Mathematics and Engineering Communications Laboratory

Technical Report

A Generalized Poor-Verdú Error Bound for Multihypothesis Testing and the Channel Reliability Function

P.-N. Chen and *F. Alajaji*

July 2011

A Generalized Poor-Verdú Error Bound for Multihypothesis Testing and the Channel Reliability Function

Po-Ning Chen and Fady Alajaji

Abstract

A lower bound on the minimum error probability for multihypothesis testing is established. The bound, which is expressed in terms of the cumulative distribution function of the *tilted* posterior hypothesis distribution given the observation with tilting parameter $\theta \geq 1$, generalizes an earlier bound due the Poor and Verdú (1995). A sufficient condition is established under which the new bound (minus a multiplicative factor) provides the exact error probability in the limit of θ going to infinity. Examples illustrating the new bound are also provided.

The application of this generalized Poor-Verdú bound to the channel reliability function is next carried out, resulting in two information-spectrum upper bounds. It is observed that, for a class of channels including the finite-input memoryless Gaussian channel, one of the bounds is tight and gives a multi-letter asymptotic expression for the reliability function, albeit its determination or calculation in single-letter form remains an open challenging problem. Numerical examples regarding the other bound are finally presented.

Index Terms

Hypothesis testing, probability of error, maximum-a-posteriori and maximum likelihood estimation, channel coding, channel reliability function, error exponent, binary-input additive white Gaussian noise channel.

P. N. Chen is with the Department of Electrical Engineering (Institute of Communications Engineering), National Chiao Tung University, Taiwan, ROC (email: poning@faculty.nctu.edu.tw). F. Alajaji is with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada (email: fady@mast.queensu.ca).

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and the National Science Council of Taiwan, R.O.C.

I. INTRODUCTION

In [12], Poor and Verdú establish a lower bound to the minimum error probability of multihypothesis testing. Specifically, given two random variables X and Y with joint distribution $P_{X,Y}$, X taking values in a finite or countably-infinite alphabet $\mathcal{X}$ and Y taking values in an arbitrary alphabet $\mathcal{Y}$, they show that the optimal maximum-a-posteriori (MAP) estimation of X given Y results in the following lower bound on the probability of estimation error P_e:

$$P_e \geq (1-\alpha) P_{X,Y} \left\{ (x,y) \in \mathcal{X} \times \mathcal{Y} : \; P_{X|Y}(x|y) \leq \alpha \right\} \tag{1}$$

for each $\alpha \in [0,1]$, where $P_{X|Y}$ denotes the posterior distribution of X given Y and the prior distribution P_X is arbitrary (not necessarily uniform). This bound has pertinent information-theoretic applications such as in the proof of the converse part of the channel coding theorem that yield formulas for both capacity and ε-capacity for general channels with memory (not necessarily information stable, stationary, etc) [14], [12]. It also improves upon previous lower bounds due to Shannon [13], [12, Eq. (7)] and to Verdú and Han [14], [12, Eq. (9)].

Furthermore, Poor and Verdú use the above bound to establish an information-spectrum based upper bound to the reliability function $E^*(R)$ – i.e., the optimal error exponent or the largest rate of asymptotic exponential decay of the error probability of channel codes [9], [5], [8], [15]– of general channels [12, Eq. (14)]. They conjecture that this bound, which is expressed in terms of a large-deviation rate function for the normalized channel information density (see Section IV-A for the definition), is tight (i.e., exactly equal to $E^*(R)$) for all rates R. In [1], it is however shown via a counterexample involving the memoryless binary erasure channel (BEC) that the bound is not tight at low rates, and a slightly tighter bound is presented [1, Corollary 1].

In this work, we generalize the above Poor-Verdú lower bound in (1) for the minimum error probability of multihypothesis testing. The new bound is expressed in terms of the cdf of the tilted posterior distribution of X given Y with tilting parameter $\theta \geq 1$, and it reduces to (1) when $\theta = 1$; see Theorem 1. We also provide a sufficient condition under which our generalized Poor-Verdú bound, without the multiplicative factor $(1-\alpha)$, is exact in the limit of θ going to infinity. Specifically, the sufficient condition requires having a unique MAP estimate of X from Y almost surely in P_Y, where P_Y is the distribution of Y; see Theorem 2. We present a few examples to illustrate the results of Theorems 1 and 2.

We proceed by applying the above results to the reliability function $E^*(R)$ of general channels. We employ Theorem 1 to establish two information-spectrum upper bounds to $E^*(R)$; see Theorem 3. One upper bound, $E_{\text{PV}}^{(\theta)}(R)$, is a function of the tilting parameter θ, while the other bound, $\bar{E}_{\text{PV}}(R)$, involves taking the limit infimum of θ. It turns out that if the channel satisfies a symmetry condition, then both upper bounds can be expressed in terms of the information density of an auxiliary channel whose transition distribution is nothing but the tilted distribution of the original channel distribution; see Observation 4.

We next use Theorem 2 to show that for the memoryless finite-input additive white Gaussian noise (AWGN) channel, the upper bound $\bar{E}_{\rm PV}(R)$ is tight, hence yielding an information-spectral formula for this channel's reliability function: $E^*(R) = \bar{E}_{\rm PV}(R)$ for all rates R between 0 and channel capacity; see Theorem 4. The calculation or determination in closed (single-letter) form of $\bar{E}_{\rm PV}(R)$ is however a formidable task and remains a notoriously open problem, as it requires solving the optimization of a large-deviation rate function in additions to two limiting operations; this makes it quite difficult to compare $\bar{E}_{\rm PV}(R)$ to well-known lower/upper bounds to $E^*(R)$ (such as the random coding lower bound and the sphere packing upper bound [9], [5][1]) for this AWGN channel. Nevertheless, the above multi-letter asymptotic expression for $E^*(R)$ may be conceptually useful for the future determination of $E^*(R)$ in computable single-letter form at low rates.[2] We also note that the equality $E^*(R) = \bar{E}_{\rm PV}(R)$ holds for a class of channels satisfying the sufficient condition of Theorem 2; see Corollary 1 and Observation 7.

Finally, we provide a lower bound to $E_{\rm PV}^{(\theta)}(R)$ for the case of memoryless channels, which is computable for a given value of θ. We use this lower bound to demonstrate numerically that for the memoryless BSC, $E_{\rm PV}^{(\theta)}(R)$ is not tight at all rates when $\theta = 1$ (which corresponds to the original Poor-Verdú reliability function upper bound). We also numerically show that for the memoryless Z-channel, $E_{\rm PV}^{(\theta)}(R)$ is not tight at high rates for all considered values of θ (including large ones).

The rest of the paper is organized as follows. In Section II, the generalized Poor-Verdú lower bound to the multihypothesis testing minimum error probability is established in terms of the tilted posterior distribution with parameter θ (Theorem 1). A sufficient condition under which an exact expression for the error probability is given in terms of an asymptotic (in θ) term of the bound (minus a multiplying factor) is also shown (Theorem 2). Examples illustrating Theorems 1 and 2 are provided in Section III. In Section IV, the two upper bounds, given by $E_{\rm PV}^{(\theta)}(R)$ and $\bar{E}_{\rm PV}(R)$, respectively, for the channel reliability function are proved (Theorem 3). Furthermore, it is noted that $\bar{E}_{\rm PV}(R)$ provides an exact asymptotic characterization for the channel reliability function at all rates for the finite-input AWGN channel as well

[1]The sphere packing bound [9] is referred to as the space partitioning bound in [5].

[2]For the finite-input AWGN channel as well as the whole class of memoryless channels, $E^*(R)$ is already exactly determined in terms of a simple (single-letter) expression at high rates (beyond some critical rate) since the random coding and sphere-packing bounds coincide in that rate region [9]. Further improvements were recently established for the memoryless binary symmetric channel (BSC) and the continuous-input AWGN channel in [2], [3], where it is shown that $E^*(R)$ is also exactly determined for rates R in some interval directly below the critical rate.

as other channels (Theorem 4 and Corollary 1). Numerical examples involving the BSC and the Z-channel indicating the looseness of $E_{\text{PV}}^{(\theta)}(R)$ for specific choices of θ are next provided. Finally, conclusions are stated in section V. Note that we will use the natural logarithm throughout.

II. A GENERALIZED ERROR LOWER BOUND FOR MULTIHYPOTHESIS TESTING

We herein generalize the Poor-Verdú lower bound in (1) for the multihypothesis testing error probability.

Consider two (correlated) random variables X and Y, where X has a discrete (i.e., finite or countably infinite) alphabet $\mathcal{X} = \{x_1, x_2, x_3, \ldots\}$ and Y takes on values in an arbitrary alphabet $\mathcal{Y}$. The minimum probability of error P_e in estimating X from Y is given by

$$P_e \triangleq \Pr\left[X \neq e(Y)\right] \tag{2}$$

where $e(Y)$ is the MAP estimate defined as

$$e(Y) = \arg\max_{x \in \mathcal{X}} P_{X|Y}(x|Y). \tag{3}$$

Theorem 1: The above minimum probability of error P_e in estimating X from Y satisfies the following inequality

$$P_e \geq (1-\alpha) P_{X,Y}\left\{(x,y) \in \mathcal{X} \times \mathcal{Y} : \; P_{X|Y}^{(\theta)}(x|y) \leq \alpha\right\} \tag{4}$$

for each $\alpha \in [0,1]$ and $\theta \geq 1$, where for each $y \in \mathcal{Y}$,

$$P_{X|Y}^{(\theta)}(x|y) \triangleq \frac{P_{X|Y}^{\theta}(x|y)}{\sum_{x' \in \mathcal{X}} P_{X|Y}^{\theta}(x'|y)}, \quad x \in \mathcal{X}, \tag{5}$$

is the tilted distribution of $P_{X|Y}(\cdot|y)$ with parameter θ [6].

Note: When $\theta = 1$, the above bound in (4) reduces to the Poor-Verdú bound in (1).

Proof: Fix $\theta \geq 1$. We only provide the proof for $\alpha < 1$ since the lower bound trivially holds when $\alpha = 1$.

From (2) and (3), the minimum error probability P_e incurred in testing among the values of X satisfies

$$\begin{aligned}
1 - P_e &= \Pr[X = e(Y)] \\
&= \int_{\mathcal{Y}} P_{X|Y}(e(y)|y)\; dP_Y(y) \\
&= \int_{\mathcal{Y}} \left(\max_{x \in \mathcal{X}} P_{X|Y}(x|y)\right) dP_Y(y) \\
&= \int_{\mathcal{Y}} \left(\max_{x \in \mathcal{X}} f_x(y)\right) dP_Y(y) \\
&= E\left[\max_{x \in \mathcal{X}} f_x(Y)\right],
\end{aligned}$$

where $f_x(y) \triangleq P_{X|Y}(x|y)$. For a fixed $y \in \mathcal{Y}$, let $h_j(y)$ be the j-th element in the set

$$\{f_{x_1}(y), f_{x_2}(y), f_{x_3}(y), \ldots\}$$

such that its elements are listed in non-increasing order; i.e.,

$$h_1(y) \geq h_2(y) \geq h_3(y) \geq \cdots$$

and

$$\{h_1(y), h_2(y), h_3(y), \ldots\} = \{f_{x_1}(y), f_{x_2}(y), f_{x_3}(y), \ldots\}.$$

Then

$$1 - P_e = E[h_1(Y)]. \tag{6}$$

Furthermore, for each $h_j(y)$ above, define $h_j^{(\theta)}(y)$ such that $h_j^{(\theta)}(y)$ be the respective element for $h_j(y)$ satisfying

$$h_j(y) = f_{x_j}(y) = P_{X|Y}(x_j|y) \;\Leftrightarrow\; h_j^{(\theta)}(y) = P_{X|Y}^{(\theta)}(x_j|y).$$

Since $h_1(y)$ is the largest among $\{h_j(y)\}_{j\geq 1}$,

$$h_1^{(\theta)}(y) = \frac{h_1^\theta(y)}{\sum_{j\geq 1} h_j^\theta(y)} = \frac{1}{1 + \sum_{j\geq 2}[h_j(y)/h_1(y)]^\theta}$$

is non-decreasing in θ for each y; this implies that

$$h_1^{(\theta)}(y) \geq h_1(y) \quad \text{for } \theta \geq 1 \text{ and } y \in \mathcal{Y}. \tag{7}$$

For any $\alpha \in [0, 1)$, we can write

$$P_{X,Y}\left\{(x, y) \in \mathcal{X} \times \mathcal{Y} : \; P_{X|Y}^{(\theta)}(x|y) > \alpha\right\} = \int_{\mathcal{Y}} P_{X|Y}\left\{x \in \mathcal{X} : \; P_{X|Y}^{(\theta)}(x|y) > \alpha\right\} dP_Y(y).$$

