Consider an arbitrary bounded function $W(\boldsymbol{u}, \boldsymbol{y})$ of the channel's input and output sequences $\boldsymbol{u}$, $\boldsymbol{y}$ respectively, which is concentrated asymptotically around its mean value with respect to the joint distribution $\Pr(\boldsymbol{u}, \boldsymbol{y})$

$$\Pr\left(\left|W(\boldsymbol{u},\boldsymbol{y}) - E_{\Pr(\boldsymbol{u},\boldsymbol{y})}\left[W(\boldsymbol{u},\boldsymbol{y})\right]\right| < N\epsilon\right) \xrightarrow{N\longrightarrow\infty} 1. \quad (41)$$

The input output pair $(\boldsymbol{u}, \boldsymbol{y})$ is called weakly ϵ-typical, if

$$W(\boldsymbol{u},\boldsymbol{y}) \geq E_{\Pr(\boldsymbol{u},\boldsymbol{y})}\left[W(\boldsymbol{u},\boldsymbol{y})\right] - N\epsilon. \quad (42)$$

This is in analogy to Forney [17], where strong ϵ-typicality is defined as

$$\log\frac{p(\boldsymbol{u},\boldsymbol{y})}{p(\boldsymbol{u})p(\boldsymbol{y})} \geq I(\boldsymbol{u};\boldsymbol{y}) - N\epsilon.$$

Under the weakly typical decoding rule and the random coding setup described in Section I, an error occurs given that message m is transmitted, either if codeword $\boldsymbol{u}_m$ is selected form the ensemble $\boldsymbol{C}$ such that $(\boldsymbol{u}_m, \boldsymbol{y})$ does not satisfy (42) or if there exists another message $m' \neq m$ for which $\boldsymbol{u}_{m'}$ is selected independently of $\boldsymbol{u}_m$ such that $(\boldsymbol{u}_{m'}, \boldsymbol{y})$ satisfies (42). Thus, the average error decoding probability, given that m is transmitted, equals

$$\bar{P}_{e,m} = \Pr\left((\boldsymbol{u}_m,\boldsymbol{y})\ \text{not}\ \epsilon\text{-typical}\bigcup_{m'\neq m}(\boldsymbol{u}_{m'},\boldsymbol{y})\ \epsilon\text{-typical}\right) \quad (43)$$

and is upper bounded due to the union bound as

$$\bar{P}_{e,m} \leq \Pr\left((\boldsymbol{u}_m, \boldsymbol{y}) \text{ not } \epsilon\text{-typical}\right) + \sum_{m' \neq m} \Pr\left((\boldsymbol{u}_{m'}, \boldsymbol{y})\ \epsilon\text{-typical}\right). \quad (44)$$

Furthermore, definition (42) allows the above bound to be equivalently expressed as

$$\bar{P}_{e,m} \leq \Pr\left(W(\boldsymbol{u}_m, \boldsymbol{y}) < E\left[W(\boldsymbol{u}, \boldsymbol{y})\right] - N\epsilon\right) + \sum_{m' \neq m} \Pr\left(W(\boldsymbol{u}_{m'}, \boldsymbol{y}) \geq E\left[W(\boldsymbol{u}, \boldsymbol{y})\right] - N\epsilon\right). \quad (45)$$

Next, we apply the Chernoff bound [11, eq. (5.4.11)] to both terms on the right-hand side of (45), to obtain for $\lambda_1, \lambda > 0$

$$\begin{aligned}\bar{P}_{e,m} &\leq E\big[\exp\big(\lambda_1\big(E\big[W(\boldsymbol{u}, \boldsymbol{y})\big] - N\epsilon - W(\boldsymbol{u}_m, \boldsymbol{y})\big)\big)\big] \\ &\quad + \sum_{m' \neq m} E_{Q(\boldsymbol{u}_{m'})\Pr(\boldsymbol{y})} \\ &\quad \times \left[\exp(\lambda(W(\boldsymbol{u}_{m'}, \boldsymbol{y}) - E[W(\boldsymbol{u}, \boldsymbol{y})] + N\epsilon))\right]. \quad (46)\end{aligned}$$

The product probability $Q(\boldsymbol{u}_{m'})\Pr(\boldsymbol{y})$ in the innermost term on the right-hand side of (46) is a direct consequence of the random coding setup. Indeed, $\boldsymbol{u}_{m'}$ is independent of $\boldsymbol{u}_m$ and consequently of $\boldsymbol{y}$. Noting that $\boldsymbol{u}'_m, \boldsymbol{u}_m$ are dummy variables in the above mean values, (46) satisfies

$$\begin{aligned}\bar{P}_{e,m} &\leq E\left[\exp\left(\lambda_1\left(E\left[W(\boldsymbol{u}, \boldsymbol{y})\right] - N\epsilon - W(\boldsymbol{u}, \boldsymbol{y})\right)\right)\right] \\ &\quad + (|\mathcal{M}| - 1)E_{Q(\boldsymbol{u}')\Pr(\boldsymbol{y})} \\ &\quad \times \left[\exp(\lambda(W(\boldsymbol{u}', \boldsymbol{y}) - E[W(\boldsymbol{u}, \boldsymbol{y})] + N\epsilon))\right]. \quad (47)\end{aligned}$$

First, we handle the second term on the right-hand side of (47). Note that $\boldsymbol{y}$ is uniquely defined by the transmitted codeword $\boldsymbol{u}$ and the noise sequence $\boldsymbol{\nu}$, and thus, $W(\boldsymbol{u}', \boldsymbol{y})$ is expressed as a Doob martingale process [19, pp. 22]. Specifically, it holds

$$W(\boldsymbol{u}', \boldsymbol{y}) = \sum_{i=1}^{N} \Delta'_i + E_{Q(\boldsymbol{u})\Pr(\boldsymbol{y})}\left[W(\boldsymbol{u}', \boldsymbol{y})\right] \quad (48)$$

where $\{\Delta'_i\}_{i=1}^{N}$ is the martingale difference sequence with respect to the joint filter F'_i

$$\emptyset = F'_0 \subseteq F'_1 \subseteq \cdots \subseteq F'_N$$

where

$$F'_i = \{u'_1, \ldots, u'_i, u_1, \ldots, u_i, \nu_1, \ldots, \nu_i\} \quad (49)$$

and the sequence $\{\Delta'_i\}_{i=1}^{N}$ is defined as

$$\begin{aligned}\Delta'_i &= X'_i - X'_{i-1} \\ X'_i &= E\left[W(\boldsymbol{u}', \boldsymbol{y})|F'_i\right] \\ X'_N &= W(\boldsymbol{u}', \boldsymbol{y}) \\ X'_0 &= E\left[W(\boldsymbol{u}', \boldsymbol{y})\right]. \quad (50)\end{aligned}$$

