The Asymptotic Generalized Poor-Verdú Bound Achieves the BSC Error Exponent at Zero Rate

Ling-Hua Chang*, Po-Ning Chen†, Fady Alajaji‡ and Yunghsiang S. Han§

Abstract—**The generalized Poor-Verdú error lower bound for multihypothesis testing is revisited. Its asymptotic expression is established in closed-form as its tilting parameter grows to infinity. It is also shown that the asymptotic generalized bound achieves the error exponent (or reliability function) of the memoryless binary symmetric channel at zero coding rates.**

Index Terms—**Binary symmetric channel, error probability bounds, error exponent, hypothesis testing, zero coding rates.**

I. Introduction

A well-known lower bound on the minimum probability of error P_e of multihypothesis testing is the so-called Poor-Verdú bound [1]. The bound was generalized in [2] by tilting, via a parameter $\theta \geq 1$, the posterior hypothesis distribution. The generalized bound was noted to progressively improve with θ; however its asymptotic formula as θ tends to infinity was not determined.

In this paper, we revisit this generalized bound and establish its asymptotic expression in closed-form. We then investigate the asymptotic generalized bound in the classical context of the error probability of block codes used over the memoryless binary symmetric channel (BSC) with crossover probability $p < \frac{1}{2}$. We prove that it is exponentially tight for arbitrary sequences of zero-rate codes and hence achieves the BSC zero-rate error exponent or reliability function (for in-depth studies of the channel reliability function, whose characterization at low rates remains a long-standing open problem, see [3]–[14] and the references therein).

In showing the exponential tightness of the asymptotic generalized Poor-Verdú bound, we first observe that when a code $\mathcal{C}_n$ with blocklength n and size $|\mathcal{C}_n| = M_n$ is transmitted over the BSC, this bound exactly equals the probability of the set $\mathtt{N}(\mathcal{C}_n)$, which consists of all input-output n-tuple pairs $(x^n, y^n) \in \mathcal{C}_n \times \mathcal{Y}^n$ satisfying

$$d(x^n, y^n) > \min_{u^n \in \mathcal{C}_n \setminus \{x^n\}} d(u^n, y^n), \tag{1}$$

where $d(\cdot,\cdot)$ is the Hamming distance and $\mathcal{Y}$ is the channel output alphabet (see Section III). By adding the probability of all ties, i.e., all $(x^n, y^n) \in \mathcal{C}_n \times \mathcal{Y}^n$ such that

$$d(x^n, y^n) = \min_{u^n \in \mathcal{C}_n \setminus \{x^n\}} d(u^n, y^n), \tag{2}$$

which are collected in the set $\mathtt{T}(\mathcal{C}_n)$, to $\Pr(\mathtt{N}(\mathcal{C}_n))$, an upper bound on the minimum probability of decoding error P_e is then obtained. The exponential tightness of $\Pr(\mathtt{N}(\mathcal{C}_n))$ to P_e can thus be confirmed by showing that $\Pr(\mathtt{T}(\mathcal{C}_n))$ has either the same error exponent as, or decreases exponentially faster than, $\Pr(\mathtt{N}(\mathcal{C}_n))$ for zero-rate codes. This property is demonstrated by constructing partitions of $\mathtt{T}(\mathcal{C}_n)$ and $\mathtt{N}(\mathcal{C}_n)$, denoted by $\{\mathcal{T}_i\}_{i=1}^{M_n}$ and $\{\mathcal{N}_i\}_{i=1}^{M_n}$, respectively, and then judiciously relating the probability of component set $\mathcal{T}_i$ to that of component set $\mathcal{N}_i$ for $i = 1, \ldots, M_n$. Specifically, we show that the probability of a finite cover of each $\mathcal{T}_i$, multiplied by $2M_n\frac{(1-p)}{p}$, is no larger than the probability of a subset of $\mathcal{N}_i$ (cf. Fig. 1 in Section III). With these key ingredients in place, we prove the exponential tightness of the asymptotic generalized Poor-Verdú bound at rate zero (i.e., when $\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{C}_n| = 0$).

The rest of the paper is organized as follows. In Section II, the exact expression of the asymptotic generalized Poor-Verdú lower bound on the error probability in multihypothesis testing is derived. In Section III, the error exponent analysis of this asymptotic bound is carried out in detail for the channel coding problem over the BSC. Finally conclusions are drawn in Section IV.