Noting that

$$\begin{aligned} P_{X|Y}\left\{x \in \mathcal{X} \;:\; P_{X|Y}^{(\theta)}(x|y) > \alpha\right\} &= \sum_{x\in\mathcal{X}} P_{X|Y}(x|y) \cdot \mathbf{1}\left(P_{X|Y}^{(\theta)}(x|y) > \alpha\right) \\ &= \sum_{j=1}^{\infty} h_j(y) \cdot \mathbf{1}\left(h_j^{(\theta)}(y) > \alpha\right), \end{aligned}$$

where $\mathbf{1}(\cdot)$ is the indicator function, yields

$$\begin{aligned} P_{X,Y}\left\{(x, y) \in \mathcal{X} \times \mathcal{Y} : \; P_{X|Y}^{(\theta)}(x|y) > \alpha\right\} &= \int_{\mathcal{Y}} \left(\sum_{j=1}^{\infty} h_j(y) \cdot \mathbf{1}\left(h_j^{(\theta)}(y) > \alpha\right)\right) dP_Y(y) \\ &\geq \int_{\mathcal{Y}} h_1(y) \cdot \mathbf{1}(h_1^{(\theta)}(y) > \alpha) dP_Y(y) \\ &\geq \int_{\mathcal{Y}} h_1(y) \cdot \mathbf{1}(h_1(y) > \alpha) dP_Y(y) \\ &= E[h_1(Y) \cdot \mathbf{1}(h_1(Y) > \alpha)], \end{aligned} \tag{8}$$

where the second inequality follows from (7). To complete the proof, we next relate $E[h_1(Y)\cdot \mathbf{1}(h_1(Y) > \alpha)]$ with $E[h_1(Y)]$, which is exactly $1-P_e$. Invoking [12, eq. (19)], we have that for any $\alpha \in [0,1]$ and any random variable U with $\Pr\{0 \le U \le 1\} = 1$, the following inequality holds with probability one

$$U \le \alpha + (1-\alpha)\cdot U \cdot \mathbf{1}(U > \alpha).$$

Thus

$$E[U] \le \alpha + (1-\alpha)E[U \cdot \mathbf{1}(U > \alpha)].$$

Applying the above inequality to (8) by setting $U = h_1(Y)$, we obtain

$$\begin{aligned} (1-\alpha)P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} \;:\; P^{(\theta)}_{X|Y}(x|y) > \alpha\right\} &\ge (1-\alpha)E[h_1(Y)\cdot\mathbf{1}(h_1(Y)>\alpha)] \\ &\ge E[h_1(Y)] - \alpha \\ &= (1-P_e) - \alpha \\ &= (1-\alpha) - P_e, \end{aligned}$$

where the first equality follows from (6). ■

We next show that if the MAP estimate $e(Y)$ of X from Y is almost surely unique in (3), then the bound of Theorem 1, without the $(1-\alpha)$ factor, is tight in the limit of θ going to infinity.

Theorem 2: Consider two random variables X and Y, where X has a finite or countably infinite alphabet $\mathcal{X} = \{x_1, x_2, x_3, \ldots\}$ and Y has an arbitrary alphabet $\mathcal{Y}$. Assume that

$$P_{X|Y}(e(y)|y) > \max_{x\in\mathcal{X}:x\ne e(y)} P_{X|Y}(x|y) \tag{9}$$

holds almost surely in P_Y, where $e(y)$ is the MAP estimate from y as defined in (3); in other words, the MAP estimate is almost surely unique in P_Y. Then, the error probability in the MAP estimation of X from Y satisfies

$$P_e = \lim_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\; P^{(\theta)}_{X|Y}(x|y) \le \alpha\right\} \tag{10}$$

for each $\alpha \in (0,1)$, where the tilted distribution $P^{(\theta)}_{X|Y}(\cdot|y)$ is given in (5) for $y \in \mathcal{Y}$.

Proof: It can be easily verified from the definitions of $h_j(\cdot)$ and $h^{(\theta)}_j(\cdot)$ that the following two limits hold for each $y \in \mathcal{Y}$:

$$\lim_{\theta\to\infty} h^{(\theta)}_1(y) = \frac{1}{\ell(y)},$$

where

$$\ell(y) \triangleq \max\{j \in \mathbb{N} : h_j(y) = h_1(y)\} \tag{11}$$

and $\mathbb{N} \triangleq \{1, 2, 3, \ldots\}$ is the set of positive integers, and

$$\lim_{\theta\to\infty} h_j(y)\cdot \mathbf{1}\left(h_j^{(\theta)}(y) > \alpha\right) = \begin{cases} h_j(y)\cdot \mathbf{1}\left(\frac{1}{\ell(y)} > \alpha\right) & \text{for } j = 1, 2, \cdots, \ell(y) \\ 0 & \text{for } j > \ell(y) \end{cases} \tag{12}$$

where $\mathbf{1}(\cdot)$ is the indicator function.

As a result, we obtain that for any $\alpha \in [0, 1)$,

$$\begin{aligned} &\lim_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\ P^{(\theta)}_{X|Y}(x|y) > \alpha\right\} \\ &= \lim_{\theta\to\infty}\int_{\mathcal{Y}}\left(\sum_{j=1}^{\infty} h_j(y)\cdot\mathbf{1}\left(h_j^{(\theta)}(y) > \alpha\right)\right) dP_Y(y) \\ &= \int_{\mathcal{Y}}\lim_{\theta\to\infty}\left(\sum_{j=1}^{\infty} h_j(y)\cdot\mathbf{1}\left(h_j^{(\theta)}(y) > \alpha\right)\right) dP_Y(y) \qquad (13) \\ &= \int_{\mathcal{Y}}\left(\sum_{j=1}^{\ell(y)} h_j(y)\cdot\mathbf{1}\left(\frac{1}{\ell(y)} > \alpha\right)\right) dP_Y(y), \qquad (14) \end{aligned}$$

where (13) follows from the Dominated Convergence Theorem [4, Thm. 16.4] since

$$\left|\sum_{j=1}^{\infty} h_j(y)\cdot\mathbf{1}\left(h_j^{(\theta)}(y) > \alpha\right)\right| \le \sum_{j=1}^{\infty} h_j(y) = 1.$$

Furthermore, (14) holds since the limit (in θ) of

$$a_{\theta,j} \triangleq h_j(y)\cdot\mathbf{1}\left(h_j^{(\theta)}(y) > \alpha\right)$$

exists for every $j = 1, 2, \cdots$ by (12), hence implying (as shown in Appendix A) that

$$\lim_{\theta\to\infty}\sum_{j=1}^{\infty} a_{\theta,j} = \sum_{j=1}^{\infty}\lim_{\theta\to\infty} a_{\theta,j}.$$

Now condition (9) is equivalent to

$$\Pr[\ell(Y) = 1] \triangleq P_Y\{y\in\mathcal{Y} : \ell(y) = 1\} = 1; \tag{15}$$

thus,

$$\begin{aligned} \lim_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\ P^{(\theta)}_{X|Y}(x|y) > \alpha\right\} &= \int_{\mathcal{Y}} h_1(y)\cdot\mathbf{1}(1 > \alpha)\, dP_Y(y) = E[h_1(Y)] \\ &= 1 - P_e, \qquad (16) \end{aligned}$$

where (16) follows from (6).

This immediately yields that for $0<\alpha<1$,

$$\begin{aligned} P_e &= 1-\lim_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\ P_{X|Y}^{(\theta)}(x|y)>\alpha\right\} \\ &= \lim_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\ P_{X|Y}^{(\theta)}(x|y)\le\alpha\right\}. \end{aligned}$$

■

Observation 1: We first note that since the bound in (4) holds for every $\theta\ge 1$, it also holds in the limit of θ going to infinity (the limit exists as shown in the above proof):

$$P_e \ge (1-\alpha)\lim_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\ P_{X|Y}^{(\theta)}(x|y)\le\alpha\right\} \tag{17}$$

for any $0\le\alpha\le 1$.

Furthermore, if condition (9) does not hold (or equivalently from (15), if $\Pr[\ell(Y)=1]<1$), but there exists an integer $L>1$ such that $\Pr[\ell(Y)\le L]=1$, then using (14), we can write (17) as

$$\begin{aligned} P_e &\ge (1-\alpha)\left[1-\int_{\mathcal{Y}}\left(\sum_{j=1}^{\ell(y)} h_j(y)\cdot\mathbf{1}\left(\frac{1}{\ell(y)}>\alpha\right)\right)dP_Y(y)\right] \\ &= (1-\alpha)\left[\int_{\mathcal{Y}}\left(\sum_{j=1}^{\infty} h_j(y)\right)dP_Y(y)-\int_{\mathcal{Y}}\left(\sum_{j=1}^{\ell(y)} h_j(y)\cdot\mathbf{1}\left(\frac{1}{\ell(y)}>\alpha\right)\right)dP_Y(y)\right] \\ &= (1-\alpha)\int_{\mathcal{Y}}\left(\sum_{j=1}^{\ell(y)} h_j(y)\cdot\mathbf{1}\left(\frac{1}{\ell(y)}\le\alpha\right)+\sum_{j=\ell(y)+1}^{\infty} h_j(y)\right)dP_Y(y) \qquad (18) \\ &= (1-\alpha)\left[\int_{y:\ell(y)=1}\left(\sum_{j=1}^{1} h_j(y)\cdot\mathbf{1}(1\le\alpha)+\sum_{j=2}^{\infty} h_j(y)\right)dP_Y(y)\right. \\ &\quad +\int_{y:\ell(y)=2}\left(\sum_{j=1}^{2} h_j(y)\cdot\mathbf{1}\left(\frac{1}{2}\le\alpha\right)+\sum_{j=3}^{\infty} h_j(y)\right)dP_Y(y) \\ &\quad \left.+\cdots+\int_{y:\ell(y)=L}\left(\sum_{j=1}^{L} h_j(y)\cdot\mathbf{1}\left(\frac{1}{L}\le\alpha\right)+\sum_{j=L+1}^{\infty} h_j(y)\right)dP_Y(y)\right]. \qquad (19) \end{aligned}$$

To render this lower bound as large as possible, its formula above indicates that although the multiplicative constant $(1-\alpha)$ favors a small α, the integration term in (18) actually has its smallest value when α is less than $1/L$ (see (19)). Therefore, a compromise in the choice of α has to be made in order to maximize the bound.

III. Examples for the generalized Poor-Verdú bound

In this section, we provide four examples (three of them with a finite observation alphabet and one with a continuous observation alphabet) to illustrate the results of the previous section.

A. Ternary Hypothesis Testing

We revisit the ternary hypothesis testing example examined in [12, Figs. 1 and 2], where random variables X and Y have identical alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, X is uniformly distributed ($P_X(x) = 1/3 \ \forall x \in \mathcal{X}$) and Y is related to X via

$$P_{Y|X}(y|x) = \begin{cases} 1 - v_1 - v_2 & \text{if } y = x \\ v_1 & \text{if } x = 1 \text{ and } y = 0 \\ v_2 & \text{if } x = 2 \text{ and } y = 0 \\ v_1 & \text{if } y \neq x \text{ and } y = 1 \\ v_2 & \text{if } y \neq x \text{ and } y = 2 \end{cases}$$

where we assume that $1 - v_1 - v_2 > v_2 > v_1 > 0$. In [12], $v_1 = 0.27$ and $v_2 = 0.33$ are used.

A direct calculation reveals that the MAP estimation function (3) for guessing X from Y is given by $e(y) = y$ for every $y \in \mathcal{Y}$, resulting in a probability of error of $P_e = v_1 + v_2 = 0.6$ when $v_1 = 0.27$ and $v_2 = 0.33$. Furthermore, we obtain that P_e is exactly determined via

$$\lim_{\theta \to \infty} P_{X,Y} \left\{ (x, y) \in \mathcal{X} \times \mathcal{Y} : \ P^{(\theta)}_{X|Y}(x|y) \le \alpha \right\} = v_1 + v_2 = P_e;$$

as predicted by Theorem 2, since condition (9) holds (since $\ell(Y) = 1$ almost surely in P_Y, where $\ell(\cdot)$ is defined in (11)).

We next compute the new bound in (4) for $v_1 = 0.27$, $v_2 = 0.33$ and for different values of $\theta \ge 1$ and plot it in Fig. 1, along with Fano's original bound (referred to as "Fano" in the figure) given by

$$P_e \ge \frac{\log 3 - I(X; Y) - \log 2}{\log 2} = 0.568348,$$

and Fano's weaker (but commonly used) bound

$$P_e \ge 1 - \frac{I(X; Y) + \log 2}{\log 3} = 0.358587$$

shown in [12, Fig. 2] (referred to as "Weakened Fano" in the figure). The case of $\theta = 1$ corresponds to the original Poor-Verdú bound in (1). As can be seen from the figure, bound (4) for $\theta = 20$ and 100 improves upon (1) and both Fano bounds and approaches the exact probability of error as θ is increased without bound (e.g., for $\theta = 100$ and $\alpha \downarrow 0$, the bound is quite close to P_e). In Fig. 2, bounds (4) and (1), maximized over $\alpha \in [0, 1]$, are plotted versus θ. It is observed that when $\theta \ge 16$, bound (4) improves upon (1).