The mean values appearing in martingale differences (50) are with respect to the product probability $Q(\boldsymbol{u}')\Pr(\boldsymbol{y})$. Furthermore, the mean value of the second summand on the right-hand side of (47) is equivalently expressed due to (48)–(50) as

$$E_{Q(\boldsymbol{u}')\Pr(\boldsymbol{y})}\left[\exp\left(\lambda\left(\sum_{i=1}^{N}\Delta'_i - \mathcal{K}'_\epsilon(W, \mathbf{Q}, N)\right)\right)\right] \quad (51)$$

where

$$\begin{aligned}&\mathcal{K}'_\epsilon(W, \mathbf{Q}, N) \\ &\quad = E_{\Pr(\boldsymbol{u}, \boldsymbol{y})}\left[W(\boldsymbol{u}, \boldsymbol{y})\right] - E_{Q(\boldsymbol{u})\Pr(\boldsymbol{y})}\left[W(\boldsymbol{u}, \boldsymbol{y})\right] - N\epsilon. \quad (52)\end{aligned}$$

In analogy to (22) and (35), let the conditional deviations and the maximum conditional variances satisfy, respectively

$$\begin{aligned}dev^+(\Delta'_i) &= \max_{u_i, u'_i, \nu_i} \Delta'_i \leq b' \\ \max_{F'_{i-1}} var(\Delta'_i|F'_{i-1}) &= \max_{F'_{i-1}} E\left[\left(X'_i - X'_{i-1}\right)^2 |F'_{i-1}\right] \\ &\leq V'. \quad (53)\end{aligned}$$

Then, for $V' + b'\mathcal{K}'_\epsilon(W, \mathbf{Q}, N)/N \geq 0$, according to [22, Th. 2.1], (51) is upper bounded as

$$\begin{aligned}&E_{Q(\boldsymbol{u}')\Pr(\boldsymbol{y})}\left[\exp\left(\lambda\left(\sum_{i=1}^{N}\Delta'_i - \mathcal{K}'_\epsilon(W, \mathbf{Q}, N)\right)\right)\right] \\ &\quad \leq \exp\left(-ND_{KL}\left(\frac{\frac{\hat{V} + b'\mathcal{K}'_\epsilon(W, \mathbf{Q}, N)}{N}}{(b')^2 + V'} \middle\| \frac{V'}{(b')^2 + V'}\right)\right) \quad (54)\end{aligned}$$

where

$$D_{KL}(a\|p) = (1-a)\log\frac{1-a}{1-p} + a\log\frac{a}{p}$$

denotes the Kullback–Leibler divergence.

Next, we consider the first term on the right-hand side of (47). We follow the same steps as in (48)–(54) where the new martingale difference sequence $\{\Delta'_{1,i}\}_{i=1}^{N}$ is with respect to the filters $F'_{1,i}$ consisting only of the corresponding sequences $\boldsymbol{u}, \boldsymbol{\nu}$. Since

$$W(\boldsymbol{u}, \boldsymbol{y}) = \sum_{i=1}^{N} \Delta'_{1,i} + E\left[W(\boldsymbol{u}, \boldsymbol{y})\right]$$

the first term on the right-hand side of (47) satisfies

$$\begin{aligned}&E\left[\exp\left(\lambda_1\left(E\left[W(\boldsymbol{u}, \boldsymbol{y})\right] - N\epsilon - W(\boldsymbol{u}, \boldsymbol{y})\right)\right)\right] \\ &\quad = E\left[\exp\left(\lambda_1\left(-\sum_{i=1}^{N}\Delta'_{1,i} - N\epsilon\right)\right)\right]. \quad (55)\end{aligned}$$

Since the martingale differences $-\Delta'_{1,i}$ behave in a similar manner to $\Delta'_{1,i}$ regarding the upper bound of $\max dev^+$, there exist constants b'_1, V'_1 in analogy to (53) such that applying again [22, Th. 2.1], we have

$$\begin{aligned}&E\left[\exp\left(\lambda_1\left(E\left[W(\boldsymbol{u}, \boldsymbol{y})\right] - N\epsilon - W(\boldsymbol{u}, \boldsymbol{y})\right)\right)\right] \\ &\quad \leq \exp\left(-ND_{KL}\left(\frac{\hat{V}_1 + b'_1\epsilon}{b'^2_1 + V'_1} || \frac{V'_1}{b'^2_1 + V'_1}\right)\right). \quad (56)\end{aligned}$$

The right-hand side of (56) approaches 0 as N becomes large, and thus, from (55) and (56), there exists an $\epsilon_1 > 0$ arbitrarily small such that

$$E\left[\exp\left(\lambda_1\left(E\left[W(\boldsymbol{u},\boldsymbol{y})\right] - N\epsilon - W(\boldsymbol{u},\boldsymbol{y})\right)\right)\right] \le \epsilon_1. \quad (57)$$

Replacement of (54) and (57) to (46) results in

$$\bar{P}_{e,m} \le \epsilon_1 + |\mathcal{M}| \cdot \exp\left(-ND_{KL}\left(\frac{\frac{V'+b'\mathcal{K}'_\epsilon(W,\mathbf{Q},N)}{N}}{(b')^2+V'}\middle\|\frac{V'}{(b')^2+V'}\right)\right). \quad (58)$$

For the bound (58) to be exponentially decreasing with respect to N, $\mathcal{K}'_\epsilon(W,\mathbf{Q},N)$ must be of order N. Thus, supposing that $\mathcal{K}'_\epsilon(W,\mathbf{Q},N)/N = \mathcal{K}'_W(Q)$, (58) becomes

$$\bar{P}_{e,m} \le \epsilon_1 + \exp\left(-N\left(D_{KL}\left(\frac{V'+b'\mathcal{K}'_W(Q)}{(b')^2+V'}\middle\|\frac{V'}{(b')^2+V'}\right) - R\right)\right). \quad (59)$$

Corollary 2: Considering the transmission of information over the nonlinear Volterra channel (2), all rates below

$$\max_Q D_{KL}\left(\frac{V'+b'\mathcal{K}'_W(Q)}{(b')^2+V'}\middle\|\frac{V'}{(b')^2+V'}\right)$$

are achievable, under the weakly typical set decoding rule described by function W and the assumption that $\mathcal{K}'_\epsilon(W,\mathbf{Q},N)/N = \mathcal{K}'_W(Q) > 0$.

