

# Exponential Error Bounds on Parameter Modulation–Estimation for Memoryless Channels

Neri Merhav

Dep. of Electrical Engineering  
Technion – Israel Institute of Technology  
Technion City, Haifa 32000, Israel  
Email: merhav@ee.technion.ac.il

**Abstract**—We consider the problem of modulation and estimation of a random parameter  $U$  to be conveyed across a discrete memoryless channel. Upper and lower bounds are derived for the best achievable exponential decay rate of a general moment of the estimation error,  $E|\hat{U} - U|^\rho$ ,  $\rho \geq 0$ , when both the modulator and the estimator are subjected to optimization. These exponential error bounds turn out to be intimately related to error exponents of channel coding and to channel capacity. While in general, there is some gap between the upper and the lower bound, they asymptotically coincide both for very small and for very large values of the moment power  $\rho$ . This means that our achievability scheme, which is based on simple quantization of  $U$  followed by channel coding, is nearly optimum in both limits. Some additional properties of the bounds are discussed and demonstrated, and finally, an extension to the case of a multidimensional parameter vector is outlined, with the principal conclusion that our upper and lower bound asymptotically coincide also for a high dimensionality.

## I. INTRODUCTION

Consider the problem of conveying a parameter  $u$  across a discrete memoryless channel  $p(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^n p(y_t|x_t)$ , where  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{y} = (y_1, \dots, y_n)$  are the channel input and output vectors, respectively. Our interest is in the following question: How well can one estimate  $u$  based on  $\mathbf{y}$  when one is allowed to optimize both the estimator and the modulator, that is, the function  $\mathbf{x}(u) = (x_1(u), \dots, x_n(u))$  that maps  $u$  into a channel input vector?

In principle, this problem, which is the discrete-time analogue of the classical problem of “waveform communication” [9, Chap. 8], can be viewed both from the information-theoretic and the estimation-theoretic perspectives. Classical results in neither of these disciplines, however, seem to suggest satisfactory answers.

From the information-theoretic point of view, if the parameter is random, call it  $U$ , this is a problem of joint source–channel coding, where the source emits a single variable  $U$ , whereas the channel is allowed to be used many times ( $n$  large). The separation theorem of information theory asserts that asymptotic optimality of separate source– and channel coding is guaranteed in the limit of long blocks. However, it refers to a regime of long blocks both in source coding and channel coding, whereas here the source block length is 1, and so, there is no hope to compress the source with performance that comes close to the rate–distortion function.

In the realm of estimation theory, on the other hand, there is a rich literature on Bayesian and non-Bayesian bounds, mostly concerning the mean square error (MSE) in estimating parameters from signals corrupted by an additive white Gaussian noise (AWGN) channel, as well as other channels (see, e.g., [7] for an overview). Most of these bounds lend themselves to calculation for a *given* modulator  $\mathbf{x}(u)$  and therefore they may give insights concerning optimum estimation for this specific modulator. They may not, however, be easy to use for the derivation of *universal* lower bounds, namely, lower bounds that depend neither on the modulator nor on the estimator, which are relevant when both optimum modulators and optimum estimators are sought. Two exceptions to this rule are families of bounds that stem from generalized data processing theorems (DPT’s) [10], [12], henceforth referred to as “DPT bounds”, and bounds based on hypothesis testing and channel coding considerations [1], [11], henceforth called “channel–coding bounds.”

In this paper, we use both the channel–coding techniques and DPT techniques in order to derive lower bounds on general moments of the estimation error,  $E|\hat{U} - U|^\rho$ , where  $U$  is a random parameter,  $\hat{U}$  is its estimate, and the power  $\rho$  is an arbitrary positive real (not necessarily an integer). It turns out that when  $\mathbf{x}(u)$  is subjected to optimization,  $E|\hat{U} - U|^\rho$  can decay exponentially rapidly as a function of  $n$ , and so, our focus is on the best achievable exponential rate of decay as a function of  $\rho$ , which we shall denote by  $\mathcal{E}(\rho)$ , that is,  $\inf E|\hat{U} - U|^\rho \approx e^{-n\mathcal{E}(\rho)}$ , where the infimum is over all modulators and estimators.\* Interestingly, both the upper and lower bounds on  $\mathcal{E}(\rho)$  are intimately related to well-known exponential error bounds associated with channel coding, such as Gallager’s random coding exponent (for small values of  $\rho$ ) and the expurgated exponent function (for large values of  $\rho$ ). In other words, we establish an estimation-theoretic meaning to these error exponent functions. In particular, under certain conditions, our channel–coding upper bound on  $\mathcal{E}(\rho)$  (corresponding to a lower bound on  $E|\hat{U} - U|^\rho$ ) can be presented as

$$\bar{\mathcal{E}}(\rho) = \begin{cases} E_0(\rho) & \rho < \rho_0 \\ E_{ex}(0) & \rho \geq \rho_0 \end{cases} \quad (1)$$