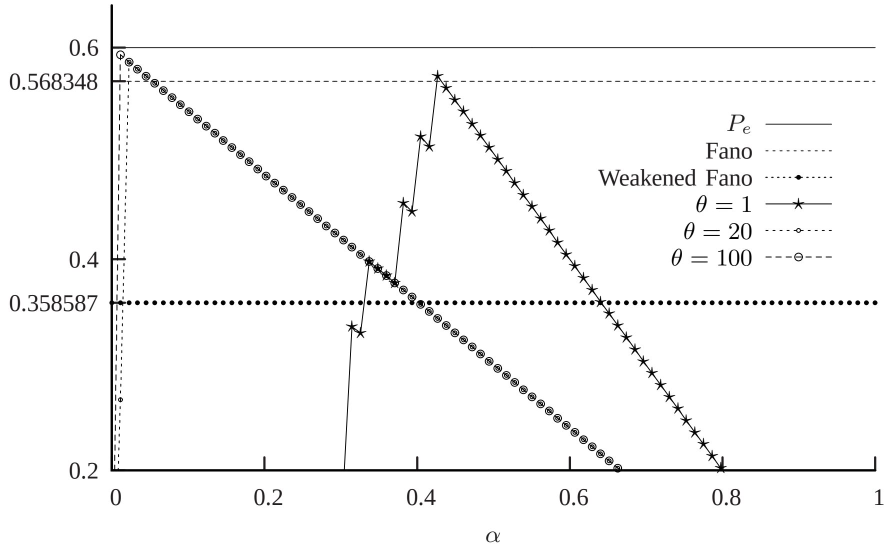

Fig. 1. Lower bounds on the minimum probability of error for Example III-A: bound (4) versus α for $\theta = 1, 20, 100$ and Fano's original and weakened bounds.

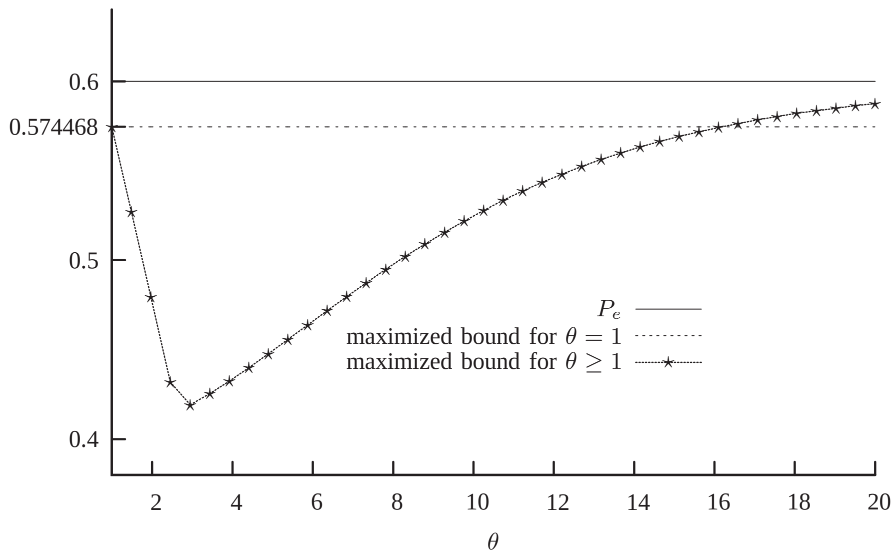

Fig. 2. Lower bounds on the minimum probability of error for Example III-A: bounds (1) and (4) versus θ optimized over α.

B. Binary Erasure Channel

Suppose that X and Y are respectively the channel input and output of a BEC with erasure probability ε, where $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, \mathtt{E}\}$. Let $\Pr[X = 0] = 1 - p$ and $\Pr[X = 1] = p$ with $0 < p < 1/2$.

Then, the MAP estimate of X from Y is given by

$$e(y) = \begin{cases} y & \text{if } y \in \{0,1\} \\ 0 & \text{if } y = \mathtt{E} \end{cases}$$

and the resulting error probability is $P_e = \varepsilon p$.

Calculating bound (4) of Theorem 1 yields

$$\begin{aligned}(1-\alpha)P_{X,Y}&\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\ P^{(\theta)}_{X|Y}(x|y)\le\alpha\right\}\\ &= \begin{cases} 0 & \text{if } 0 \le \alpha < \dfrac{p^\theta}{p^\theta+(1-p)^\theta} \\ \varepsilon p(1-\alpha) & \text{if } \dfrac{p^\theta}{p^\theta+(1-p)^\theta} \le \alpha < \dfrac{(1-p)^\theta}{p^\theta+(1-p)^\theta} \\ \varepsilon(1-\alpha) & \text{if } \dfrac{(1-p)^\theta}{p^\theta+(1-p)^\theta} \le \alpha < 1. \end{cases}\end{aligned} \tag{20}$$

Thus, taking $\theta \uparrow \infty$ and then $\alpha \downarrow 0$ in (20) results in the exact error probability εp. Note that in this example, the original Poor-Verdú bound (i.e., with $\theta = 1$) also achieves the exact error probability εp by choosing $\alpha = 1-p$; however this maximizing choice of $\alpha = 1-p$ for the original bound is a function of system's statistics (here, the input distribution p) which is undesirable. On the other hand, the generalized bound (4) can herein achieve its peak by systematically taking $\theta \uparrow \infty$ and then letting $\alpha \downarrow 0$.

Furthermore, since in this example, $\ell(y) = 1$ for every $y \in \{0,1,\mathtt{E}\}$, we have that (9) holds; hence, by Theorem 2, (10) yields

$$\begin{aligned} P_e &= \lim_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\ P^{(\theta)}_{X|Y}(x|y)\le\alpha\right\} \\ &= \varepsilon p \text{ for } 0 \le \alpha < 1, \end{aligned}$$

where the last equality follows directly from (20) without the $(1-\alpha)$ factor.

C. Multiple-Use BEC

We now extend the previous example of the single-use BEC to the case of using the memoryless BEC n times with an input n-tuple $X^n = (X_1, \cdots, X_n)$ of independent and identically distributed (i.i.d.) random variables X_i with $\Pr[X_i = 1] = p$, where $0 < p < 1/2$. Here again we determine the MAP estimation of X^n by observing the channel output Y^n. For a received output n-tuple y^n,

$$P_{X^n|Y^n}(x^n|y^n) = \begin{cases} (1-p)^{d_{0\mathtt{E}}(x^n,y^n)}p^{d_{1\mathtt{E}}(x^n,y^n)} & \text{if } d_{01}(x^n,y^n) = d_{10}(x^n,y^n) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where $d_{0\mathtt{E}}(x^n, y^n)$ is the number of occurrences of $(x_j, y_j) = (0, \mathtt{E})$ in (x^n, y^n), and the other d-terms are defined similarly. The above equation indicates that for a given y^n, $P_{X^n|Y^n}(x^n|y^n)$ always peaks for $d_{1\mathtt{E}}(x^n, y^n) = 0$ since $0 < p < 1/2$. Thus the MAP estimator $e(y^n)$ replaces all erasures in y^n by 0's while keeping the 0's and 1's in y^n unchanged (e.g., if $n = 5$ and $y^n = (0, 0, \mathtt{E}, \mathtt{E}, 1)$, then $e(y^n) = (0, 0, 0, 0, 1)$). The resulting probability of error is given by

$$\begin{aligned} P_e &= 1 - \sum_{y^n \in \mathcal{Y}^n} P_{X^n}(e(y^n)) P_{Y^n|X^n}(y^n|e(y^n)) \\ &= 1 - \sum_{k=0}^{n} \sum_{i=0}^{n-k} \binom{n}{k} \binom{n-k}{i} (1-p)^{n-i} p^i \varepsilon^k (1-\varepsilon)^{n-k} \\ &= 1 - (1 - \varepsilon p)^n \end{aligned}$$

where k is the number of erasures $\mathtt{E}$ in y^n and i is the number of 1's in y^n.

On the other hand, we directly obtain from (21) that condition (9) holds (or equivalently condition (15), i.e., $\ell(y^n) = 1$ with probability one in P_{Y^n}). We can then apply Theorem 2 to obtain from (10) that

$$\begin{aligned} P_e &= 1 - (1 - \varepsilon p)^n \\ &= \lim_{\theta \to \infty} P_{X^n, Y^n} \left\{ (x^n, y^n) \in \mathcal{X} \times \mathcal{Y} : \ P^{(\theta)}_{X^n|Y^n}(x^n|y^n) \le \alpha \right\}. \end{aligned}$$

We next consider the case of $p = 1/2$, i.e,. the input X^n is uniformly distributed. In this case, (21) yields that

$$h_1(y^n) = h_2(y^n) = \cdots = h_{2^k}(y^n) = 2^{-k}$$

and

$$h_{2^k+1}(y^n) = h_{2^k+2}(y^n) = \cdots = h_{2^n}(y^n) = 0$$

where k is the number of erasures $\mathtt{E}$ in y^n. Thus $\ell(y^n) = 2^k$ and Theorem 2 no longer holds. Furthermore, $h_j^{(\theta)}(y^n) = h_j(y^n)$ for every $\theta \ge 1$; this implies that for the uniform-input multiple-use BEC, the generalized bound (4) does not improve upon the original Poor-Verdú bound (1).

D. Binary Input Observed in Gaussian Noise

We herein consider an example with a continuous observation alphabet $\mathcal{Y} = \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers. Specifically, let the observation be given by $Y = X + N$, where X is uniformly distributed over $\mathcal{X} = \{-1, +1\}$ and N is a zero-mean Gaussian random variable with variance σ^2. Assuming that

X and N are independent from each other, then

$$
\begin{aligned}
P_{X|Y}(x|y) &= \frac{\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(y-x)^2}{2\sigma^2}\}}{\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(y-1)^2}{2\sigma^2}\}+\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(y+1)^2}{2\sigma^2}\}} \\
&= \frac{\exp\{\frac{xy}{\sigma^2}\}}{\exp\{\frac{y}{\sigma^2}\}+\exp\{-\frac{y}{\sigma^2}\}} = \frac{1}{1+\exp\{-\frac{2xy}{\sigma^2}\}}
\end{aligned} \tag{22}
$$

for $x \in \{-1,+1\}$, $y \in \mathbb{R}$. This directly implies that the MAP estimate of X from Y is given by $e(y) = +1$ if $y > 0$ and $e(y) = -1$ if $y \le 0$. The resulting error probability is $P_e = \Phi(-1/\sigma)$, where $\Phi(z) \triangleq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp -\frac{t^2}{2}dt$ is the cdf of the standard (zero-mean unit-variance) Gaussian distribution. Furthermore, since $x \in \{-1,+1\}$, we have

$$
P^{(\theta)}_{X|Y}(x|y) = \frac{\left(\frac{\exp\{\frac{xy}{\sigma^2}\}}{\exp\{\frac{y}{\sigma^2}\}+\exp\{-\frac{y}{\sigma^2}\}}\right)^\theta}{\left(\frac{\exp\{\frac{y}{\sigma^2}\}}{\exp\{\frac{y}{\sigma^2}\}+\exp\{-\frac{y}{\sigma^2}\}}\right)^\theta+\left(\frac{\exp\{\frac{-y}{\sigma^2}\}}{\exp\{\frac{y}{\sigma^2}\}+\exp\{-\frac{y}{\sigma^2}\}}\right)^\theta} = \frac{1}{1+\exp\{-\frac{2xy}{\sigma^2/\theta}\}},
$$

and the generalized Poor-Verdú bound (4) yields

$$
\begin{aligned}
P_e &\ge (1-\alpha)P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} \ : \ P^{(\theta)}_{X|Y}(x|y)\le\alpha\right\} \\
&= (1-\alpha)P_X(-1)\int_{y\in\mathbb{R}\ :\ \frac{1}{1+\exp\left\{\frac{2y}{\sigma^2/\theta}\right\}}\le\alpha}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y+1)^2}{2\sigma^2}\right\}dy \\
&\quad +(1-\alpha)P_X(1)\int_{y\in\mathbb{R}\ :\ \frac{1}{1+\exp\left\{-\frac{2y}{\sigma^2/\theta}]\right\}}\le\alpha}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y-1)^2}{2\sigma^2}\right\}dy \\
&= \frac{(1-\alpha)}{2}\int_{\frac{\sigma^2}{2\theta}\log\left(\frac{1}{\alpha}-1\right)}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y+1)^2}{2\sigma^2}\right\}dy \\
&\quad +\frac{(1-\alpha)}{2}\int_{-\infty}^{-\frac{\sigma^2}{2\theta}\log\left(\frac{1}{\alpha}-1\right)}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y-1)^2}{2\sigma^2}\right\}dy \\
&= (1-\alpha)\int_{-\infty}^{-\frac{\sigma^2}{2\theta}\log\left(\frac{1}{\alpha}-1\right)-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{t^2}{2\sigma^2}\right\}dt \\
&= (1-\alpha)\Phi\left(-\frac{\sigma}{2\theta}\log\left(\frac{1}{\alpha}-1\right)-\frac{1}{\sigma}\right).
\end{aligned} \tag{23}
$$

Now taking the limits $\theta \uparrow \infty$ followed by $\alpha \downarrow 0$ for the right-hand side term in (23) yields exactly $\Phi\left(-\frac{1}{\sigma}\right) = P_e$; hence the generalized Poor-Verdú bound (4) is asymptotically tight. The bound is illustrated in Fig. 3 for $\sigma = 0.429858$ which gives $P_e = 0.01$. It can be seen that for $\theta = 100$ and $\alpha \downarrow 0$, bound (4) is quite close to P_e. Finally note that (22) directly ascertains that condition (9) of Theorem 2 holds; thus P_e is given by (10).