In the next subsection, the above analysis is properly modified to treat a specific unbounded function $W(\boldsymbol{u},\boldsymbol{y})$ under the assumption that the noise samples are i.i.d. and normally distributed $\mathcal{N}(0,\sigma_\nu^2)$.

B. Weakly Typical Set Decoding for Nonlinear Systems

Note that the exponent in (59) replaces mutual information $I(\boldsymbol{u};\boldsymbol{y})$ in the classical treatment of typical set decoding as an average measure of dependence. Function $W(\boldsymbol{u},\boldsymbol{y})$ can be selected so that the corresponding exponent denotes a normalized correlation between the channel's input and output. More specifically, let

$$W(\boldsymbol{u},\boldsymbol{y}) = (D\boldsymbol{u})^T\boldsymbol{y} = (D\boldsymbol{u})^T(D\boldsymbol{u}+\boldsymbol{\nu}). \quad (60)$$

Since the components of all codewords are i.i.d, it holds through the analysis in (18)–(20) and (60)

$$\begin{aligned}\mathcal{K}'_\epsilon(W,\mathbf{Q},N) &= E\left[D\boldsymbol{u}^T(D\boldsymbol{u}+\boldsymbol{\nu})\right] \\ &\quad - E\left[D\boldsymbol{u}^T\right]E\left[D\boldsymbol{u}+\boldsymbol{\nu}\right] - N\epsilon \\ &= E\left[D\boldsymbol{u}^T D\boldsymbol{u}\right] \\ &\quad - E\left[D\boldsymbol{u}\right]^T E\left[D\boldsymbol{u}\right] - N\epsilon \\ &= E\left[\|D\boldsymbol{u}\|_2^2\right] - \|E\left[D\boldsymbol{u}\right]\|_2^2 - N\epsilon \\ &= 2N\mathcal{D}_v(Q) - N\epsilon. \end{aligned} \quad (61)$$

If we set

$$d_i = [Du']_i([Du]_i + \nu_i), \quad 1 \le i \le N \quad (62)$$

then the martingale differences Δ'_i given by (50) satisfy

$$\begin{aligned}\Delta'_i &= X'_i - X'_{i-1} \\ &= \sum_{j=1}^{N} E\left[d_j|F'_i\right] - \sum_{j=1}^{N} E\left[d_j|F'_{i-1}\right]. \end{aligned} \quad (63)$$

Expanding X'_i in the same way as in the proof of Lemma 1, we obtain

$$X'_i = \sum_{j=1}^{i-1} E\left[d_j|F'_i\right] + \sum_{j=i}^{i+\mu} E\left[d_j|F'_i\right] + \sum_{j=i+\mu+1}^{N} E\left[d_j|F'_i\right]. \quad (64)$$

Under the same assumptions of Lemma 1, all samples d_j (62), $j \ge i+\mu+1$ and $j \le i-1$ are independent of the samples u_i, u'_i (the noise samples are i.i.d). Consequently

$$X'_i = \sum_{j=1}^{i-1} E\left[d_j|F'_{i-1}\right] + \sum_{j=i}^{i+\mu} E\left[d_j|F'_i\right] + \sum_{j=i+\mu+1}^{N} E\left[d_j|F'_{i-1}\right]. \quad (65)$$

Combining (62)–(65), we obtain

$$\Delta'_i = \sum_{j=i}^{i+\mu} E\left[d_j|F'_i\right] - \sum_{j=i}^{i+\mu} E\left[d_j|F'_{i-1}\right]. \quad (66)$$

The martingale differences Δ'_i are further expanded due to independence of the transmitted and noise sequences as

$$\begin{aligned}\Delta'_i &= [Du]_i[Du']_i + [Du']_i\nu_i \\ &\quad + \sum_{j=i+1}^{i+\mu}\Big(E\left[[Du]_j[Du']_j|F'_i\right] \\ &\qquad + E\left[[Du']_j|F'_i\right]E\left[\nu_j|F'_i\right]\Big) \\ &\quad - \sum_{j=i}^{i+\mu}\Big(E\left[[Du]_j[Du']_j|F'_{i-1}\right] \\ &\qquad + E\left[[Du']_j|F'_{i-1}\right]E\left[\nu_j|F'_{i-1}\right]\Big). \end{aligned} \quad (67)$$

Since the noise samples $\nu_i,\ i \in [1,N]$ are i.i.d. and normally distributed $\mathcal{N}(0,\sigma_v^2)$, it holds

$$E\left[\nu_j|F'_i\right] = 0, \quad j \ne i \quad (68)$$

and thus, Δ'_i in (67) satisfies through (68)

$$\Delta'_i = [Du']_i\nu_i + \Delta''_i \quad (69)$$

where

$$\Delta''_i = \sum_{j=i}^{i+\mu}\left(E\left[[Du]_j[Du']_j\,|F'_i\right] - E\left[[Du]_j[Du']_j|F'_{i-1}\right]\right).$$

Due to the peak amplitude constraint (3) and its direct consequence (5), we have as in the proof of Lemma 1

$$\Delta_i'' \leq |\Delta_i''| \leq 2(\mu+1)g_D(r)^2 = b' \tag{70}$$

while

$$var(\Delta_i''|F_{i-1}') \leq V' = \left(2(\mu+1)g_D(r)^2\right)^2 = (b')^2. \tag{71}$$

Lemma 2: Suppose that all noise samples $\nu_i,\ i \in [1,N]$ are normally distributed $\mathcal{N}(0,\sigma_\nu^2)$. Then for any $\lambda > 0$ and $0 < k < 1$

$$E_{u_i,u_i'}\left[\exp\left(\lambda\Delta_i'\right)|F_{i-1}'\right] \leq \exp\left(\frac{\lambda^2}{2k}g_D(r)^2\sigma_\nu^2\right) \cdot \left(\frac{1}{2}\exp\left(-\frac{\lambda}{1-k}b'\right)+\frac{1}{2}\exp\left(\frac{\lambda}{1-k}b'\right)\right). \tag{72}$$

The proof of Lemma 2 is provided in the Appendix and takes advantage of the same technique used for the proof of [22, Th. 2.1] that results in (54).