II. Asymptotic Expression of the Generalized Poor-Verdú Bound

In 1995, Poor and Verdú established a lower bound on the error probability of multihypothesis testing [1]. This bound was generalized in [2] in terms of a *tilted* posterior hypothesis distribution with tilting parameter $\theta \geq 1$ (with the original bound in [1] recovered when $\theta = 1$).

Lemma 1 (Generalized Poor-Verdú bound [2]): Consider random variables X and Y, governed by the joint distribution $P_{X,Y}$, and that take values in a discrete (i.e., finite or countably infinite) alphabet $\mathcal{X}$ and an arbitrary alphabet $\mathcal{Y}$, respectively. The minimum probability of error P_{e} in estimating X from Y satisfies

$$P_{\mathrm{e}} \geq (1-\alpha)\cdot P_{X,Y}\left\{(x,y) \in \mathcal{X}\times\mathcal{Y} : P^{(\theta)}_{X|Y}(x|y) \leq \alpha\right\} \tag{3}$$

for each $\alpha \in [0,1]$ and arbitrary $\theta \geq 1$, where

$$P^{(\theta)}_{X|Y}(x|y) \triangleq \frac{(P_{X|Y}(x|y))^\theta}{\sum_{u\in\mathcal{X}}(P_{X|Y}(u|y))^\theta} \tag{4}$$

is the tilted distribution of $P_{X|Y}(x|y)$ with parameter θ.

*Dept. Elec. Eng., Yuan-Ze Univ., Taiwan, R.O.C.
†Dept. Elec. and Comp. Eng., National Chiao Tung Univ., Taiwan, R.O.C.
‡Dept. Mathematics & Statistics, Queen's Univ., Kingston, ON, Canada.
§School of Elec. Eng. & Intelligentization, Dongguan Univ. of Tech., China.
{iamjaung, qponing}@gmail.com, fa@queensu.ca, yunghsiangh@gmail.com
This work was supported by the Ministry of Science and Technology of Taiwan (MOST 107-2218-E-155-012-MY2//108-2221-E-009-024-MY2), the Natural Sciences and Engineering Research Council of Canada and the National Natural Science Foundation of China (Grant No. 61671007).

It is illustrated via examples in [2] that the lower bound in (3) improves in general as θ grows. However, the asymptotic expression of (3), as θ goes to infinity, was not established in closed-form. This issue is resolved in what follows.

Lemma 2: For joint distribution $P_{X,Y}$ with marginal distribution P_X having finite support $\mathcal{C} \subseteq \mathcal{X}$, we have that for $\alpha < 1/|\mathcal{C}|$,

$$\limsup_{\theta\to\infty} P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} : P^{(\theta)}_{X|Y}(x|y)\le\alpha\right\}$$
$$= P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} : P_{X|Y}(x|y) < \max_{u\in\mathcal{C}} P_{X|Y}(u|y)\right\}. \tag{5}$$

Proof: Setting $\alpha = e^{-\kappa}$ in the right-hand side (RHS) probability term in (3) yields

$$P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} :\ P^{(\theta)}_{X|Y}(x|y)\le e^{-\kappa}\right\}$$
$$= P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} :\ \frac{(P_{X|Y}(x|y))^\theta}{\sum_{u\in\mathcal{C}}(P_{X|Y}(u|y))^\theta}\le e^{-\kappa}\right\}$$
$$= P_{X,Y}\Bigg\{(x,y)\in\mathcal{X}\times\mathcal{Y} : \log P_{X|Y}(x|y) \le \frac{1}{\theta}\log\left(\sum_{u\in\mathcal{C}}(P_{X|Y}(u|y))^\theta\right) - \frac{\kappa}{\theta}\Bigg\} \tag{6}$$
$$= P_{X,Y}\Bigg\{(x,y)\in\mathcal{X}\times\mathcal{Y} : \log P_{X|Y}(x|y)\le\log\|P_{X|Y}(\cdot|y)\|_\theta - \frac{\kappa}{\theta}\Bigg\}, \tag{7}$$

where $\|P_{X|Y}(\cdot|y)\|_\theta \triangleq \left(\sum_{u\in\mathcal{C}}(P_{X|Y}(u|y))^\theta\right)^{1/\theta}$ is the θ-norm of $P_{X|Y}(\cdot|y)$ for a fixed $y\in\mathcal{Y}$. Noting that

$$\log P_{X|Y}(x|y)\le\log\|P_{X|Y}(\cdot|y)\|_\theta - \frac{\kappa}{\theta}$$
$$\iff \frac{\kappa}{\theta}\le\log\|P_{X|Y}(\cdot|y)\|_\theta - \log P_{X|Y}(x|y), \tag{8}$$

we separately consider the following two cases.