\*This is still an informal and non-rigorous description. More precise definitions will be given in the sequel.

where  $E_0(\rho) = \max_q E_0(\rho, q)$ ,  $E_0(\rho, q)$  being Gallager's function,  $E_{ex}(0)$  is the expurgated exponent at zero rate, and  $\rho_0$  is value of  $\rho$  for which  $E_0(\rho) = E_{ex}(0)$  (so that  $\bar{E}(\rho)$  is continuous). In addition, we derive a DPT bound and discuss its advantages and disadvantages compared to the above bound.

We also suggest a lower bound,  $\underline{E}(\rho)$ , on  $\mathcal{E}(\rho)$ , which is achieved by a simple, separation-based modulation and estimation scheme (see also [3, Section III.A]). While there is a certain gap between  $\bar{E}(\rho)$  and  $\underline{E}(\rho)$  for every finite  $\rho$ , it turns out that this gap disappears (in the sense that  $\underline{E}(\rho)/\bar{E}(\rho) \rightarrow 1$  both for large  $\rho$  and for small  $\rho$ , and so, we have exact asymptotics of  $\mathcal{E}(\rho)$  in these two extremes: For large  $\rho$ ,  $\mathcal{E}(\rho)$  tends to  $E_{ex}(0)$  and for small  $\rho$ ,  $\mathcal{E}(\rho) \sim \rho C$ , where  $C$  is the channel capacity. Our simple achievability scheme is then nearly optimum at both extremes, which means that a separation theorem essentially holds for small and for large values of  $\rho$ , in spite of the earlier discussion (see also [5, Section III.D]). The results are demonstrated for the example of a "very noisy channel," [2, Example 3, pp. 147–149] which is convenient to analyze, as it admits closed-form expressions.

Finally, we suggest an extension of our results to the case of a multidimensional parameter vector  $\mathbf{U} = (U_1, \dots, U_d)$ . It turns out that the effect of the dimension  $d$  is in reducing the effective value of  $\rho$  by a factor of  $d$ . In other words,  $\bar{E}(\rho)$  is replaced by  $\bar{E}(\rho/d)$  and the extension of the achievability result is straightforward. This means that for fixed  $\rho$ , the limit of large  $d$  (where the effective value  $\rho/d$  is very small) also admits exact asymptotics, where  $\mathcal{E}(\rho) \sim \rho C/d$ .

## II. NOTATION CONVENTIONS AND FORMALIZATION

Throughout this paper, scalar random variables (RV's) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and their sample values which will be denoted with same symbols in a bold face font. For example,  $y \in \mathcal{Y}$  is a realization of a random variable  $Y$ , whereas  $\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{Y}^n$  ( $n$  being a positive integer and  $\mathcal{Y}^n$  being the  $n$ -th Cartesian power of  $\mathcal{Y}$ ) is a realization of a random vector  $\mathbf{Y} = (Y_1, \dots, Y_n)$ . Let  $U$  be a uniformly distributed<sup>†</sup> random variable over the interval  $[-1/2, +1/2]$ , which we will also denote by  $\mathcal{U}$ . We refer to  $U$  as the parameter to be conveyed from the source to the destination, via a given noisy channel. A given realization of  $U$  will be denoted by  $u$ .

A discrete memoryless channel (DMC) is characterized by a matrix of conditional probabilities  $p = \{p(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ , where the channel input and output alphabets,  $\mathcal{X}$  and  $\mathcal{Y}$ , are assumed finite.<sup>‡</sup> When a DMC  $p = \{p(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$  is fed by an input vector  $\mathbf{x} \in \mathcal{X}^n$ , it produces an

output vector  $\mathbf{y} \in \mathcal{Y}^n$  according to  $p(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^n p(y_t|x_t)$ . A modulator is a measurable mapping  $\mathbf{x} = f_n(u)$  from  $\mathcal{U} = [-1/2, +1/2]$  to  $\mathcal{X}^n$  and an estimator is a mapping  $\hat{u} = g_n(\mathbf{y})$  from  $\mathcal{Y}^n$  back to  $\mathcal{U}$ . The random vector  $f_n(U)$  will also be denoted by  $\mathbf{X}$ . Similarly, the random variable  $g_n(\mathbf{Y})$  will also be denoted by  $\hat{U}$ . Our figure of merit for communication systems is the expectation of  $\rho$ -th power of the estimation error, i.e.,  $\mathbf{E}\{|\hat{U} - U|^\rho\}$ , where  $\rho$  is a positive real (not necessarily an integer). The capability of attaining an exponential decay in  $\mathbf{E}\{|\hat{U} - U|^\rho\}$  by certain choices of a modulator  $f_n$  and an estimator  $g_n$ , motivates the definition of the following exponential rates