Theorem 2: Let the transmission of an arbitrary set of messages $\mathcal{M}$ over an additive noise nonlinear channel, under the same random encoding setup of Theorem 1. Let also the noise samples be i.i.d. and normally distributed $\mathcal{N}(0,\sigma_\nu^2)$, independent from the channel input. Then, for any ϵ, $\epsilon_1 > 0$ arbitrarily small constants, the average error decoding probability $\bar{P}_e$ is upper bounded as

$$\bar{P}_e \leq \epsilon_1 + e^{-N\left(E_c'(Q,D,\sigma_\nu^2)-R\right)} \tag{73}$$

where

$$\begin{aligned} &E_c'(Q,D,\sigma_\nu^2) \\ &= \max_{0<k<1}\max_{0<\lambda}\lambda(2\mathcal{D}_v(Q)-\epsilon)-\frac{\lambda^2}{2k}g_D(r)^2\sigma_\nu^2 \\ &\quad -\ln\left(\frac{1}{2}\exp\left(-\frac{\lambda}{1-k}b'\right)+\frac{1}{2}\exp\left(\frac{\lambda}{1-k}b'\right)\right) > 0. \end{aligned} \tag{74}$$

Proof: Due to the specific selection of $W(\boldsymbol{u},\boldsymbol{y})$ in (60) and property (61), the upper bound in (47) satisfies for $\lambda_1, \lambda > 0$ in analogy to (51) and (55)

$$\begin{aligned} \bar{P}_{e,m} &\leq \exp(-N\epsilon)\,E\left[\exp\left(-\lambda_1\sum_{i=1}^{N}\Delta_{1,i}'\right)\right] \\ &+\exp\left(NR-\lambda N(2\mathcal{D}_v(Q)-\epsilon)\right)E\left[\exp\left(\lambda\sum_{i=1}^{N}\Delta_i'\right)\right]. \end{aligned} \tag{75}$$

The second mean value in the right-hand side of (75) is equivalently expressed as

$$\begin{aligned} &E\left[\exp\left(\lambda\sum_{i=1}^{N}\Delta_i'\right)\right] \\ &= E\left[\exp\left(\sum_{i=1}^{N-1}\Delta_i'\right)E_{u_N,u_N'}\left[\exp\left(h\Delta_i'\right)|F_{i-1}'\right]\right]. \end{aligned} \tag{76}$$

We apply Lemma 2 in the inner mean value on the right-hand side of (76). Then

$$\begin{aligned} &E\left[\exp\left(\lambda\sum_{i=1}^{N}\Delta_i'\right)\right] \\ &\leq \exp\left(\frac{\lambda^2}{2k}g_D(r)^2\sigma_\nu^2\right) \\ &\quad\cdot\left(\frac{1}{2}\exp\left(-\frac{\lambda}{1-k}b'\right)+\frac{1}{2}\exp\left(\frac{\lambda}{1-k}b'\right)\right) \\ &\quad\cdot E\left[\exp\left(\sum_{i=1}^{N-1}\Delta_i'\right)\right]. \end{aligned} \tag{77}$$

Applying the above argument N times and combining (76) with (77), we have

$$E\left[\exp\left(\lambda\sum_{i=1}^{N}\Delta_i'\right)\right] \leq \exp\left(-N\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)\right) \tag{78}$$

where

$$\begin{aligned} &\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2) \\ &\quad=\lambda(2\mathcal{D}_v(Q)-\epsilon)-\frac{\lambda^2}{2k}g_D(r)^2\sigma_\nu^2 \\ &\qquad-\ln\left(\frac{1}{2}\exp\left(-\frac{\lambda}{1-k}b'\right)+\frac{1}{2}\exp\left(\frac{\lambda}{1-k}b'\right)\right). \end{aligned} \tag{79}$$

For fixed Q, k, the first and second derivatives of $\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)$ with respect to $\lambda \geq 0$ satisfy respectively

$$\begin{aligned} &\frac{\vartheta\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)}{\vartheta\lambda} \\ &\quad=2\mathcal{D}_v(Q)-\epsilon-\frac{\lambda(g_D(r))^2\sigma_\nu^2}{k}-\frac{b'\left(1-e^{-\frac{2b'\lambda}{1-k}}\right)}{(1-k)\left(1+e^{-\frac{2b'\lambda}{1-k}}\right)}, \\ &\frac{\vartheta^2\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)}{\vartheta\lambda^2} \\ &\quad=-\frac{4(b')^2e^{\frac{2b'\lambda}{1-k}}k+\left(1+e^{\frac{2b'\lambda}{1-k}}\right)^2(1-k)^2(g_D(r))^2\sigma_\nu^2}{\left(1+e^{\frac{2b'\lambda}{1-k}}\right)^2(1-k)^2k}. \end{aligned}$$

The second derivative of $\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)$ with respect to $\lambda > 0$ is strictly negative. Therefore, $\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)$ is strictly concave, and thus, the first derivative is decreasing with respect to λ. Furthermore, the first derivative satisfies

$$\left.\frac{\vartheta\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)}{\vartheta\lambda}\right|_{\lambda=0} = 2\mathcal{D}_v(Q)-\epsilon > 0$$

so that there exists a region of λ where

$$2\mathcal{D}_v(Q)-\epsilon \geq \frac{\vartheta\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2)}{\vartheta\lambda} \geq 0.$$

Thus, we conclude that

$$\max_{\lambda>0}\psi(\lambda,k,b',\mathcal{D}_v(Q),\sigma_\nu^2) > 0$$

and consequently $E_c'(Q,D,\sigma_\nu^2) > 0$. For the first summand on the right-hand side of (75), following the same steps as in