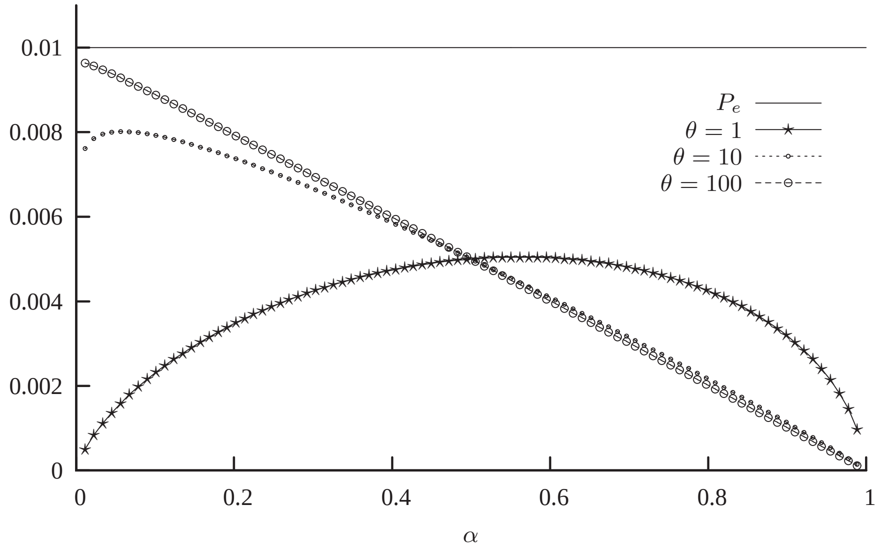

Fig. 3. Example III-D: bound (4) versus α for $\theta = 1, 10, 100$; $\sigma = 0.429858$ and $P_e = 0.01$.

IV. Channel reliability function

We next use the results of Section II to study the channel reliability function.

A. Preliminaries

Consider an arbitrary input process $\mathbf{X}$ defined by a sequence of finite-dimensional distributions [14], [10]

$$\mathbf{X} \triangleq \left\{ X^n = \left(X_1^{(n)}, \cdots, X_n^{(n)} \right) \right\}_{n=1}^{\infty}.$$

Denote by

$$\mathbf{Y} \triangleq \left\{ Y^n = \left(Y_1^{(n)}, \cdots, Y_n^{(n)} \right) \right\}_{n=1}^{\infty}$$

the corresponding output process induced by $\mathbf{X}$ via a general channel with memory

$$\mathbf{W} \triangleq \{ W^n = P_{Y^n|X^n} : \mathcal{X}^n \to \mathcal{Y}^n \}_{n=1}^{\infty}$$

which is an arbitrary sequence of n-dimensional conditional distributions from $\mathcal{X}^n$ to $\mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively.

We assume throughout this section that $\mathcal{X}$ is finite and that $\mathcal{Y}$ is arbitrary. Note though that for the sake of clarity, we adopt the notations of a discrete probability space for $\mathcal{Y}$ with the usual caveats (such as replacing summations with integrals and working with the appropriate probability measures, e.g., see [10, Remark 3.2.1]).

Definition 1 (Channel block code): An (n, M) code $\mathcal{C}_n$ for channel $\boldsymbol{W}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is a pair of maps (f, g), where

$$f : \{1, 2, \cdots, M\} \to \mathcal{X}^n$$

is the encoding function yielding codewords $f(1), f(2), \cdots, f(M) \in \mathcal{X}^n$, each of length n, and

$$g : \mathcal{Y}^n \to \{1, 2, \cdots, M\}$$

is the decoding function. The set of the M codewords is called the codebook and we also usually write $\mathcal{C}_n = \{f(1), f(2), \cdots, f(M)\}$ to list the codewords.

The set $\{1, 2, \ldots, M\}$ is called the message set and we assume that a message V is drawn from the message set according to the uniform distribution. To convey message V over channel $\boldsymbol{W}$, its corresponding codeword $X^n = f(V)$ is sent over the channel. Then Y^n is received at the channel output and $\hat{V} = g(Y^n)$ is yielded as the message estimate.

The code's average error probability (or average probability of decoding error) is given by

$$P_e(\mathcal{C}_n) \triangleq \frac{1}{M} \sum_{m=1}^{M} \sum_{\{y^n : g(y^n) \neq m\}} W^n(y^n | f(m)).$$

Since message V is uniformly distributed over $\{1, 2, \ldots, M\}$, we have that $P_e(\mathcal{C}_n) = \Pr[V \neq \hat{V}]$.

Definition 2 (Channel reliability function [12]): For any $R > 0$, define the channel reliability function $E^*(R)$ for a channel $\boldsymbol{W}$ as the largest scalar $\beta > 0$ such that there exists a sequence of $\mathcal{C}_n = (n, M_n)$ codes with[3]

$$\beta \leq \liminf_{n \to \infty} -\frac{1}{n} \log P_e(\mathcal{C}_n)$$

and

$$R < \liminf_{n \to \infty} \frac{1}{n} \log M_n. \tag{24}$$

Observation 2: We have adopted the above definition of channel reliability function from [12] for the sake of consistency. Note that this definition is not exactly identical to the traditional definition of the channel reliability function. If $P_{e,\min}(n, R)$ denotes the probability of error of the best $(n, \lceil 2^{nR} \rceil)$ code (i.e., the code with smallest error probability) for channel $\boldsymbol{W}$, then the channel's reliability function is

[3] When no $\beta > 0$ satisfies the definition, we simply set $E^*(R) = 0$.

traditionally defined as [4]

$$E(R) = \liminf_{n\to\infty} -\frac{1}{n}\log P_{e,\min}(n,R).$$

However, the following relation can be shown between $E^*(R)$ and $E(R)$:

$$E(R) \geq E^*(R) \geq \lim_{\delta\downarrow 0} E(R+\delta).$$

Thus the above two definitions are equivalent except possibly for discontinuity rate points (of which there are at most countably many as $E^*(R)$ and $E(R)$ are non-increasing in R).

Definition 3 ([14]): Given that Y^n is the output of channel $W^n = P_{Y^n|X^n}$ due to input X^n with distribution P_{X^n}, the channel information density is defined as

$$i_{X^nW^n}(x^n;y^n) \triangleq \log\frac{W^n(y^n|x^n)}{P_{Y^n}(y^n)} = \log\frac{P_{Y^n|X^n}(y^n|x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n} P_{X^n}(\hat{x}^n)P_{Y^n|X^n}(y^n|\hat{x}^n)} \tag{25}$$

for $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$.

Definition 4: Fix $R > 0$. For an input $\mathbf{X}$ and a channel $\mathbf{W}$,

$$\pi_{\boldsymbol{X}}(R) \;\triangleq\; \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}i_{X^nW^n}(x^n;y^n)\leq R\right\} \tag{26}$$

is called a large-deviation rate function for the normalized information density $\frac{1}{n}i_{X^nW^n}(\cdot,\cdot)$.

Proposition 1 (Poor-Verdú upper bound to $E^*(R)$): For a given channel $\boldsymbol{W}$, its reliability function $E^*(R)$ satisfies [12, Eq. (14)], [1, Theorem 1]

$$E^*(R) \leq \sup_{\mathbf{X}} \pi_{\mathbf{X}}(R) \tag{27}$$

for any $R > 0$, where $\pi_{\boldsymbol{X}}(R)$ is the large-deviation rate function for $\frac{1}{n}i_{X^nW^n}(\cdot,\cdot)$ as defined in (26).

Furthermore, the bound in (27) can be slightly tightened by restricting the supremum operation over a smaller set of inputs [1, Corollary 1]:

$$E^*(R) \leq E_{\mathrm{PV}}(R) \triangleq \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \pi_{\boldsymbol{X}}(R), \tag{28}$$

for any $R > 0$, where

$$\mathcal{Q}(R) \triangleq \Big\{\boldsymbol{X} \;:\; \text{Each } X^n \text{ in } \boldsymbol{X} \text{ is uniformly distributed over its support } \mathcal{S}(X^n),$$
$$\text{and } R < \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{S}(X^n)|\Big\}. \tag{29}$$

[4]The limit supremum is also commonly used instead of the limit infimum in the definition of $E(R)$, e.g., see [9, p. 160]. We could have also used the limit supremum in the inequality on β in Definition 2; in that case the results of this section would still hold by replacing $\liminf_n$ with $\limsup_n$ in Theorems 3 and 4 and Corollary 1.

B. Upper Bounds for the Channel Reliability Function

Using Theorem 1, we provide a lower bound for the probability of decoding error of any (n, M) channel code and establish two information-spectrum upper bounds for the channel reliability function.

Theorem 3: Every $\mathscr{C}_n = (n, M)$ code for channel $\boldsymbol{W}$ has its probability of decoding error satisfying

$$P_e(\mathscr{C}_n) \geq (1-\alpha)\, P_{X^nW^n}\left\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : j^{(\theta)}_{X^nW^n}(x^n; y^n) \leq \log(M\alpha)\right\} \tag{30}$$

for every $\alpha \in [0, 1]$ and $\theta \geq 1$, where channel input X^n places probability mass $1/M$ on each codeword of $\mathscr{C}_n$ and

$$j^{(\theta)}_{X^nW^n}(x^n; y^n) \triangleq \log \frac{P^{\theta}_{Y^n|X^n}(y^n|x^n)}{\sum_{\hat{x}^n \in \mathcal{X}^n} P_{X^n}(\hat{x}^n) P^{\theta}_{Y^n|X^n}(y^n|\hat{x}^n)}. \tag{31}$$

Furthermore, the channel's reliability function satisfies

$$\begin{aligned} E^*(R) &\leq \sup_{\boldsymbol{X} \in \mathcal{Q}(R)} \liminf_{n\to\infty} -\frac{1}{n} \log P_{X^nW^n}\left\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n} j^{(\theta)}_{X^nW^n}(x^n; y^n) \leq R\right\} \\ &\triangleq E^{(\theta)}_{\text{PV}}(R) \end{aligned} \tag{32}$$

for every $R > 0$ and $\theta \geq 1$, and

$$\begin{aligned} E^*(R) &\leq \sup_{\boldsymbol{X} \in \mathcal{Q}(R)} \liminf_{n\to\infty} \lim_{\theta\to\infty} -\frac{1}{n} \log P_{X^nW^n}\left\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n} j^{(\theta)}_{X^nW^n}(x^n; y^n) \leq R\right\} \\ &\triangleq \bar{E}_{\text{PV}}(R) \end{aligned} \tag{33}$$

for every $R > 0$, where the set $\mathcal{Q}(R)$ is given in (29).