$$\bar{\mathcal{E}}(\rho) = \limsup_{n \rightarrow \infty} \left[ -\frac{1}{n} \ln \left( \inf_{f_n, g_n} \mathbf{E}\{|\hat{U} - U|^\rho\} \right) \right] \quad (2)$$

and

$$\underline{\mathcal{E}}(\rho) = \liminf_{n \rightarrow \infty} \left[ -\frac{1}{n} \ln \left( \inf_{f_n, g_n} \mathbf{E}\{|\hat{U} - U|^\rho\} \right) \right]. \quad (3)$$

This paper is about upper bounds on  $\bar{\mathcal{E}}(\rho)$  and lower bounds on  $\underline{\mathcal{E}}(\rho)$ , with special interest in situations where these bounds come close.

## III. UPPER AND LOWER BOUNDS BASED ON CHANNEL CODING

Let  $q = \{q(x), x \in \mathcal{X}\}$  be a given probability vector of a random variable  $X$  taking on values in  $\mathcal{X}$ , and let  $p = \{p(y|x), \mathcal{X}, y \in \mathcal{Y}\}$  define the given DMC. Let  $E_0(\rho, q)$  be the Gallager function [2, p. 138, eq. (5.6.14)], [8, p. 133, eq. (3.1.18)], defined as

$$E_0(\rho, q) = -\ln \left( \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} q(x) p(y|x)^{1/(1+\rho)} \right]^{1+\rho} \right), \quad (4)$$

for  $\rho \geq 0$ . Next, define  $E_0(\rho) = \max_q E_0(\rho, q)$ , where the maximum is over the simplex of probability vectors, and let  $\bar{E}_0(\rho)$  be the upper concave envelope<sup>§</sup> (UCE) of  $E_0(\rho)$ . Next define

$$E_x(\varrho) = -\varrho \ln \sum_{x, x' \in \mathcal{X}} q(x) q(x') \left[ \sum_{y \in \mathcal{Y}} \sqrt{\varrho p(y|x) \varrho p(y|x')} \right]^{1/\varrho}$$

where the parameter  $\varrho$  should be distinguished from the power  $\rho$  of the estimation error in discussion. The expurgated exponent function [2, p. 153, eq. (5.7.11)], [8, p. 146, eq. (3.3.13)] is defined as  $E_{ex}(R) = \sup_{\varrho \geq 1} [E_x(\varrho) - \varrho R]$ . It is well known (and a straightforward exercise to show) that

$$\begin{aligned} E_{ex}(0) &= \sup_{\varrho \geq 1} E_x(\varrho) = \lim_{\varrho \rightarrow \infty} E_x(\varrho) \\ &= - \sum_{x, x' \in \mathcal{X}} q(x) q(x') \ln \left[ \sum_{y \in \mathcal{Y}} \sqrt{p(y|x) p(y|x')} \right] \end{aligned} \quad (5)$$

<sup>§</sup>While the Gallager function  $E_0(\rho, q)$  is known to be concave in  $\rho$  for every fixed  $q$  [8, p. 134, eq. (3.2.5a)], we are not aware of an argument asserting that  $E_0(\rho)$  is concave in general. On the other hand, there are many situations where  $E_0(\rho)$  is, in fact, concave and then  $\bar{E}_0(\rho) = E_0(\rho)$ , for example, when the achiever  $q^*$  of  $\max_q E_0(\rho, q)$  is independent of  $\rho$ , like the case of the binary input output-symmetric (BIOS) channel [8, p. 153].

<sup>†</sup>This specific assumption concerning the density of  $U$  and its support is made for convenience only. Our results extend to more general densities.

<sup>‡</sup>The finite alphabet assumption is used mainly for reasons of simplicity. The extension to continuous alphabets is possible, though some caution should be exercised at several places.

Finally, define

$$\bar{E}(\rho) = \begin{cases} \bar{E}_0(\rho) & \rho \leq \rho_0 \\ E_{ex}(0) & \rho > \rho_0 \end{cases} \quad (6)$$

where  $\rho_0$  is the (unique) solution to the equation  $\bar{E}_0(\rho) = E_{ex}(0)$ .

Our first theorem (see [6, Appendix A] for the proof) asserts that  $\bar{E}(\rho)$  is an upper bound on the best achievable exponential decay rate of  $\rho$ -th moment of the estimation error.