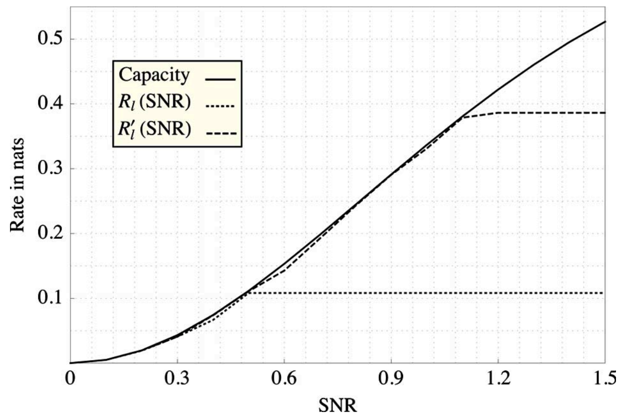

Fig. 1. Comparison between channel capacity $\max_Q I(U;Y)$ and maximal achievable rates $R_l(\mathrm{SNR})$ (82) and $R_l'(\mathrm{SNR})$ (84) with respect to the bound (32) for binary input u, memoryless channels with additive gaussian noise $\nu \sim \mathcal{N}(0, \sigma_\nu^2)$.

(76)–(78) and noting that $-\Delta'_{1,i}$ behaves similar to $\Delta'_{1,i}$ for the upper bound of $\max dev^+$

$$\exp(-N\epsilon)\, E\left[\exp\left(-\lambda_1 \sum_{i=1}^{N} \Delta'_{1,i}\right)\right] \leq \exp(-N \max_{0<k<1} \max_{0<\lambda} \psi(\lambda, k, b', \epsilon, \sigma_\nu^2)). \quad (80)$$

Since the exponent is strictly negative, for arbitrary large N

$$\exp(-N\epsilon)\, E\left[\exp\left(-\lambda_1 \sum_{i=1}^{N} \Delta'_{1,i}\right)\right] \leq \epsilon_1. \quad (81)$$

The proof is complete if we replace (78) and (81) in (75).

Corollary 3: Considering the transmission of information over a nonlinear Volterra additive gaussian noise channel, all rates below $\max_Q E'_c(Q, D, \sigma_\nu^2)$(74) are achievable for the weakly typical set decoding rule described by (60).

The tightness of the random coding exponent, given by Theorem 2, depends on lower bounds for the error decoding probability of the form provided in [7], for the specific functions $W(\boldsymbol{u}, \boldsymbol{y})$.

IV. Applications

In this section, three special cases of the nonlinear system (1) are considered and the martingale based random coding bounds are illustrated. First the additive white Gaussian noise channel with finite inputs and real outputs is studied. This has the form $y = u + \nu$ where $\nu \sim \mathcal{N}(0, \sigma_\nu^2)$ and u is binary with values in $\{-A, A\}$. The channel capacity C is obtained by maximizing the mutual information $I(U;Y)$ with respect to $Q = \Pr(u = A)$. The capacity C is a well defined function of the signal to noise ratio $\mathrm{SNR} = A^2/\sigma_\nu^2$, although not given by a closed form solution. It is depicted in Fig. 1. The achievable rate with respect to the bound (32) and Corollary 1 is also plotted. Since $\mathcal{D}_v(Q) = 4A^2\pi(1-\pi)$ and $\kappa = 4A^2$ the achievable rate is expressed in terms of the signal to noise ratio as

$$R_l(\mathrm{SNR}) = \begin{cases} \frac{\mathrm{SNR}}{2} - \left(e^{\mathrm{SNR}} - 1 - \mathrm{SNR}\right), & \mathrm{SNR} < \ln\frac{3}{2} \\ \frac{3}{2}\ln\frac{3}{2} - \frac{1}{2}, & \mathrm{SNR} \geq \ln\frac{3}{2}. \end{cases} \quad (82)$$

The simplicity of the linear memoryless channel filter enables us to do better. Indeed we calculate the terms in (20), (22), and (35) explicitly and use them in (38) instead of the more crude bounds in Lemma 1. Thus

$$\begin{aligned} D_v &= 4A^2\pi(1-\pi) \\ \max_i dev^+(\Delta_i) &= 4A^2(1 - 2\pi(1-\pi)) \\ \frac{1}{N}\max_{F_N}\sum_{i=1}^{N} var(\Delta_i|F_{i-1}) &= 32A^2\pi(1-\pi) \\ &\quad \times (1 - 2\pi(1-\pi)). \end{aligned} \quad (83)$$

Then according to (40) and Corollary 1, maximal achievable rates based on martingales are expressed with respect to SNR as

$$R_l'(\mathrm{SNR}) = \max_{0\leq\pi\leq 1}\max_{0\leq\rho\leq 1} 2\pi(1-\pi)\left(\rho\mathrm{SNR} - \frac{e^{\rho\mathrm{SNR}(1-2\pi(1-\pi))} - 1 - \rho\mathrm{SNR}(1-2\pi(1-\pi))}{1-2\pi(1-\pi)}\right). \quad (84)$$

In Fig. 1, the channel capacity and the maximal achievable rates (82) and (84) are numerically evaluated in relation with SNR. The analytic calculation of the parameters in (83) extends the SNR region for which the martingale achievable rates matches channel capacity. Furthermore, powerful capacity approaching codes that achieve low error decoding probability, such as turbo codes presented in [23], are shown to exist for the specific range of SNR values.