Proof: When the channel input X^n is uniformly distributed over the code $\mathscr{C}_n \subseteq \mathcal{X}^n$ of size M, the tilted distribution $P^{(\theta)}_{X^n|Y^n}$ of Theorem 1 becomes

$$\begin{aligned} P^{(\theta)}_{X^n|Y^n}(x^n|y^n) &= \frac{P^{\theta}_{X^n|Y^n}(x^n|y^n)}{\sum_{\hat{x}^n \in \mathcal{X}^n} P^{\theta}_{X^n|Y^n}(\hat{x}^n|y^n)} \\ &= \frac{P^{\theta}_{X^n}(x^n) P^{\theta}_{Y^n|X^n}(y^n|x^n)/P^{\theta}_{Y^n}(y^n)}{\sum_{\hat{x}^n \in \mathcal{X}^n} P^{\theta}_{X^n}(\hat{x}^n) P^{\theta}_{Y^n|X^n}(y^n|\hat{x}^n)/P^{\theta}_{Y^n}(y^n)} \\ &= \frac{P^{\theta}_{Y^n|X^n}(y^n|x^n)}{\sum_{\hat{x}^n \in \mathcal{X}^n} P^{\theta}_{Y^n|X^n}(y^n|\hat{x}^n)} \\ &= \frac{P^{\theta}_{Y^n|X^n}(y^n|x^n)/M}{\sum_{\hat{x}^n \in \mathcal{X}^n} P_{X^n}(\hat{x}^n) P^{\theta}_{Y^n|X^n}(y^n|\hat{x}^n)} \end{aligned} \tag{34}$$

for all $x^n \in \mathscr{C}_n$. Hence inequality (30) follows directly from Theorem 1 and (34). We next prove (33); the proof of (32) is identical by omitting the limit over θ. Setting $\alpha = e^{-n\gamma}$ in (30) yields

$$\begin{aligned} -\frac{1}{n}\log P_e(\mathscr{C}_n) \quad &\leq \quad -\frac{1}{n}\log\left(1-e^{-n\gamma}\right) \\ &\quad -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\leq\frac{1}{n}\log M-\gamma\right\}, \end{aligned}$$

which implies in light of (17)

$$\begin{aligned} \liminf_{n\to\infty} -\frac{1}{n}\log P_e(\mathscr{C}_n) \quad &\leq \quad \liminf_{n\to\infty}\lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\Big\{(x^n;y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \\ &\qquad \frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\leq\frac{1}{n}\log M-\gamma\Big\}. \end{aligned}$$

We can then conclude by definition of the channel reliability function that

$$\begin{aligned} E^*(R) \quad &= \quad \sup_{\{\mathscr{C}_n=\mathcal{S}(X^n)\}_{n\geq 1}:\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} -\frac{1}{n}\log P_e(\mathscr{C}_n) \\ &\leq \quad \sup_{\boldsymbol{X}\in\mathcal{Q}(R)}\liminf_{n\to\infty}\lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\Big\{(x^n;y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \\ &\qquad \frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\leq\frac{1}{n}\log|\mathcal{S}(X^n)|-\gamma\Big\}. \end{aligned}$$

When considering only the sequence of codes in $\mathcal{Q}(R)$, we can replace $\frac{1}{n}\log|\mathcal{S}(X^n)|-\gamma$ by R (if γ is chosen to be small enough such that $R<\liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{S}(X^n)|-\gamma$ is valid for the considered input $\boldsymbol{X}$) as such a replacement can only (ultimately) increase the upper bound; we thus obtain

$$E^*(R)\leq\sup_{\boldsymbol{X}\in\mathcal{Q}(R)}\liminf_{n\to\infty}\lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\Big\{(x^n;y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\leq R\Big\}.$$

■

Observation 3: When $\theta=1$, $j^{(\theta)}_{X^nW^n}(x^n;y^n)$ in (31) reduces to

$$\log\frac{P_{Y^n|X^n}(y^n|x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n}P_{X^n}(\hat{x}^n)P_{Y^n|X^n}(y^n|\hat{x}^n)}=\log\frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}=i_{X^nW^n}(x^n;y^n)$$

which is the channel information density as defined in (25).

In this case, the generalized upper bound for the channel reliability function $E^{(\theta)}_{\mathrm{PV}}(R)$ of (32) reduces to the Poor-Verdú upper bound $E_{\mathrm{PV}}(R)$ of (28) (as expected, since for $\theta=1$, (4) reduces to (1)).

Observation 4: Note that when $\theta>1$, the denominator of the fraction in (31) (in other words, $\sum_{\hat{x}^n\in\mathcal{X}^n}P_{X^n}(\hat{x}^n)P^{\theta}_{Y^n|X^n}(y^n|\hat{x}^n)$) is not a legitimate distribution since it does not sum to one over $y^n\in\mathcal{Y}^n$. However, if

$$\sum_{\hat{y}^n\in\mathcal{Y}^n}P^{\theta}_{Y^n|X^n}(\hat{y}^n|x^n)=\sum_{\hat{y}^n\in\mathcal{Y}^n}P^{\theta}_{Y^n|X^n}(\hat{y}^n|\hat{x}^n)\qquad\forall\ x^n,\hat{x}^n\in\mathcal{X}^n, n=1,2,\cdots, \tag{35}$$

then $j^{(\theta)}_{X^nW^n}(x^n;y^n)$ can be reformulated as follows

$$\begin{aligned} j^{(\theta)}_{X^nW^n}(x^n;y^n) &= \log \frac{\dfrac{P^{\theta}_{Y^n|X^n}(y^n|x^n)}{\sum_{\hat{y}^n\in\mathcal{Y}^n} P^{\theta}_{Y^n|X^n}(\hat{y}^n|x^n)}}{\sum_{\hat{x}^n\in\mathcal{X}^n} P_{X^n}(\hat{x}^n)\dfrac{P^{\theta}_{Y^n|X^n}(y^n|\hat{x}^n)}{\sum_{\hat{y}^n\in\mathcal{Y}^n} P^{\theta}_{Y^n|X^n}(\hat{y}^n|\hat{x}^n)}} \\ &= \log \frac{P^{(\theta)}_{Y^n|X^n}(y^n|x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n} P_{X^n}(\hat{x}^n)P^{(\theta)}_{Y^n|X^n}(y^n|x^n)} \\ &\triangleq i^{(\theta)}_{X^n,Y^n}(x^n;y^n), \end{aligned} \tag{36}$$

where for each $y^n \in \mathcal{Y}^n$,

$$P^{(\theta)}_{Y^n|X^n}(y^n|x^n) \triangleq \frac{P^{\theta}_{Y^n|X^n}(y^n|x^n)}{\sum_{\hat{y}^n\in\mathcal{Y}^n} P^{\theta}_{Y^n|X^n}(\hat{y}^n|x^n)} \quad x^n \in \mathcal{X}^n$$

is the *tilted* distribution with parameter θ of the channel statistics $P_{Y^n|X^n}(\cdot|x^n)$. Note that $P^{(\theta)}_{Y^n|X^n}$ is a legitimate distribution (like $P^{(\theta)}_{X|Y}$ defined in Theorem 1). As a result, the new denominator of the fraction in (36) (i.e., $\sum_{\hat{x}^n\in\mathcal{X}^n} P_{X^n}(\hat{x}^n)P^{(\theta)}_{Y^n|X^n}(y^n|x^n)$) is a true distribution on $\mathcal{Y}^n$; it is indeed the distribution of the output due to an input with distribution P_{X^n} sent over a channel with (legitimate) tilted statistics $P^{(\theta)}_{Y^n|X^n}$. We thus conclude that for channels satisfying the invariance condition of (35), the upper bounds for the channel reliability function in (32) and (33) are actually based on the channel information density $i^{(\theta)}_{X^nW^n}(x^n;y^n)$ of an *auxiliary channel* whose transition probability $P^{(\theta)}_{Y^n|X^n}$ is the *tilted* counterpart of the original channel transition probability $P_{Y^n|X^n}$.

When the output alphabet is finite, the channel $\boldsymbol{W}$ satisfies (35) if it is *row-symmetric*, i.e., if the rows of its transition matrix $[p_{x^ny^n}]$ of size $|\mathcal{X}^n| \times |\mathcal{Y}^n|$, where $p_{x^ny^n} \triangleq P_{Y^n|X^n}(y^n|x^n)$, are permutations of each other for each n. Note that channels whose transition matrix $[p_{x^ny^n}]$ is symmetric in the Gallager sense [9, p. 94] for each n are row-symmetric; such channels include the memoryless BSC and BEC.

When the output alphabet is continuous (i.e., with $\mathcal{Y} = \mathbb{R}$) and the channel is described by a sequence of n-dimensional transition (conditional) probability density functions (pdfs) $f_{Y^n|X^n}$, the invariance condition of (35) translates into

$$\int_{\hat{y}^n\in\mathbb{R}^n} f^{\theta}_{Y^n|X^n}(\hat{y}^n|x^n)d\hat{y}_1\cdots d\hat{y}_n = \int_{\hat{y}^n\in\mathbb{R}^n} f^{\theta}_{Y^n|X^n}(\hat{y}^n|\hat{x}^n)d\hat{y}_1\cdots d\hat{y}_n \tag{37}$$

$\forall\ x^n, \hat{x}^n \in \mathcal{X}^n$, $n = 1, 2, \cdots$. The memoryless finite-input AWGN channel and the memoryless binary-input (with $\mathcal{X} = \{-1, +1\}$) *output-symmetric* channel, i.e., whose transition pdf satisfies $f_{Y|X}(y|-1) = f_{Y|X}(-y|+1)\ \forall\ y \in \mathbb{R}$, fulfill (37).

Observation 5: It can be shown along similar lines as the proof of [1, Theorem 1] that one can interchange the supremum and limit infimum (over n) in $E^{(\theta)}_{\rm PV}(R)$ and $\bar{E}_{\rm PV}(R)$ and obtain

$$\lim_{\gamma\downarrow 0}\underline{E}^{(\theta)}_{\rm PV}(R+\gamma)\le E^{(\theta)}_{\rm PV}(R)\le \underline{E}^{(\theta)}_{\rm PV}(R)\quad\text{and}\quad \lim_{\gamma\downarrow 0}\underline{E}_{\rm PV}(R+\gamma)\le \bar{E}_{\rm PV}(R)\le \underline{E}_{\rm PV}(R),\tag{38}$$

where

$$\underline{E}_{\rm PV}(R)\triangleq \liminf_{n\to\infty}\sup_{X^n\in\mathcal{Q}_n(R)}\lim_{\theta\to\infty}-\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n:\ \frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\le R\right\},$$

$$\underline{E}^{(\theta)}_{\rm PV}(R)\triangleq \liminf_{n\to\infty}\sup_{X^n\in\mathcal{Q}_n(R)}-\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n:\frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\le R\right\}$$

and

$$\mathcal{Q}_n(R)\ \triangleq\ \left\{X^n\ :\ P_{X^n}(x^n)\ =\ \frac{1}{|\mathcal{S}(X^n)|}\text{ for }x^n\ \in\ \mathcal{S}(X^n)\text{ and }R\ <\ \frac{1}{n}\log|\mathcal{S}(X^n)|\right\}.$$

The new expressions that take the supremum over $\mathcal{Q}_n(R)$ before letting n approaching infinity provide an alternative possibility for the evaluation of the two bounds. In particular, $\mathcal{Q}_n(R)$ becomes a finite set as the input alphabet is finite; hence, taking the supremum over $\mathcal{Q}_n(R)$ can be replaced with a maximization operation. Inequality (38) nevertheless implies that $E^{(\theta)}_{\rm PV}(R)=\underline{E}^{(\theta)}_{\rm PV}(R)$ and $\bar{E}_{\rm PV}(R)=\underline{E}_{\rm PV}(R)$ almost everywhere in R (since these functions are non-increasing in R).

C. Information-Spectral Characterization of the Reliability Function for a Class of Channels

We next employ Theorem 2 to show that the upper bound in (33) is tight for the memoryless finite-input AWGN channel as well as a larger class of channels, hence providing an information-spectral characterization for the reliability function of these channels. This exact expression $E^*(R)=\bar{E}_{\rm PV}(R)$ holds for all rates R (below channel capacity), albeit its determination in single-letter form (i.e., solving the optimization of a large-deviation rate function) remains a challenging open problem.

We first focus on the Gaussian channel and then present the result for a wider class of channels. Consider a finite-input AWGN channel described by $Y_i=X_i+Z_i$, $i=1,2,\cdots$, where X_i, Y_i and Z_i are the channel's input, output and noise at time i, respectively. We assume that the noise process $\boldsymbol{Z}$ is i.i.d. with each Z_i being a zero-mean Gaussian random variable with variance $\sigma^2>0$. We also assume that the noise and input processes are independent from each other.

Theorem 4: The channel reliability function $E^*(R)$ of the above finite-input AWGN channel satisfies

$$\begin{aligned} E^*(R) &= \bar{E}_{\rm PV}(R) \\ &= \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} \lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n} j^{(\theta)}_{X^nW^n}(x^n;y^n)\le R\right\} \\ &= \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} \lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n} i^{(\theta)}_{X^nW^n}(x^n;y^n)\le R\right\} \end{aligned}$$

for any $0<R<C$, where C denotes the channel's capacity, and $j^{(\theta)}_{X^nW^n}(x^n,y^n)$ and $i^{(\theta)}_{X^nW^n}(x^n,y^n)$ are given in (31) and (36), respectively.