*Theorem 1:* Let  $U$  be uniformly distributed over  $\mathcal{U} = [-1/2, +1/2]$  and let  $p = \{p(y|x) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$  be a given DMC. Then, for every  $\rho \geq 0$ ,  $\bar{\mathcal{E}}(\rho) \leq \bar{E}(\rho)$ .

We now proceed to present a lower bound  $\underline{E}(\rho)$  to  $\underline{\mathcal{E}}(\rho)$ . Let  $R_-$  be the smallest  $R$  such that  $E_{ex}(R)$  is attained with  $\varrho = 1$  and let  $R_+$  denote the largest  $R$  such that

$$E_r(R) = \max_{0 \leq \varrho \leq 1} [E_0(\varrho, q) - \varrho R] \quad (7)$$

is attained for  $\varrho = 1$ .<sup>¶</sup> Next, define  $\rho_+ = [E_0(1) - R_+]/R_+$  and  $\rho_- = [E_0(1) - R_-]/R_-$ , and finally,

$$\underline{E}(\rho) = \begin{cases} \sup_{0 \leq \varrho \leq 1} \rho E_0(\varrho)/(\varrho + \rho) & \rho \leq \rho_+ \\ \rho E_0(1)/(1 + \rho) = \rho E_x(1)/(1 + \rho) & \rho_+ < \rho \leq \rho_- \\ \sup_{\varrho \geq 1} \rho E_x(\varrho)/(\varrho + \rho) & \rho > \rho_- \end{cases}$$

Our next theorem (see [6, Appendix B] for the proof) tells us that  $\underline{E}(\rho)$  is a lower bound on the best attainable exponential decay rate of  $\mathbf{E}\{| \hat{U} - U |^\rho\}$ .

*Theorem 2:* Let  $U$  be uniformly distributed over  $\mathcal{U} = [-1/2, +1/2]$  and let  $p = \{p(y|x) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$  be a given DMC. Then, for every  $\rho \geq 0$ ,  $\underline{\mathcal{E}}(\rho) \geq \underline{E}(\rho)$ .

The derivations of both  $\bar{E}(\rho)$  and  $\underline{E}(\rho)$  rely on channel coding considerations. In particular, the derivation of  $\bar{E}(\rho)$  builds strongly on the method of [5], which extends the derivation of the Ziv–Zakai bound [11] and the Chazan–Zakai–Ziv bound [1]. While the two latter bounds are based on considerations associated with binary hypotheses testing, here and in [5], the general idea is extended to exponentially many hypotheses pertaining to channel decoding.

We see that both bounds exhibit different types of behavior in different ranges of  $\rho$  (i.e., “phase transitions”), but in a different manner. For both  $\bar{E}(\rho)$  and  $\underline{E}(\rho)$  the behavior is related to the ordinary Gallager function in some range of small  $\rho$ , and to the expurgated exponent in a certain range of large  $\rho$ .

The communication system that achieves  $\underline{E}(\rho)$  works as follows (see also [5]: Define  $R(\rho) = \underline{E}(\rho)/\rho$  and construct a uniform grid of  $M = e^{nR(\rho)}/2$  evenly spaced points along  $\mathcal{U}$ , denoted  $\{u_1, u_2, \dots, u_M\}$ . If  $\rho > \rho_-$  assign to each grid point  $u_i$  a codeword of a code of rate  $R(\rho)$  that achieves the expurgated exponent  $E_{ex}[R(\rho)]$  (see [2, Theorem 5.7.1] or [8, Theorem 3.3.1]). If  $\rho \leq \rho_-$ , do the same with a code that achieves  $E_r[R(\rho)]$  (see [2, p. 139, Corollary 1] or [8, Theorem

3.2.1]). Given  $u$ , let  $f_n(u)$  be the codeword  $x_i$  that is assigned to the grid point  $u_i$ , which is closest to  $u$ . Given  $\mathbf{y}$ , let  $g_n(\mathbf{y})$  be the grid point  $u_j$  that corresponds to the codeword  $x_j$  that has been decoded based on  $\mathbf{y}$  using the ML decoder for the given DMC.