We consider next the transmission of coded modulated information over nonlinear channels with the Volterra kernels depicted in Tables I and II, respectively. The Volterra systems D_1,

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

TABLE II
Kernels of the 4th Order Volterra System D_2 With Memory 3

kernel	$h_1(0)$	$h_1(1)$	$h_1(2)$	$h_1(3)$
value	1.0	0.5	-0.8	1.6

kernel	$h_2(0,0)$	$h_2(1,1)$	$h_2(0,1)$
value	1.0	0.3	0.6

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

kernel	$h_3(0,1,1)$	$h_3(0,1,2)$
value	0.8	0.6

kernel	$h_4(0,0,0,0)$	$h_4(0,0,0,1)$	$h_4(0,0,1,1)$
value	1.0	1.2	0.8

kernel	$h_4(0,0,1,2)$	$h_4(0,1,2,3)$
value	-0.5	0.6

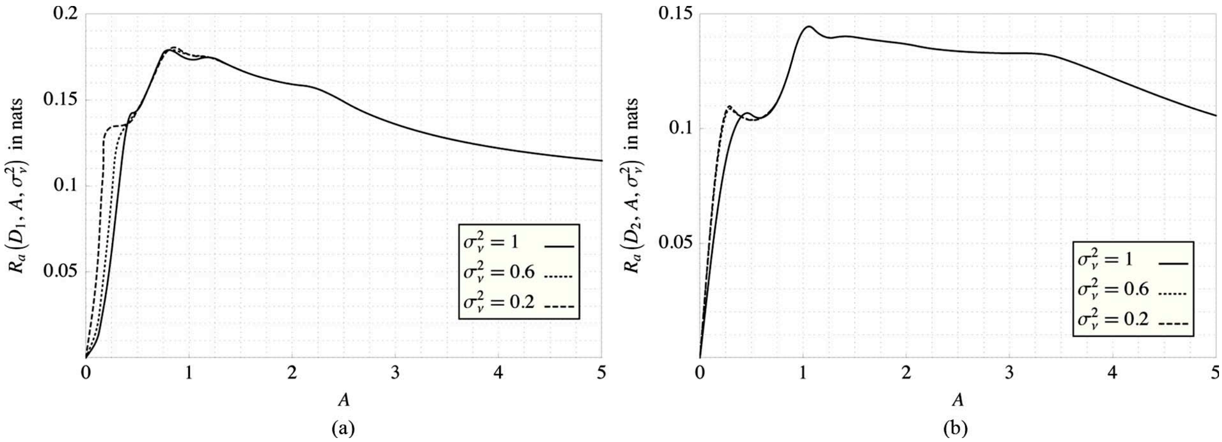

Fig. 2. Maximal achievable rates $R_a(D, A, \sigma_\nu^2)$ with respect to the bound (32) under ML decoding for BPSK modulated transmission over additive nonlinear gaussian noise channels of zero mean and variance σ_ν^2. (a) Variation of $R_a(D_1, A, \sigma_\nu^2)$ with respect to the modulation parameter A for the Volterra system D_1 of Table I. (b) Variation of $R_a(D_2, A, \sigma_\nu^2)$ with respect to the modulation parameter A for the Volterra system D_2 of Table II.

D_2 are of corresponding order $d_1 = 3$ and $d_2 = 4$, with respective memories $\mu_1 = 2$ and $\mu_2 = 3$. Moreover, all Volterra kernels up to order 3 are the same for D_1, D_2 apart from $h_2(1,1)$ and $h_1(3)$. The components of each codeword $u_m(i)$, $1 \le i \le N$, $1 \le m \le |\mathcal{M}|$ are chosen independently from the set $\{-A, A\}$ with probability Q (BPSK modulated). For the additive gaussian noise channel, the pair of bounds (b, V) (35) on the maximum conditional deviation and maximum sum of conditional variance is calculated analytically for the martingale differences Δ_i (15). Specifically, for the maximum sum of conditional variances, since $var(\Delta_i|F_{i-1}) > 0$, $1 \le i \le N$, it holds

$$\max_{F_N} \sum_{i=1}^{N} var(\Delta_i|F_{i-1}) \le \sum_{i=1}^{N} \max_{F_N} var(\Delta_i|F_{i-1}) = \sum_{i=1}^{N} \max_{F_{i-1}} var(\Delta_i|F_{i-1}).$$

Since the channel is stationary, we calculate analytically only the terms $\tilde{V} = \max_{F_{i-1}} var(\Delta_i|F_{i-1})$ and $\tilde{b} = dev^+(\Delta_i)$. These values are then utilized in the exponential martingale inequality (38), instead of the upper bounds of Lemma 1. According to Theorem 1 and Corollary 1, maximal achievable rates with respect to the bound (32) are given by

$$R_a(D, A, \sigma_\nu^2) = \max_{Q} \max_{0 \le \rho \le 1} \left(\rho \frac{D_v}{2\sigma_\nu^2} - \left(\exp\left(\frac{\rho}{4\sigma_\nu^2} \tilde{b} \right) - 1 - \frac{\rho}{4\sigma_\nu^2} \tilde{b} \right) \frac{\tilde{V}}{(\tilde{b})^2} \right). \tag{85}$$

For the Volterra systems D_1, D_2, the achievable rates $R_a(D_1, \sigma_\nu^2)$ and $R_a(D_2, \sigma_\nu^2)$ respectively are depicted, in terms of the modulation parameter A and for various noise variances σ_ν^2, in Fig. 2. Achievable rates get their maximum values for low values of A, while as the latter increases, the rates are noted to decrease and behave in a similar way for all the studied noise variances. Under the typical coding theorem setup and the specific selection $W(\boldsymbol{u}, \boldsymbol{y}) = D\boldsymbol{u}^T(D\boldsymbol{u} + \boldsymbol{\nu})$, we calculate again analytically the quantities $\tilde{V}' = \max_{F'_{i-1}} var(\Delta''_i|F'_{i-1})$ and $\tilde{b}' = dev^+(\Delta''_i)$ and utilize them properly under the guidance of (93) in Lemma 2 instead of (70) and (71). According to Theorem 2 and Corollary 3, maximal achievable rates with respect to bound (73) are expressed as

$$R'_a(D, A, \sigma_\nu^2) = \max_{Q} \max_{0<k<1} \max_{0<\lambda} \left(\lambda(2D_v - \epsilon) - \frac{\lambda^2}{k} g_D(r)^2 \sigma_\nu^2 - \ln\left(\frac{(\tilde{b}')^2}{(\tilde{b}')^2 + \tilde{V}'} e^{-\frac{\lambda}{1-k}\frac{\tilde{V}'}{\tilde{b}'}} + \frac{\tilde{V}'}{(\tilde{b}')^2 + \tilde{V}'} e^{\frac{\lambda}{1-k}\tilde{b}'} \right) \right). \tag{86}$$