Proof: Fix $0<R<C$. Let its channel input X^n be uniformly distributed over a codebook $\mathcal{C}_n\subset\mathcal{X}^n$ and let Y^n be the corresponding channel output. Then, for $x^n\in\mathcal{C}_n$,

$$\begin{aligned} P_{X^n|Y^n}(x^n|y^n) &= \frac{P_{X^n}(x^n)f_{Y^n|X^n}(y^n|x^n)}{f_{Y^n}(y^n)} \\ &= \frac{1}{|\mathcal{C}_n|\cdot f_{Y^n}(y^n)}\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{\|y^n-x^n\|^2}{2\sigma^2}\right\}, \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm. For a given y^n received at the channel output, if $\ell(y^n)$ as defined in (11) is greater than or equal to 2, then there exist distinct codewords x^n and $\tilde{x}^n$ in $\mathcal{C}_n$ such that

$$\|y^n-x^n\|^2=\|y^n-\tilde{x}^n\|^2,\ \text{equivalently}\ \sum_{i=1}^n(x_i-\tilde{x}_i)y_i=\frac{1}{2}\sum_{i=1}^n(x_i^2-\tilde{x}_i^2);$$

hence such y^n belongs to an (affine) hyperplane in $\mathbb{R}^n$. In other words, we have that

$$\{y^n\in\mathbb{R}^n:\ell(y^n)\ge 2\}\subseteq\mathcal{Y}(\mathcal{C}_n),$$

where

$$\mathcal{Y}(\mathcal{C}_n)\triangleq\left\{y^n\in\mathbb{R}^n:\|y^n-x^n\|^2=\|y^n-\tilde{x}^n\|^2 \text{ for some } x^n,\tilde{x}^n\in\mathcal{C}_n \text{ and } x^n\ne\tilde{x}^n\right\}$$

consists of the union of $\binom{|\mathcal{C}_n|}{2}$ hyperplanes in $\mathbb{R}^n$. But as the Lebesgue measure of every hyperplane in $\mathbb{R}^n$ is zero (since its volume is zero), we then obtain that the above finite union of hyperplanes has Lebesgue measure zero. Thus, $P_{Y^n}\{\mathcal{Y}(\mathcal{C}_n)\}=0$ which directly yields that $\Pr[\ell(Y^n)\ge 2]=0$, and hence $\Pr[\ell(Y^n)=1]=1$. Theorem 2 then implies that

$$P_e(\mathcal{C}_n)=\lim_{\theta\to\infty}P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n: j^{(\theta)}_{X^nW^n}(x^n;y^n)\le\log M+\log\alpha\right\}$$

for $\alpha\in[0,1)$. As a result, with $\alpha=e^{-n\gamma}$ for arbitrarily small $\gamma>0$,

$$\begin{aligned} &\liminf_{n\to\infty}-\frac{1}{n}\log P_e(\mathcal{C}_n) \\ &=\liminf_{n\to\infty}\lim_{\theta\to\infty}-\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n:\frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\le\frac{1}{n}\log|\mathcal{C}_n|-\gamma\right\}, \end{aligned}$$

where $j^{(\theta)}_{X^nW^n}(x^n,y^n)$ is as defined in (31). As stated in the proof of Theorem 3, the channel input that achieves the channel reliability should has the chosen γ and supports satisfying $\liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{S}(X^n)|-\gamma$ strictly larger but arbitrarily close to R. This concludes to

$$\begin{aligned} E^*(R) &= \sup_{\{\mathscr{C}_n=\mathcal{S}(X^n)\}_{n\geq 1}:\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} -\frac{1}{n}\log P_e(\mathscr{C}_n) \\ &= \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} \lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\leq R\right\} \\ &\triangleq \bar{E}_{\text{PV}}(R). \end{aligned}$$

Furthermore, since this channel satisfies (37), we can replace $j^{(\theta)}_{X^nW^n}(x^n;y^n)$ with $i^{(\theta)}_{X^nW^n}(x^n;y^n)$ in the expression of $\bar{E}_{\text{PV}}(R)$ as shown in Observation 4 to obtain that

$$E^*(R) = \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} \lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}i^{(\theta)}_{X^nW^n}(x^n;y^n)\leq R\right\}.$$

■

An information-spectral representation of $E^*(R)$ for the memoryless finite-input AWGN channel is thus established for all rates, although its solution in closed (single-letter) form is still a daunting task.

We emphasize that the above finding also holds for any channel satisfying $\ell(Y^n)=1$ almost surely in P_{Y^n} as shown above; we hence have the following result (which directly follows from Theorem 2 along the same lines as the above proof).

Corollary 1: Given a channel $\boldsymbol{W}$, if for its input $\boldsymbol{X}$ uniform over any block codebook $\mathscr{C}_n$, the following holds almost surely in P_{Y^n}

$$\max_{x^n\in\mathscr{C}_n} P_{Y^n|X^n}(y^n|x^n) > \max_{x^n\in\mathscr{C}_n\setminus\{e(y^n)\}} P_{Y^n|X^n}(y^n|x^n) \tag{39}$$

for each $n=1,2,\cdots$, where $e_{ML}(y^n)=\arg\max_{x^n\in\mathscr{C}_n} P_{Y^n|X^n}(y^n|x^n)$ is the maximum likelihood estimate of the transmitted codeword from the received channel output y^n, then the channel reliability function of $\boldsymbol{W}$ is given by

$$\begin{aligned} E^*(R) &= \bar{E}_{\text{PV}}(R) \\ &= \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} \lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}j^{(\theta)}_{X^nW^n}(x^n;y^n)\leq R\right\} \end{aligned}$$

for any $0<R<C$, where C is the channel's capacity.

Furthermore, if the channel satisfies the invariance conditions (35) or (37), then $j^{(\theta)}_{X^nW^n}(x^n;y^n) = i^{(\theta)}_{X^nW^n}(x^n;y^n)$, which is the information density for the auxiliary channel with transition distribution

$P^{(\theta)}_{Y^n|X^n}$ (i.e., the tilted distribution of the original channel distribution $P_{Y^n|X^n}$). In this case the channel reliability function becomes

$$\begin{aligned} E^*(R) &= \bar{E}_{\rm PV}(R) \\ &= \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} \lim_{\theta\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}i^{(\theta)}_{X^nW^n}(x^n;y^n)\le R\right\} \end{aligned}$$

for any $0<R<C$.

Observation 6: Corollary 1 requires condition (39) to be valid for any block codebook $\mathscr{C}_n$ and for each $n=1,2,\cdots$. One can immediately weaken the condition by considering only sufficiently large n; but without further knowledge on the optimal codebook (equivalently, the optimal channel input $\boldsymbol{X}$ that achieves $\bar{E}_{\rm PV}(R)$), it may be hard to derive an alternative condition for (39) that holds unanimously for *any* codebook. In particular, for discrete memoryless channels (DMC) with finite or countably infinite output alphabets, a codebook that fails condition (39) can always be constructed except if the channels are not noiseless (i.e., perfect).[5] Hence, in its current form, Corollary 1 is not useful for discrete-output channels; instead, it is of interest for continuous-output channels.

Observation 7: In light of the above observation, we further consider channels with continuous-output alphabets. For a channel that admits a channel transition pdf, the proof of Theorem 4 actually indicates that as long as $P_{Y^n}\{\mathcal{Y}(\mathscr{C}_n)\}=0$ for any block codebook $\mathscr{C}_n$, where

$$\mathcal{Y}(\mathscr{C}_n)\triangleq\left\{y^n\in\mathbb{R}^n : f_{Y^n|X^n}(y^n|x^n)=f_{Y^n|X^n}(y^n|\tilde{x}^n) \text{ for some } x^n,\tilde{x}^n\in\mathscr{C}_n \text{ and } x^n\neq\tilde{x}^n\right\},$$

we have $\Pr[\ell(Y^n)=1]=1$ and (39) holds. We note that this is indeed valid for any sequence of transition pdf's for which the number of solutions in y_n satisfying

$$f_{Y^n|X^n}(y^n|x^n)=f_{Y^n|X^n}(y^n|\tilde{x}^n)$$

[5] As a simple proof, note that for a noisy DMC there exist two inputs $a,a'\in\mathcal{X}$ and an output $b\in\mathcal{Y}$ satisfying $\min\{P_{Y|X}(b|a),P_{Y|X}(b|a')\}>0$. Then for a codebook $\mathscr{C}_n$ consisting of two distinct codewords x^n and $\tilde{x}^n$, where one of them is the permutation of the other, and their components are either a or a', we obtain

$$P_{X^n|Y^n}(x^n|y^n)=\frac{P_{X^n}(x^n)P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}=\frac{P_{Y^n|X^n}(y^n|x^n)}{|\mathscr{C}_n|\cdot P_{Y^n}(y^n)}=\frac{P_{X^n}(\tilde{x}^n)P_{Y^n|X^n}(y^n|\tilde{x}^n)}{P_{Y^n}(y^n)}=P_{X^n|Y^n}(\tilde{x}^n|y^n)$$

for the channel output y^n satisfying $y_i=b$ for every $1\le i\le n$; hence, $\ell(y^n)\ge 2$ with $P_{Y^n}(y^n)=\frac{1}{2}P_{Y^n|X^n}(y^n|x^n)+\frac{1}{2}P_{Y^n|X^n}(y^n|\tilde{x}^n)>0$. This codebook therefore violates condition (39).

Notably, for a channel satisfying $\min\{P_{Y|X}(b|a),P_{Y|X}(b|a')\}=0$ for every unequal $a,a'\in\mathcal{X}$ and $b\in\mathcal{Y}$, the error rate is zero for any codebook $\mathscr{C}_n$. So, only under such a noiseless situation can the finite- or countable-output DMC meet the strict requirement that $\ell(Y^n)=1$ with probability one for any codebook $\mathscr{C}_n$.

for given codewords x^n, $\tilde{x}^n$ in $\mathscr{C}_n$ and given y^{n-1} is either finite or countable (as this condition immediately implies that $\mathcal{Y}(\mathscr{C}_n)$ has Lebesgue measure zero). A large class of channels satisfy this condition. For example, channels with memoryless additive noise, where the noise pdf is not uniform or piecewise-uniform, satisfy this condition and hence (39) and Corollary 1. This allows for most standard continuous distributions for the noise, such as the generalized-Gaussian distribution with shape parameter $c > 0$ (e.g., cf. [11]); this distribution includes the Gaussian and Laplacian distributions as special cases, realized for $c = 2$ and $c = 1$, respectively.

D. Examples of Channels for which the $E_{PV}^{(\theta)}(R)$ Bound Is Not Tight

As already mentioned, the (analytical or numerical) computation of both upper bounds, $E_{\text{PV}}^{(\theta)}(R)$ and $\bar{E}_{\text{PV}}(R)$, to the channel reliability function, given in (32) and (33), respectively, is formidable since they involve a difficult supremum operation of input processes in $\mathcal{Q}(R)$ in addition to the limit operations.

We can however lower-bound $E_{\text{PV}}^{(\theta)}(R)$, for a given (fixed) θ, using an auxiliary class of i.i.d. inputs and compare this lower bound to $E_{\text{PV}}^{(\theta)}(R)$ with familiar channel reliability function upper bounds (such as the sphere-packing upper bound). If the former is shown to be strictly larger than the latter for a range of rates, then this indicates that for that particular θ, $E_{\text{PV}}^{(\theta)}(R)$ is not tight. The lower bound to $E_{\text{PV}}^{(\theta)}(R)$, which we denote by $F(R,\theta)$, is derived in Appendix B and given in (43) for the case of memoryless channels. We herein calculate $F(R,\theta)$ numerically to demonstrate that $E_{\text{PV}}^{(\theta)}(R)$ is not tight within a rate range and for certain choices of θ (including $\theta = 1$ which gives the Poor-Verdú bound of (28)); this is shown for two standard binary-input memoryless channels: the BSC and the Z-channel.

1) Memoryless BSC: For the BSC with crossover probability ε, setting $p \triangleq P_{\bar{X}}(1)$ and $s = \frac{1}{1-\rho}$ in (43) yields

$$\begin{aligned} E_{\text{PV}}^{(\theta)}(R) &\geq F(R,\theta) \\ &= \sup_{0<s<1} \Bigg\{ \left(1 - \frac{1}{s}\right) R - \inf_{p: h_{\text{b}}(p) > R} \log \Bigg[\frac{(1-p)(1-\varepsilon)^{1+\theta-\theta/s} + p\varepsilon^{1+\theta-\theta/s}}{\left[(1-p)(1-\varepsilon)^{\theta} + p\varepsilon^{\theta}\right]^{(1-1/s)}} \\ &\qquad + \frac{(1-p)\varepsilon^{1+\theta-\theta/s} + p(1-\varepsilon)^{1+\theta-\theta/s}}{\left[(1-p)\varepsilon^{\theta} + p(1-\varepsilon)^{\theta}\right]^{(1-1/s)}} \Bigg] \Bigg\} \end{aligned}$$

for reals $\theta \geq 1$ and $0 < R < C = \log(2) - h_{\text{b}}(\varepsilon)$, where C is the channel capacity and $h_{\text{b}}(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ is the binary entropy function.

We compare $F(R,\theta)$ with the sphere packing upper bound to the BSC's reliability function (e.g., [9], [5]), which is denoted by $E_{\text{sp}}(R)$ and given by

$$E_{\text{sp}}(R) = \sup_{0<s\leq 1} \left\{ \left(1 - \frac{1}{s}\right)(R - \log 2) - \frac{1}{s} \log\left[(1-\varepsilon)^s + \varepsilon^s\right] \right\}$$

for $0 < R < C$. In Fig. 4, we plot $E_{\text{sp}}(R)$ and $F(R,\theta)$ for $\theta = 1$ and 2 and $\varepsilon = 0.01$. The figure indicates that for $\theta = 1$, $F(R,\theta) > E_{\text{sp}}(R)$ for all rates R. This directly implies that

$$E_{\text{PV}}(R) = E_{\text{PV}}^{(\theta=1)}(R) \geq F(R,\theta) > E_{\text{sp}}(R)$$

for all $0 < R < C$. Now recall that the sphere-packing upper bound $E_{\text{sp}}(R)$ is loose at low rates (for rates R less than the critical rate [9]) and tight (i.e., exactly equal to the channel reliability function $E^*(R)$) at high rates (rates between the critical rate and capacity). Thus for the BSC, the Poor-Verdú bound of (28) is not tight for all rates. Furthermore, note from the figure that since $F(R,\theta) < E_{\text{sp}}(R)$ for $\theta = 2$, we cannot make a conclusion regarding the tightness of $E_{\text{PV}}^{(\theta)}(R)$ in this case (this is also observed for $\theta > 2$).