Let us examine the behavior of these bounds as  $\rho \rightarrow 0$  and as  $\rho \rightarrow \infty$ . For very large values of  $\rho$ , where the upper bound  $\bar{E}(\rho)$  is obviously given by  $E_{ex}(0)$ , the lower bound is given by

$$\lim_{\rho \rightarrow \infty} \underline{E}(\rho) = \lim_{\rho \rightarrow \infty} \sup_{\varrho \geq 1} \frac{\rho E_x(\varrho)}{\varrho + \rho} \quad (8)$$

$$\geq \lim_{\rho \rightarrow \infty} \frac{\rho E_x(\sqrt{\rho})}{\sqrt{\rho} + \rho} \quad (9)$$

$$= \lim_{\rho \rightarrow \infty} E_x(\sqrt{\rho}) = E_{ex}(0), \quad (10)$$

which means that for large  $\rho$  all the exponents asymptotically coincide:

$$\lim_{\rho \rightarrow \infty} \underline{E}(\rho) = \lim_{\rho \rightarrow \infty} \underline{\mathcal{E}}(\rho) = \lim_{\rho \rightarrow \infty} \bar{\mathcal{E}}(\rho) = \lim_{\rho \rightarrow \infty} \bar{E}(\rho) = E_{ex}(0).$$

In the achievability scheme described above,  $R(\rho)$  is a very low coding rate. On the other hand, for very small values of  $\rho$ , where  $\bar{E}(\rho) = \bar{E}_0(\rho) = \rho C + o(\rho)$ ,  $C$  being the channel capacity, we have

$$\lim_{\rho \rightarrow 0} \frac{\underline{E}(\rho)}{\rho} = \lim_{\rho \rightarrow 0} \sup_{0 \leq \varrho \leq 1} \frac{E_0(\varrho)}{\varrho + \rho} \quad (11)$$

$$\geq \lim_{\rho \rightarrow 0} \frac{E_0(\sqrt{\rho})}{\sqrt{\rho} + \rho} \quad (12)$$

$$= \lim_{\rho \rightarrow 0} \frac{E_0(\sqrt{\rho})}{\sqrt{\rho}} \cdot \frac{1}{1 + \sqrt{\rho}} \quad (13)$$

$$= \lim_{\rho \rightarrow 0} \frac{E_0(\sqrt{\rho})}{\sqrt{\rho}} = C, \quad (14)$$

which means that for small  $\rho$  all the exponents behave like  $\rho C$ , i.e.,

$$\lim_{\rho \rightarrow 0} \frac{\underline{E}(\rho)}{\rho} = \lim_{\rho \rightarrow 0} \frac{\underline{\mathcal{E}}(\rho)}{\rho} = \lim_{\rho \rightarrow 0} \frac{\bar{\mathcal{E}}(\rho)}{\rho} = \lim_{\rho \rightarrow 0} \frac{\bar{E}(\rho)}{\rho} = C. \quad (15)$$

It is then interesting to observe that not only channel-coding error exponents, but also channel capacity plays a role in the characterization of the best achievable modulation–estimation performance. In the achievability scheme described above,  $R(\rho)$  is a very high coding rate, very close to the capacity  $C$ .

#### IV. UPPER BOUND BASED ON DATA PROCESSING INEQUALITIES

We next derive an alternative upper bound on  $\bar{\mathcal{E}}(\rho)$  that is based on generalized data processing inequalities, following Ziv and Zakai [12] and Zakai and Ziv [10]. The idea behind these works is that it is possible to define generalized mutual information functionals satisfying a DPT, by replacing the negative logarithm function of the ordinary mutual information, by a general convex function. This enables to obtain tighter

<sup>¶</sup>For example, in the case of the BSC with a crossover parameter  $p$ ,  $R_- = \ln 2 - h_2(Z/(1+Z))$ , with  $Z = \sqrt{4p(1-p)}$ , and  $R_+ = \ln 2 - h_2(\sqrt{p}/(\sqrt{p} + \sqrt{1-p}))$ , where  $h_2(x) = -x \ln x - (1-x) \ln(1-x)$  [8, pp. 151–152].

distortion bounds for communication systems with short block length.

In [4] it was shown that the following generalized mutual information functional, between two generic random variables,  $A$  and  $B$ , admits a DPT for every positive integer  $k$  and for every vector  $(\alpha_1, \dots, \alpha_k)$  whose components are non-negative and sum to unity:

$$\tilde{I}(A; B) = - \sum_{b \in \mathcal{B}} \prod_{i=1}^k \sum_{a_i \in \mathcal{A}} q(a_i) p(b|a_i)^{\alpha_i}. \quad (16)$$