The maximal achievable rates $R'_a(D_1, A, \sigma_\nu^2)$ and $R'_a(D_2, A, \sigma_\nu^2)$ for the respective Volterra systems D_1, D_2 are illustrated, in terms of the modulation parameter A and for various noise variances σ_ν^2, in Fig. 3. The corresponding functions $g_{D_1}(x)$, $g_{D_2}(x)$ (6) used for the calculation of the above achievable rates are given due to Tables I and II as

$$g_{D_1}(x) = 2.3x + 1.9x^2 + 4.1x^3$$
$$g_{D_2}(x) = 3.9x + 1.9x^2 + 4.1x^3 + 4.1x^4.$$

Comparing Fig. 2(a) with Figs. 3(a) and 2(b) with Fig. 3(b), we note that maximal achievable rates under the selected weakly typical set decoding are higher than those under optimum ML decoding for large values of the modulation parameter A. The main reason for this difference is the utilization of a tighter exponential martingale inequality in Lemma 2 compared to the one used in Lemma 1. Thus, tighter achievable rates both under ML and weakly typical set decoding are expected to be achieved by optimal martingale inequalities. Numerical optimization for the calculation of (85) and (86) was performed in MATHEMATICA using the Nelder–Mead algorithm.

V. Conclusion

Random coding theorems for nonlinear communication channels are presented in this work. Martingale theory and corresponding martingale exponential inequalities are properly utilized for this purpose. Apart from the additive gaussian noise channel setup, the martingale approach allows us to define weakly typical set decoding theorems for nonlinear channels with unknown transition probabilities. Specifically, due to computation reasons mainly, the information density is replaced by

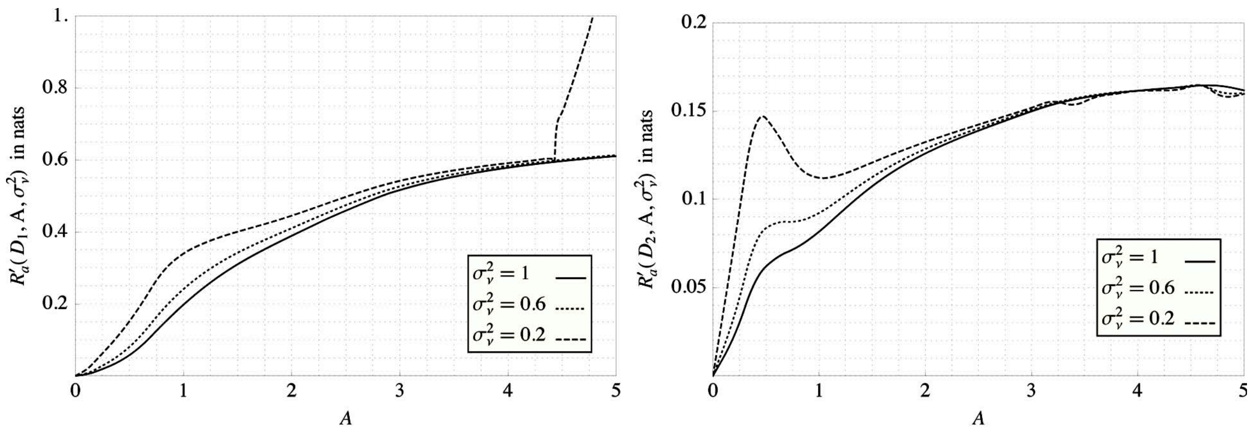

Fig. 3. Maximal achievable rates $R'_a(D, A, \sigma_\nu^2)$ with respect to the bound (73) under the weak typical coding theorem and $W(\boldsymbol{u}, \boldsymbol{y}) = D\boldsymbol{u}^T(D\boldsymbol{u} + \boldsymbol{\nu})$ for BPSK modulated transmission over additive nonlinear gaussian noise channels of zero mean and variance σ_ν^2. (a) Variation of $R'_a(D_1, A, \sigma_\nu^2)$ with respect to the modulation parameter A for the Volterra system D_1 of Table I. (b) Variation of $R'_a(D_2, A, \sigma_\nu^2)$ with respect to the modulation parameter A for the Volterra system D_2 of Table II.

a simpler generalized function. The optimality of the deduced theorems depend on lower bounds for the selected functions and are currently investigated. The analysis is proper also for nonstationary environments since the martingale differences capture the possible changes in the parameters of the channels.

The proposed technique can also be used for linear communication channels with intersymbol interference (ISI). Due to the union bound effect apparent in our analysis, we expect that the deduced coding exponents for the gaussian noise case, are only lower bounds to the true exponents of the channels. Thus, in linear channels with ISI, the maximization of our random coding exponent constitutes only a lower bound in the channel's capacity and our coding exponent is lower than the corresponding deduced from the finite state approach. Nevertheless, for nonlinear channels, little is known, to our knowledge, about their random coding theorems. Thus, the current methodology is a first approach towards the aforementioned problem. Moreover, possible ways, in relation with martingale theory, to leverage the union bound bound effect are currently studied. They are expected to yield tighter random coding exponents for linear and nonlinear channels with memory.

Appendix
Proof of Lemma 2

We expand the term on the left-hand side of (72) as

$$E\left[\exp\left(\lambda\Delta'_i\right)|F'_{i-1}\right] = E\Big[\exp\left(\lambda[Du']_i\nu_i\right)\cdot\exp\left(\lambda\Delta''_i\right)|F'_{i-1}\Big] \quad (87)$$

and apply the variant of Hölder's inequality [11, eq. (4.15(c))] to the right hand side of (87). Then due to the expansion in (69) it holds for $0 < k < 1$

$$\begin{aligned} E\left[\exp\left(\lambda\Delta'_i\right)|F'_{i-1}\right] &\le \left(E\left[\exp\left(\frac{\lambda}{k}[Du']_i\nu_i\right)\Big|F'_{i-1}\right]\right)^k \\ &\quad\cdot\left(E\left[\exp\left(\frac{\lambda}{1-k}\Delta''_i\right)\Big|F'_{i-1}\right]\right)^{1-k}. \end{aligned} \quad (88)$$