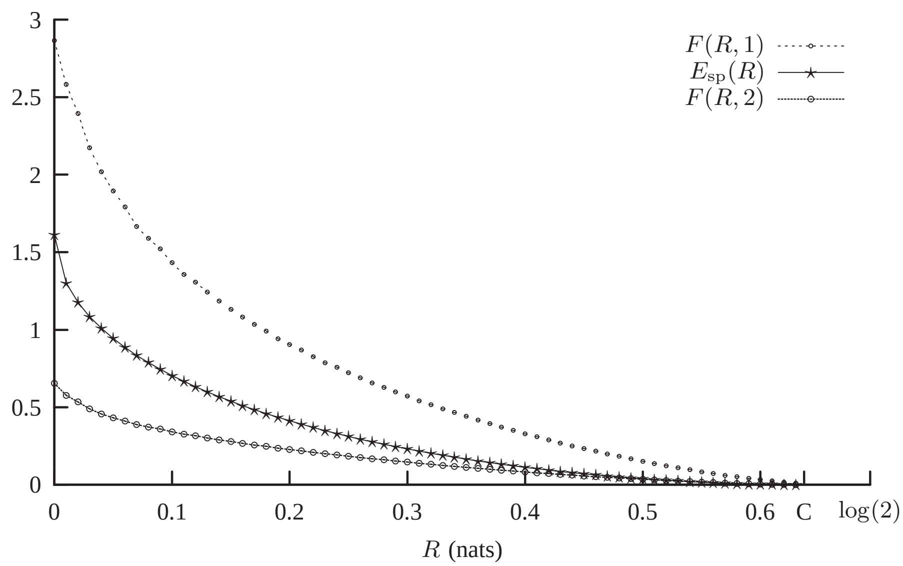

Fig. 4. BSC with crossover probability $\varepsilon = 0.01$: lower bound $F(R,\theta)$ to $E_{\text{PV}}^{(\theta)}(R)$ for $\theta = 1, 2$ and the sphere packing bound $E_{\text{sp}}(R)$.

2) Memoryless Z-Channel: We next consider the memoryless binary Z-channel described by $P_{Y|X}(0|1) = \varepsilon$ and $P_{Y|X}(0|0) = 1$. Again, setting $p \triangleq P_{\bar{X}}(1)$ and $s = \frac{1}{1-\rho}$ in (43) yields

$$\begin{aligned} E_{\text{PV}}^{(\theta)}(R) &\geq F(R,\theta) \\ &= \sup_{0<s<1} \left\{ \left(1 - \frac{1}{s}\right) R - \inf_{p: h_{\text{b}}(p) > R} \log \left[\frac{1 - p + p\varepsilon^{1+\theta-\theta/s}}{[1 - p + p\varepsilon^{\theta}]^{1-1/s}} + p^{1/s}(1-\varepsilon) \right] \right\} \end{aligned}$$

for $\theta \geq 1$ and $0 < R < C = \log\left(1 + (1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}\right)$. Furthermore, the channel's sphere packing upper bound is given by

$$E_{\text{sp}}(R) = \sup_{0<s\leq 1} \left\{ \left(1 - \frac{1}{s}\right) R - \inf_{0\leq p\leq 1} \log \left[(1 - p + p\varepsilon^{s})^{1/s} + p^{1/s}(1-\varepsilon) \right] \right\}$$

for $0 < R < C$. In Fig. 5, we plot $E_{\text{sp}}(R)$ and $F(R,\theta)$ for $\theta = 1, 3, 10, 100$ and $\varepsilon = 0.01$. We remark from the figure that for all considered values of θ (including θ very large not shown herein), $F(R,\theta) > E_{\text{sp}}(R)$ for high rates. This leads us to conclude that for the Z-channel, bound $E_{\text{PV}}^{(\theta)}(R)$ of (32) is not tight at high rates even when θ approaches infinity.

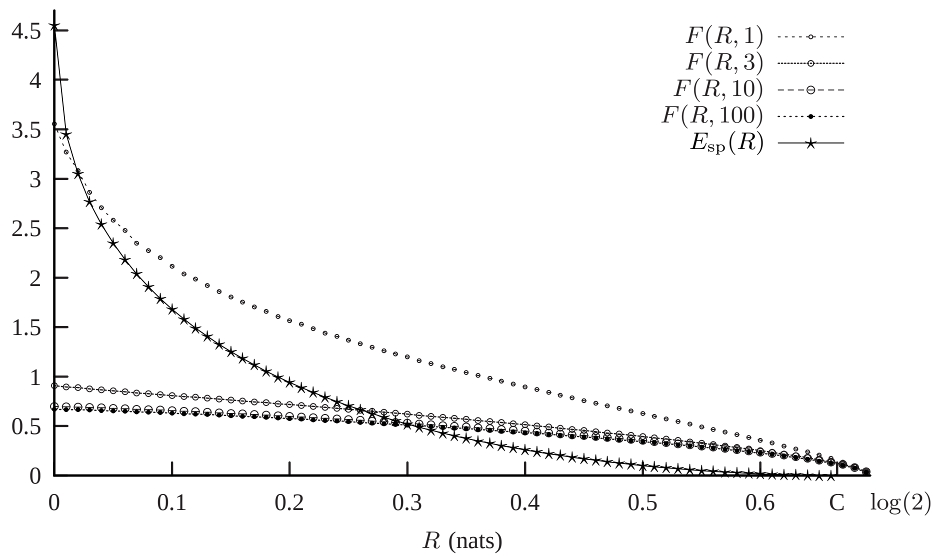

Fig. 5. Z-channel with crossover probability $\varepsilon = 0.01$: lower bound $F(R,\theta)$ to $E_{\text{PV}}^{(\theta)}(R)$ for $\theta = 1, 3, 10, 100$ and the sphere packing bound $E_{\text{sp}}(R)$.

Observation 8: It should be emphasized that the above numerical examples regarding the looseness of $E_{\text{PV}}^{(\theta)}(R)$ within a rate region and for given values of θ do not shed any light on the tightness of $\bar{E}_{\text{PV}}(R)$ given in (33), since the expression of $\bar{E}_{\text{PV}}(R)$ requires taking the limit with respect to θ *before* taking the limit with respect to the blocklength n.

V. Conclusion

In this work, we generalized the Poor-Verdú lower bound for the multihypothesis testing error probability. The new bound, which involves the tilted posterior distribution of the hypothesis given the observation with tilting parameter θ, reduces to the original Poor-Verdú bound when $\theta = 1$. We established a sufficient condition under which the bound (without its multiplicative factor) provides the exact error probability when $\theta \to \infty$. We also provided some examples to illustrate the tightness of the bound in terms of θ.

We next applied the new bound to obtain two new upper information-spectrum based bounds to the reliability function of general channels with memory, $E_{\text{PV}}^{(\theta)}(R)$ and $\bar{E}_{\text{PV}}(R)$, given in (32) and (33), respectively. It was shown that $\bar{E}_{\text{PV}}(R)$ is tight at all rates (below channel capacity) for a class of channels that include the finite-input memoryless Gaussian channel, hence providing an information-spectral characterization for these channels' reliability function. The determination of $\bar{E}_{\rm PV}(R)$ in closed form and its calculation remains a challenging problem (specially at low rates) as it involves taking the limit with respect to θ followed by optimizing the resulting large-deviation rate function over a constrained set of input processes (see (33)). It is anticipated that i.i.d. channel inputs are unlikely to be a valid optimizer for $\bar{E}_{\rm PV}(R)$. Although the evaluation of $\bar{E}_{\rm PV}(R)$ for non–i.i.d. channel inputs appears difficult, the judicious use of Markovian inputs might be worthwhile investigating in the future.

Appendix A

Lemma 1: If the limit (in n) of $a_{n,j}$ exists for every $j = 1, 2, 3, \ldots$, then

$$\lim_{n\to\infty}\sum_{j=1}^{\infty} a_{n,j} = \sum_{j=1}^{\infty}\lim_{n\to\infty} a_{n,j}.$$

Proof: Since for any sequences $\{b_n\}$ and $\{c_n\}$,

$$\liminf_{n\to\infty}(b_n + c_n) \geq \liminf_{n\to\infty} b_n + \liminf_{n\to\infty} c_n,$$

we recursively have that

$$\begin{aligned}\liminf_{n\to\infty}\sum_{j=1}^{\infty} a_{n,j} &\geq \liminf_{n\to\infty} a_{n,1} + \liminf_{n\to\infty}\sum_{j=2}^{\infty} a_{n,j}\\ &\geq \liminf_{n\to\infty} a_{n,1} + \liminf_{n\to\infty} a_{n,2} + \liminf_{n\to\infty}\sum_{j=3}^{\infty} a_{n,j}\\ &\geq \cdots\\ &\geq \sum_{j=1}^{\infty}\liminf_{n\to\infty} a_{n,j}.\end{aligned}$$

Similarly, since

$$\limsup_{n\to\infty}(b_n + c_n,) \leq \limsup_{n\to\infty} b_n + \limsup_{n\to\infty} c_n,$$

we obtain that

$$\limsup_{n\to\infty}\sum_{j=1}^{\infty} a_{n,j} \leq \sum_{j=1}^{\infty}\limsup_{n\to\infty} a_{n,j}.$$

Since

$$\limsup_{n\to\infty} a_{n,j} = \liminf_{n\to\infty} a_{n,j} = \lim_{n\to\infty} a_{n,j} \text{ for every } j,$$

we have

$$\sum_{j=1}^{\infty}\lim_{n\to\infty} a_{n,j} \geq \limsup_{n\to\infty}\sum_{j=1}^{\infty} a_{n,j} \geq \liminf_{n\to\infty}\sum_{j=1}^{\infty} a_{n,j} \geq \sum_{j=1}^{\infty}\lim_{n\to\infty} a_{n,j},$$

which immediately yields the desired result. ∎

Appendix B

We derive a lower bound to $E_{\rm PV}^{(\theta)}(R)$ given in (32), which can be numerically evaluated for different values of θ when the channel is memoryless.

Consider a general channel $\boldsymbol{W} = \{W^n\}_{n=1}^{\infty}$ with finite input alphabet $\mathcal{X}$ and arbitrary output alphabet $\mathcal{Y}$. Fix $R > 0$. Given an i.i.d. process $\bar{\boldsymbol{X}} = \{\bar{X}^n\}_{n=1}^{\infty}$ with alphabet $\mathcal{X}$ and entropy $H(\bar{X}) > R$ and a constant $0 < \delta < H(\bar{X}) - R$ arbitrarily small, define the (weakly) δ-typical set as:

$$\begin{aligned}\mathcal{F}_n(\delta|\bar{X}) &\triangleq \left\{x^n \in \mathcal{X}^n : \left|-\frac{1}{n}\log P_{\bar{X}^n}(x^n) - H(\bar{X})\right| \le \delta\right\} \\ &= \left\{x^n \in \mathcal{X}^n : \left|-\frac{1}{n}\sum_{i=1}^{n}\log P_{\bar{X}}(x_i) - H(\bar{X})\right| \le \delta\right\}.\end{aligned}$$

We now recall the consequence of the Asymptotic Equipartition Property for i.i.d. (memoryless) sources (e.g., see [5], [7]).

Proposition 2: Given an i.i.d. source $\{\bar{X}_n\}_{n=1}^{\infty}$ with entropy $H(\bar{X})$ and any δ greater than zero, then its δ-typical set $\mathcal{F}_n(\delta|\bar{X})$ satisfies the following.

1) If $x^n \in \mathcal{F}_n(\delta|\bar{X})$, then $e^{-n(H(\bar{X})+\delta)} \le P_{\bar{X}^n}(x^n) \le e^{-n(H(\bar{X})-\delta)}$.
2) $P_{\bar{X}^n}\left(\mathcal{F}_n^c(\delta|\bar{X})\right) < \delta$ for sufficiently large n, where the superscript "c" denotes the complement set operation.
3) $|\mathcal{F}_n(\delta|\bar{X})| > (1-\delta)e^{n(H(\bar{X})-\delta)}$ for sufficiently large n, and $|\mathcal{F}_n(\delta|\bar{X})| \le e^{n(H(\bar{X})+\delta)}$ for every n, where $|\mathcal{F}_n(\delta|\bar{X})|$ denotes the number of elements in $\mathcal{F}_n(\delta|\bar{X})$.