In particular, since  $U \rightarrow \mathbf{Y} \rightarrow \hat{U}$  is a Markov chain, then by the generalized DPT,  $\tilde{I}(U; \hat{U}) \leq \tilde{I}(U; \mathbf{Y})$ . The idea is to further upper bound  $\tilde{I}(U; \mathbf{Y})$  and to further lower bound  $\tilde{I}(U; \hat{U})$  subject to the constraint  $\mathbf{E}|\hat{U} - U|^\rho = D$ , which leads to a generalized rate-distortion function, and thereby to obtain an inequality on  $\mathbf{E}|\hat{U} - U|^\rho$ . Specifically,  $\tilde{I}(U; \mathbf{Y})$  is upper bounded as follows:

$$\begin{aligned} \tilde{I}(U; \mathbf{Y}) &= - \sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^k \int_{-1/2}^{+1/2} du_i p(\mathbf{y} | f_n(u_i))^{\alpha_i} \\ &= - \sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^k \int_{-1/2}^{+1/2} du_i \prod_{t=1}^n p(y_t | [f_n(u_i)]_t)^{\alpha_i} \\ &= - \prod_{t=1}^n \sum_{y_t \in \mathcal{Y}} \prod_{i=1}^k \int_{-1/2}^{+1/2} du_i p(y_t | [f_n(u_i)]_t)^{\alpha_i} \\ &\leq - \prod_{t=1}^n \min_q \sum_{y_t \in \mathcal{Y}} \prod_{i=1}^k \sum_{x_i \in \mathcal{X}} q(x_i) p(y_t | x_i)^{\alpha_i} \\ &= - \left[ \min_q \sum_{y \in \mathcal{Y}} \prod_{i=1}^k \sum_{x_i \in \mathcal{X}} q(x_i) p(y | x_i)^{\alpha_i} \right]^n \\ &= - \exp\{-n \max_q E(\alpha_1, \dots, \alpha_k, q)\}, \end{aligned} \quad (17)$$

where  $[f_n(u_i)]_t$  denotes the  $t$ -th component of the vector  $\mathbf{x} = f_n(u_i)$  and where

$$E(\alpha_1, \dots, \alpha_k, q) = - \ln \left[ \sum_{y \in \mathcal{Y}} \prod_{i=1}^k \left( \sum_{x_i \in \mathcal{X}} q(x_i) p(y | x_i)^{\alpha_i} \right) \right].$$

Note that for  $k = 1 + \varrho$  ( $\varrho$  - integer),

$$E\left(\frac{1}{1+\varrho}, \dots, \frac{1}{1+\varrho}, q\right) = E_0(\varrho, q). \quad (18)$$

In [6, Appendix C], we show that

$$\min\{\tilde{I}(U, \hat{U}) : \mathbf{E}|\hat{U} - U|^\rho = D\} \triangleq \tilde{R}(D) \geq -c \cdot D \sum_{i=1}^k \zeta_\rho(\alpha_i) \quad (19)$$

where  $c$  is a constant that depends solely on  $\rho$ ,  $k$  and  $\alpha_1, \dots, \alpha_k$ , and where

$$\zeta_\rho(\alpha) = \min \left\{ \alpha, \frac{1-\alpha}{\rho} \right\}. \quad (20)$$

The function  $\tilde{R}(D)$  in eq. (19) is referred to as a “generalized rate-distortion function” in the terminology of [12] and [10].

Thus, from the generalized DPT,

$$\mathbf{E}|\hat{U} - U|^\rho \equiv D \geq c' \cdot e^{-n \bar{E}_{DPT}(\rho)} \quad (21)$$

where  $c'$  is another constant and

$$\bar{E}_{DPT}(\rho) \triangleq \inf_{k \geq 1} \inf_{\alpha_1, \dots, \alpha_k} \sup_q \frac{E(\alpha_1, \dots, \alpha_k, q)}{\sum_{i=1}^k \zeta_\rho(\alpha_i)}. \quad (22)$$

As an example, assume that the channel is such that the function  $E_0(\varrho)$  is concave, so that  $E_0(\varrho) = E_0(\varrho)$ . In this case,  $\rho_0 \geq 1$  since  $E_0(1) \leq E_{ex}(0)$  and  $E_0(\varrho)$  is monotonically increasing. Now, let  $\rho \leq \rho_0$  be an integer (for example,  $\rho = 1$  is always a legitimate choice). Then,

$$\bar{E}(\rho) = E_0(\rho) \quad (23)$$

$$= \sup_q \frac{E(1/(1+\rho), \dots, 1/(1+\rho), q)}{(1+\rho)\zeta_\rho(1/(1+\rho))} \quad (24)$$

$$\geq \inf_{k \geq 1} \inf_{\alpha_1, \dots, \alpha_k} \sup_q \frac{E(\alpha_1, \dots, \alpha_k, q)}{\sum_{i=1}^k \zeta_\rho(\alpha_i)} \quad (25)$$