Since the noise samples $\nu_i,\ i \in [1, N]$ are mutual independent and independent from the transmitted sequence, the first term of the product on the right-hand side of (88) satisfies

$$\begin{aligned} &\left(E\left[\exp\left(\frac{\lambda}{k}[Du']_i\nu_i\right)\Big|F'_{i-1}\right]\right)^k \\ &\quad= \left(E\left[E_{\nu_i}\left[\exp\left(\frac{\lambda}{k}[Du']_i\nu_i\right)\right]\Big|F'_{i-1}\right]\right)^k. \end{aligned} \quad (89)$$

The innermost mean value on the right-hand side of (89) satisfies due to the normal distribution of ν_i

$$E_{\nu_i}\left[\exp\left(\frac{\lambda}{k}[Du']_i\nu_i\right)\right] = \exp\left(\frac{\lambda^2}{2k^2}([Du']_i)^2\sigma_\nu^2\right) \quad (90)$$

and thus, replacing (90) in (89), we have

$$\begin{aligned} &\left(E\left[\exp\left(\frac{\lambda}{k}[Du']_i\nu_i\right)\Big|F'_{i-1}\right]\right)^k \\ &\quad= \left(E\left[\exp\left(\frac{\lambda^2}{2k^2}[Du']_i^2\sigma_\nu^2\right)\Big|F'_{i-1}\right]\right)^k. \end{aligned} \quad (91)$$

Combining (89)–(91) with $([Du']_i)^2 \le g_D(r)^2$, a direct consequence of (5), we have

$$\left(E\left[\exp\left(\frac{\lambda}{k}[Du']_i\nu_i\right)\Big|F'_{i-1}\right]\right)^k \le \exp\left(\frac{\lambda^2}{2k}g_D(r)^2\sigma_\nu^2\right). \quad (92)$$

Consider next the second term of the product on the right-hand side of (88). Application of inequality [24, Lemma 2] yields due to $E\left[\Delta''_i|F'_{i-1}\right] = 0$, (70) and (71)

$$\begin{aligned} &\left(E\left[\exp\left(\frac{\lambda}{1-k}\Delta''_i\right)\Big|F'_{i-1}\right]\right)^{1-k} \\ &\le \frac{(b')^2}{(b')^2+V'}\exp\left(-\frac{\lambda}{1-k}\frac{V'}{b'}\right) \\ &\quad+ \frac{V'}{(b')^2+V'}\exp\left(\frac{\lambda}{1-k}b'\right) \quad (93) \\ &= \frac{1}{2}\exp\left(-\frac{\lambda}{1-k}b'\right) + \frac{1}{2}\exp\left(\frac{\lambda}{1-k}b'\right). \quad (94) \end{aligned}$$

Relation (94) results from the direct application of (71) to (93). The proof is complete if we replace (92) and (94) in (88).

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor Prof. A. Lapidoth and the anonymous reviewers for their helpful comments which greatly enhanced the present work.

REFERENCES

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

[2] A. Goldsmith, *Wireless Communications*. New York: Cambridge Univ. Press, 2005.

[3] A. C. Singer, N. R. Shanbhag, and H. M. Bae, "Electronic dispersion compensation," *IEEE Signal. Process. Mag.*, vol. 25, no. 6, pp. 110–130, Nov. 2008.

[4] N. Kalouptsidis, *Signal Processing Systems: Theory and Design*. New York: Wiley, 1997.

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

[6] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Malabar, FL: Krieger, 2006.

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

[8] M. Maqusi, "Performance of baseband digital data transmission in nonlinear channels with memory," *IEEE Trans. Commun.*, vol. COM-33, pp. 715–719, 1985.

[9] T. L. Lim and J. K. Omura, "Error rates estimates in digital communication over a nonlinear channel with memory," *IEEE Trans. Commun.*, vol. COM-31, pp. 407–412, 1983.

[10] Y. G. Liu, I. Oka, and E. Biglieri, "Error probability for digital transmission over nonlinear channels with application to TCM," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1101–1110, Sep. 1990.

[11] R. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ: Wiley, 1968.

[12] 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, Sep. 1995.

[13] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer, 1998.

[14] I. Csiszar and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[15] A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[16] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[17] D. J. Forney, Jr., "Capacity of AWGN channels," M.I.T. Lecture Notes, 2005.

[18] C. B. Bell, "Mutual information and maximal correlation as measures of dependence," *Ann. Math. Statist.*, vol. 33, no. 2, pp. 587–595, 1962.

[19] C. McDiarmid, "Concentration," *Probab. Meth. Alg. Discrete Math.*, pp. 195–248, 1998.

[20] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Telatar, "A simple converse of Burnashev's reliability function," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3074–3080, Jul. 2009.

[21] M. V. Burnashev, "Data transmission over a discrete channel with feedback," *Probl. Inf. Transm.*, pp. 250–265, 1976.

[22] V. Bentkus, "On Hoeffding's inequalities," *Ann. Probab.*, vol. 32, no. 2, pp. 1650–1673, 2004.

[23] D. Divsalar and F. Pollara, Multiple turbo codes for deep-space communications JPL, 1995, TDA Prog. Rep. 42-121.

[24] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, Mar. 1963.

Kostis Xenoulis (S'08) received the B.Sc. degree in informatics and telecommunications and the M.Sc. degree in signal processing for telecommunications and multimedia from the University of Athens, Athens, Greece, in 2003 and 2005, respectively.

He is currently pursuing the Ph.D. degree in the Department of Informatics and Telecommunications, University of Athens, Greece. His research interests are in the area of information theory.

Nicholas Kalouptsidis (M'82–SM'85) received the B.Sc. degree in mathematics from the University of Athens, Athens, Greece, in 1973, and the Ph.D. degree in systems science and mathematics from Washington University, St. Louis, MO, in 1976.

He has held visiting positions with Washington University; Politecnico di Torino; Northeastern University, Boston, MA; CNET, Lannion, France; and University of Utah, Salt Lake City. In spring 2008, he was a visiting scholar with Harvard University, Cambridge, MA. He is currently a Professor with the Department of Informatics and Telecommunications, University of Athens. He is the author of the textbook *Signal Processing Systems: Theory and Design* (Wiley, 1997) and coeditor, with S. Theodoridis, of the book Adaptive System Identification and Signal Processing Algorithms (Englewood Cliffs, NJ: Prentice-Hall, 1993). His research interests are in system theory and signal processing.