Let $\hat{\boldsymbol{X}} = \{\hat{X}^n\}_{n=1}^{\infty}$ be a process that is uniformly distributed over $\mathcal{F}_n(\delta|\bar{X})$ for each n; i.e., $P_{\hat{X}^n}(x^n) = \frac{1}{|\mathcal{F}(\delta|\bar{X})|}$ for $x^n \in \mathcal{F}_n(\delta|\bar{X})$ and $n = 1, 2, \cdots$. From Proposition 2, we also obtain that for n sufficiently large and $x^n \in \mathcal{F}_n(\delta|\bar{X})$,

$$(1-\delta)e^{-2n\delta} \le P_{\bar{X}^n}(x^n)|\mathcal{F}_n(\delta|\bar{X})| = \frac{P_{\bar{X}^n}(x^n)}{P_{\hat{X}^n}(x^n)} \le e^{2n\delta}. \tag{40}$$

For $\hat{\boldsymbol{X}}$ to belong to the set $\mathcal{Q}(R)$ as defined in (29), it is required that

$$\liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{S}(\hat{X}^n)| = \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{F}(\delta|\bar{X})| > R. \tag{41}$$

But condition (41) can be guaranteed by setting $H(\bar{X}) > R$ and taking $\delta < H(\bar{X}) - R$ (as already assumed) since

$$\liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{F}(\delta|\bar{X})| \ge \liminf_{n\to\infty}\frac{1}{n}\log(1-\delta)e^{n(H(\bar{X})-\delta)} = H(\bar{X}) - \delta > R,$$

where the inequality follows from property 1 of Proposition 2. Hence, such $\{\hat{X}^n\}_{n=1}^{\infty}$ process, uniformly distributed over its support, belongs to $\mathcal{Q}(R)$. Thus, we can lower-bound $E_{\text{PV}}^{(\theta)}(R)$ for channel $\boldsymbol{W} = \{W^n\}_{n=1}^{\infty}$ and a given $\theta > 1$ as follows

$$\begin{aligned} E_{\text{PV}}^{(\theta)}(R) &\triangleq \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n} j_{X^nW^n}^{(\theta)}(x^n;y^n)\le R\right\} \\ &\ge \liminf_{n\to\infty} -\frac{1}{n}\log P_{\hat{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n} j_{\hat{X}^nW^n}^{(\theta)}(x^n;y^n)\le R\right\}. \end{aligned}$$

For n sufficiently large, we can write

$$\begin{aligned} & j_{\hat{X}^nW^n}^{(\theta)}(x^n;y^n) \\ &= \log \frac{P_{Y^n|X^n}^{\theta}(y^n|x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n} P_{\hat{X}^n}(\hat{x}^n)P_{Y^n|X^n}^{\theta}(y^n|\hat{x}^n)} \\ &= \log \frac{P_{Y^n|X^n}^{\theta}(y^n|x^n)}{\sum_{\hat{x}^n\in\mathcal{F}_n(\delta|\bar{X})} P_{\hat{X}^n}(\hat{x}^n)P_{Y^n|X^n}^{\theta}(y^n|\hat{x}^n) + \sum_{\hat{x}^n\notin\mathcal{F}_n(\delta|\bar{X})} P_{\hat{X}^n}(\hat{x}^n)P_{Y^n|X^n}^{\theta}(y^n|\hat{x}^n)} \\ &= \log \frac{P_{Y^n|X^n}^{\theta}(y^n|x^n)}{\sum_{\hat{x}^n\in\mathcal{F}_n(\delta|\bar{X})} P_{\hat{X}^n}(\hat{x}^n)P_{Y^n|X^n}^{\theta}(y^n|\hat{x}^n)} \\ &\ge \log \frac{P_{Y^n|X^n}^{\theta}(y^n|x^n)}{\frac{e^{2n\delta}}{1-\delta}\sum_{\hat{x}^n\in\mathcal{F}_n(\delta|\bar{X})} P_{\bar{X}^n}(\hat{x}^n)P_{Y^n|X^n}^{\theta}(y^n|\hat{x}^n)} \\ &\ge \log \frac{(1-\delta)e^{-2n\delta}P_{Y^n|X^n}^{\theta}(y^n|x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n} P_{\bar{X}^n}(\hat{x}^n)P_{Y^n|X^n}^{\theta}(y^n|\hat{x}^n)} \\ &= \log(1-\delta) - 2n\delta + j_{\bar{X}^nW^n}^{(\theta)}(x^n;y^n), \end{aligned}$$

where the first inequality follows from the lower bound in (40). Accordingly,

$$\begin{aligned} E_{\text{PV}}^{(\theta)}(R) &\ge \liminf_{n\to\infty} -\frac{1}{n}\log P_{\hat{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n} j_{\hat{X}^nW^n}^{(\theta)}(x^n;y^n)\le R\right\} \\ &\ge \liminf_{n\to\infty} -\frac{1}{n}\log P_{\hat{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}\log(1-\delta) - 2\delta \right. \\ &\qquad\qquad \left. + \frac{1}{n} j_{\bar{X}^nW^n}^{(\theta)}(x^n;y^n)\le R\right\} \\ &= \liminf_{n\to\infty} -\frac{1}{n}\log P_{\hat{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n} j_{\bar{X}^nW^n}^{(\theta)}(x^n;y^n) \right. \\ &\qquad\qquad \left. \le R - \frac{1}{n}\log(1-\delta) + 2\delta\right\}. \end{aligned} \tag{42}$$

Observe that

$$P_{\hat{X}^nW^n}(x^n,y^n) = P_{\hat{X}^n}(x^n)P_{Y^n|X^n}(y^n|x^n) \le \frac{e^{2n\delta}}{1-\delta}P_{\bar{X}^n}(x^n)P_{Y^n|X^n}(y^n|x^n),$$

where the inequality follows from (40). Then, we can further lower-bound the right-hand side term of (42) to obtain

$$\begin{aligned} E_{\mathrm{PV}}^{(\theta)}(R) &\geq \liminf_{n\to\infty} -\frac{1}{n}\log\left(\frac{e^{2n\delta}}{1-\delta}P_{\bar{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \right.\right.\\ &\qquad\qquad \left.\left.\frac{1}{n}j^{(\theta)}_{\bar{X}^nW^n}(x^n;y^n)\leq R-\frac{1}{n}\log(1-\delta)+2\delta\right\}\right)\\ &\geq \liminf_{n\to\infty} -\frac{1}{n}\log P_{\bar{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}j^{(\theta)}_{\bar{X}^nW^n}(x^n;y^n)\leq R+\gamma\right\}-2\delta, \end{aligned}$$

where it suffices to take $\gamma > 2\delta$ to have $\gamma > -\frac{1}{n}\log(1-\delta)+2\delta$ for n sufficiently large.

In summary, we have shown that for any channel $\boldsymbol{W} = \{W^n\}_{n=1}^{\infty}$, the upper bound $E_{\mathrm{PV}}^{(\theta)}(R)$ to its channel reliability function satisfies

$$E_{\mathrm{PV}}^{(\theta)}(R) \geq \liminf_{n\to\infty} -\frac{1}{n}\log P_{\bar{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}j^{(\theta)}_{\bar{X}^nW^n}(x^n;y^n)\leq R+\gamma\right\}-2\delta$$

for $\theta \geq 1$ and any i.i.d. input process $\bar{X}$ with

$$\begin{cases} H(\bar{X}) > R \\ 0 < \delta < H(\bar{X}) - R \\ \gamma > 2\delta. \end{cases}$$

We next specialize the above lower bound for the case when channel $\boldsymbol{W}$ is memoryless. For a memoryless channel with an i.i.d. input, we have for $\rho < 0$,

$$\begin{aligned} &P_{\bar{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n}j^{(\theta)}_{\bar{X}^nW^n}(x^n;y^n)\leq R+\gamma\right\}\\ &= P_{\bar{X}^nW^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \rho\sum_{i=1}^{n}\log\frac{P^{\theta}_{Y|X}(y_i|x_i)}{\sum_{x'\in\mathcal{X}}P_{\bar{X}}(x')P^{\theta}_{Y|X}(y_i|x')}\geq n\rho(R+\gamma)\right\}\\ &\leq \left(e^{-\rho(R+\gamma)}\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}P_{\bar{X}}(x)P_{Y|X}(y|x)e^{\rho\log\frac{P^{\theta}_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}}P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')}}\right]\right)^n\\ &= \left(e^{-\rho(R+\gamma)}\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}P_{\bar{X}}(x)P_{Y|X}(y|x)\left(\frac{P^{\theta}_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}}P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')}\right)^{\rho}\right]\right)^n\\ &= \left(e^{-\rho(R+\gamma)}\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}P_{\bar{X}}(x)\frac{P^{1+\rho\theta}_{Y|X}(y|x)}{\left(\sum_{x'\in\mathcal{X}}P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')\right)^{\rho}}\right]\right)^n, \end{aligned}$$

where the inequality follows from Markov's inequality. Thus, for $\rho < 0$, we have

$$\begin{aligned} E_{\text{PV}}^{(\theta)}(R) &\geq \liminf_{n\to\infty} -\frac{1}{n}\log P_{\bar{X}^n W^n}\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \frac{1}{n} j^{(\theta)}_{\bar{X}^n W^n}(x^n;y^n)\leq R+\gamma\right\} - 2\delta \\ &\geq \liminf_{n\to\infty} -\frac{1}{n}\log\left(e^{-\rho(R+\gamma)}\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{\bar{X}}(x)\frac{P^{1+\rho\theta}_{Y|X}(y|x)}{\left(\sum_{x'\in\mathcal{X}} P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')\right)^{\rho}}\right]\right)^{n} - 2\delta \\ &\geq \liminf_{n\to\infty}\left(\rho(R+\gamma) - \log\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{\bar{X}}(x)\frac{P^{1+\rho\theta}_{Y|X}(y|x)}{\left(\sum_{x'\in\mathcal{X}} P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')\right)^{\rho}}\right]\right) - 2\delta \\ &= \rho(R+\gamma) - \log\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{\bar{X}}(x)\frac{P^{1+\rho\theta}_{Y|X}(y|x)}{\left(\sum_{x'\in\mathcal{Y}} P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')\right)^{\rho}}\right] - 2\delta. \end{aligned}$$

Since $\rho < 0$, γ should be made as small as possible. But as $\gamma > 2\delta$, it should thus approach 2δ to obtain

$$E_{\text{PV}}^{(\theta)}(R) \geq \rho R - \log\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{\bar{X}}(x)\frac{P^{1+\rho\theta}_{Y|X}(y|x)}{\left(\sum_{x'\in\mathcal{Y}} P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')\right)^{\rho}}\right] - 2(1-\rho)\delta.$$

Taking $\delta \downarrow 0$ yield the following lower bound to $E_{\text{PV}}^{(\theta)}(R)$ for a memoryless channel

$$\begin{aligned} E_{\text{PV}}^{(\theta)}(R) &\geq \sup_{P_{\bar{X}}:H(\bar{X})>R}\ \sup_{\rho<0}\left\{\rho R - \log\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{\bar{X}}(x)\frac{P^{1+\rho\theta}_{Y|X}(y|x)}{\left(\sum_{x'\in\mathcal{Y}} P_{\bar{X}}(x')P^{\theta}_{Y|X}(y|x')\right)^{\rho}}\right]\right\} \\ &\triangleq F(R,\theta) \end{aligned} \tag{43}$$

for $\theta \geq 1$.

REFERENCES

[1] F. Alajaji, P.-N. Chen and Z. Rached, "A note on the Poor-Verdú upper bound for the channel reliability function," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 309–313, Jan. 2002.

[2] A. Barg and A. McGregor, "Distance distribution of binary codes and the error probability of decoding," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4237–4246, Dec. 2005.

[3] Y. Ben Haim and S. Litsyn, "Improved upper bounds on the reliability function of the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 54., no. 1, pp. 5–12, Jan. 2008.

[4] P. Billingsley, *Probability and Measure*, Second Edition, Wiley, NY, 1986.

[5] R. Blahut, *Principles and Practice of Information Theory*, Addison Wesley, MA, 1988.

[6] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*, Wiley, NY, 1990.

[7] T. M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: Wiley, 2nd Ed., 2006.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, NY, 1981.

[9] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, NY, 1968.

[10] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer, 2003.

[11] J. H. Miller and J. B. Thomas, “Detectors for discrete-time signals in non-Gaussian noise,” *IEEE Trans. Inform. Theory*, vol. 18, no. 2, pp. 241–250, Mar. 1972.

[12] H. V. Poor and S. Verdú, “A lower bound on the probability of error in multi-hypothesis testing,” *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 1992–1994, Nov. 1995.

[13] C. E. Shannon, “Certain results in coding theory for noisy channels,” *Inform. Contr.*, vol. 1, pp. 6–25, Sep. 1957.

[14] S. Verdú and T. S. Han, “A general formula for channel capacity,” *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[15] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, NY, 1979.