$$= \bar{E}_{DPT}(\rho). \quad (26)$$

Thus, at least in this case, the DPT bound is guaranteed to be no worse than the channel-coding bound  $\bar{E}(\rho)$ . Nonetheless, in our numerical studies, we have not found examples where the DPT bound strictly improves on the channel-coding bound, i.e.,  $\bar{E}_{DPT}(\rho) < \bar{E}(\rho)$ , and it remains an open question whether the DPT bound can offer improvement in any situation, thanks to its additional degrees of freedom. It should be pointed out that the achiever  $(\alpha_1, \dots, \alpha_k)$  of  $E_{DPT}(\rho)$  is not always given by  $(1/(k+1), \dots, 1/(k+1))$  because the function  $E(\alpha_1, \dots, \alpha_k, q)$  is not convex in  $(\alpha_1, \dots, \alpha_k)$ . At any rate, in all cases where  $\bar{E}_{DPT}(\rho) = \bar{E}(\rho)$ , this is interesting on its own right since the two bounds are based on completely different considerations. One advantage of the DPT approach is that it seems to lend itself more comfortably to extensions that account for moments of more general functions of the estimation error, i.e.,  $\mathbf{E}\{g(|\hat{U} - U|)\}$ , for a large class of monotonically increasing functions  $g$ . On the other hand, the optimization associated with calculation of the DPT bound is not trivial.

## V. EXAMPLE: VERY NOISY CHANNEL

As an example, consider the *very noisy channel*, which is characterized by

$$p(y|x) = p(y)[1 + \epsilon(x, y)], \quad |\epsilon(x, y)| \ll 1 \quad (27)$$

for every  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ . As is shown in [8, Sect. pp. 155–158], to the first order, we have the following:

$$C = \frac{1}{2} \max_q \sum_{x, y} q(x) p(y) \epsilon^2(x, y) \quad (28)$$

$$E_0(\varrho) = \frac{\varrho}{1+\varrho} \cdot C, \quad (29)$$

and therefore

$$\begin{aligned} E_r(R) &= \max_{0 \leq \varrho \leq 1} \left( \frac{\varrho}{1+\varrho} \cdot C - \varrho R \right) \\ &= \begin{cases} \frac{C}{2} - R & R < \frac{C}{4} \\ (\sqrt{C} - \sqrt{R})^2 & \frac{C}{4} \leq R \leq C \\ 0 & R > C \end{cases} \quad (30) \end{aligned}$$

As for the expurgated exponent, we have  $E_x(\varrho) = E_0(1) = \frac{C}{2}$  and so,  $E_{ex}(R) = \sup_{\varrho \geq 1} [E_x(\varrho) - \varrho R] = \frac{C}{2} - R$ , which means that expurgation does not help for very noisy channels. This implies that  $\rho_0 = 1$  and so

$$\bar{E}(\rho) = \begin{cases} \frac{\rho}{1+\rho} \cdot C & \rho \leq 1 \\ \frac{C}{2} & \rho > 1 \end{cases} \quad (31)$$

As for the lower bound, we have the following: For  $\rho < 1$ ,

$$\underline{E}(\rho) = \sup_{0 \leq \varrho \leq 1} \frac{\rho}{\rho + \varrho} \cdot \frac{\varrho}{1 + \varrho} \cdot C = \frac{\rho}{(1 + \sqrt{\rho})^2} \cdot C. \quad (32)$$

For  $\rho \geq 1$ ,

$$\underline{E}(\rho) = \sup_{\varrho \geq 1} \frac{\rho E_x(\varrho)}{\varrho + \rho} = \sup_{\varrho \geq 1} \frac{\rho}{\varrho + \rho} \cdot \frac{C}{2} = \frac{\rho}{1 + \rho} \cdot \frac{C}{2}. \quad (33)$$

Thus, in summary

$$\underline{E}(\rho) = \begin{cases} \frac{\rho}{(1 + \sqrt{\rho})^2} \cdot C & \rho < 1 \\ \frac{\rho}{1 + \rho} \cdot \frac{C}{2} & \rho \geq 1 \end{cases}$$

We see how the bounds asymptotically coincide (in the sense that  $\bar{E}(\rho)/\underline{E}(\rho) \approx 1$ ) both for very large values of  $\rho$  and for very small values of  $\rho$ .

As for the DPT bound, using the approximations (27) and (28), we have the following (see [6] for details):

$$\sup_q E(\alpha_1, \dots, \alpha_k, q) \approx C \left( 1 - \sum_{i=1}^k \alpha_i^2 \right), \quad (34)$$

and then

$$\bar{E}_{DPT}(\rho) \approx C \cdot \inf_{k > 1} \inf_{\alpha_1, \dots, \alpha_k} \frac{1 - \sum_{i=1}^k \alpha_i^2}{\sum_{i=1}^k \zeta_\rho(\alpha_i)}. \quad (35)$$

For  $\rho = 1$  and  $k = 2$ , we have  $\zeta_1(\alpha) = \min\{\alpha, 1 - \alpha\}$ :

$$\bar{E}_{DPT}(1) \leq C \cdot \inf_{0 \leq \alpha \leq 1} \frac{1 - \alpha^2 - (1 - \alpha)^2}{2 \min\{\alpha, 1 - \alpha\}} = \frac{C}{2}, \quad (36)$$

which agrees with  $\bar{E}(\rho)$ . For  $\rho = 2$  and  $k = 2$ , the minimum is attained for  $\alpha = 1/3$ , and the result is  $\bar{E}_{DPT}(2) \leq 8C/9$ . However for  $k = 3$ , the bound improves to  $C/3$ .

## VI. EXTENSION TO THE MULTIDIMENSIONAL CASE

Consider now the case of a parameter vector  $\mathbf{U} = (U_1, \dots, U_d)$ , uniformly distributed across the unit hypercube  $[-1, 2, +1/2]^d$ . A reasonable figure of merit in this case would be a linear combination of  $\mathbf{E}\{|\hat{U}_i - U_i|^\rho\}$ ,  $i = 1, 2, \dots, d$ . Since each one of these terms is exponential in  $n$ , it makes sense to let the coefficients of this linear combination also be exponential functions of  $n$ , as otherwise, the results will be exponentially insensitive to the choice of the coefficients. This

means that we consider the criterion  $\sum_{i=1}^d e^{nr_i} \cdot \mathbf{E}\{|\hat{U}_i - U_i|^\rho\}$ , where, without loss of generality, we take  $r_i \geq 0$ ,  $\min_i r_i = 0$ .

As we show in [6], the extension of the channel-coding bound to the  $d$ -dimensional case reads

$$\bar{E}(\rho, d, r_1, \dots, r_d) = \begin{cases} E_0\left(\frac{\rho}{d}\right) - \frac{1}{d} \sum_{i=1}^d r_i & \rho/d \leq \rho_0 \\ E_{ex}(0) - \frac{\rho_0}{\rho} \sum_{i=1}^d r_i & \rho/d > \rho_0 \end{cases}$$

We see that when  $r_i = 0$  for all  $i$  (i.e., all weights are 1), it is the same channel-coding bound as before, except that  $\rho$  is replaced by  $\rho/d$ , that is,  $\bar{E}(\rho/d)$ . For  $\rho \rightarrow \infty$ , the bound tends to  $E_{ex}(0)$ , which can be approached again by a low-rate code for a Cartesian grid in the parameter space. At the other extreme, when  $d$  is very large compared to  $\rho$ , so  $\rho/d$  is small, construct a grid of  $e^{n(C-\epsilon)/d} \times e^{n(C-\epsilon)/d} \times \dots \times e^{n(C-\epsilon)/d}$ , quantize  $\mathbf{U}$  and assign to each grid point a codeword of a typical random code at rate  $C - \epsilon$ . Then the performance will be about  $e^{-n\rho C/d}$ . Therefore, as a corollary of the above result, we have

$$\sum_{i=1}^d \mathbf{E}\{|\hat{U}_i - U_i|^\rho\} \geq e^{-n[\bar{E}(\rho/d) + o(n)]}. \quad (37)$$

## REFERENCES

- [1] D. Chazan, M. Zakai, and J. Ziv, "Improved lower bounds on signal parameter estimation," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 1, pp. 90–93, January 1975.
- [2] R. G. Gallager, *Information Theory and Reliable Communication*, New York, Wiley 1968.
- [3] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1412–1424, September 1997.
- [4] N. Merhav, "Data processing inequalities based on a certain structured class of information measures with application to estimation theory," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 5287–5301, August 2012.
- [5] N. Merhav, "On optimum parameter modulation–estimation from a large deviations perspective," *IEEE Trans. Inform. Theory*, vol. 58, no. 12, pp. 7215–7225, December 2012.
- [6] N. Merhav, "Exponential error bounds on parameter modulation–estimation for discrete memoryless channels," submitted to *IEEE Trans. Inform. Theory*, December 2012. Available on-line: [http://arxiv.org/pdf/1212.4649.pdf].
- [7] H. L. Van Trees and K. L. Bell (Eds), *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, IEEE Press, Published by John Wiley & Sons, 2007.
- [8] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw–Hill, New York, 1979.
- [9] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, 1965. Reissued by Waveland Press, 1990.
- [10] M. Zakai and J. Ziv, "A generalization of the rate-distortion theory and applications," in: *Information Theory New Trends and Open Problems*, edited by G. Longo, Springer-Verlag, 1975, pp. 87–123.
- [11] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inform. Theory*, vol. IT-15, no. 3, pp. 386–391, May 1969.
- [12] J. Ziv and M. Zakai, "On functionals satisfying a data-processing theorem," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 3, pp. 275–283, May 1973.