Case 1: For (x,y) with $P_{X|Y}(x|y) < \max_{u\in\mathcal{C}} P_{X|Y}(u|y)$, the RHS of (8) tends to $\log\max_{u\in\mathcal{C}} P_{X|Y}(u|y) - \log P_{X|Y}(x|y)$ (which is strictly positive) as θ grows without bound (since the θ-norm of $P_{X|Y}(\cdot|y)$ approaches its infinity-norm given by $\max_{u\in\mathcal{C}} P_{X|Y}(u|y)$), while the left-hand side (LHS) of (8) tends to zero. Hence, (8) holds for θ sufficiently large.

Case 2: For (x,y) with $P_{X|Y}(x|y) = \max_{u\in\mathcal{C}} P_{X|Y}(u|y)$, we have

$$\log\|P_{X|Y}(\cdot|y)\|_\theta - \log P_{X|Y}(x|y)$$
$$= \log\left(\sum_{u\in\mathcal{C}}(P_{X|Y}(u|y))^\theta\right)^{1/\theta} - \log P_{X|Y}(x|y) \tag{9}$$
$$= \log\frac{\left(\sum_{u\in\mathcal{C}}(P_{X|Y}(u|y))^\theta\right)^{1/\theta}}{\left((P_{X|Y}(x|y))^\theta\right)^{1/\theta}} \tag{10}$$
$$= \frac{1}{\theta}\log\frac{\sum_{u\in\mathcal{C}}(P_{X|Y}(u|y))^\theta}{(P_{X|Y}(x|y))^\theta} \tag{11}$$
$$\le \frac{1}{\theta}\log|\mathcal{C}|, \tag{12}$$

where the last inequality holds since $P_{X|Y}(u|y)\le P_{X|Y}(x|y)$ for all $u\in\mathcal{C}$. Thus (8) is violated since $\kappa = -\log\alpha > \log|\mathcal{C}|$.

Verifying the above two cases completes the proof. ■

In light of Lemma 2, we can fix $\kappa = -\log\alpha > \log|\mathcal{C}|$, take θ to infinity, and obtain from (3) that

$$P_{\mathrm{e}} \ge (1-e^{-\kappa})$$
$$P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} : P_{X|Y}(x|y) < \max_{u\in\mathcal{C}} P_{X|Y}(u|y)\right\}. \tag{13}$$

Since (13) holds for $\kappa > \log|\mathcal{C}|$ arbitrarily large, we have the following asymptotic expression of the generalized Poor-Verdú bound.

Corollary 1: The minimum error probability P_{e} in estimating X from Y satisfies

$$P_{\mathrm{e}} \ge P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} : P_{X|Y}(x|y) < \max_{u\in\mathcal{C}} P_{X|Y}(u|y)\right\}. \tag{14}$$

Two remarks are made based on Corollary 1. First, the optimal estimate of X from observing Y is known to be the maximum *a posteriori* estimate, given by

$$e(y) = \arg\max_{x\in\mathcal{C}} P_{X|Y}(x|y), \tag{15}$$

where (15) can in fact directly yield the lower bound in (14). This indicates that tilting the *a posteriori* distribution in the generalized Poor-Verdú bound can indeed approach[1]

$$1 - P_{X,Y}\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} : P_{X|Y}(x|y) = P_{X|Y}(e(y)|y)\right\}. \tag{16}$$

As a consequence, the lower bound in (14) is tight if and only if the x that maximizes $P_{X|Y}(x|y)$ is unique for all $y\in\mathcal{Y}$. This elucidates why in the example of [2, Fig. 1] the generalized Poor-Verdú bound achieves the minimum probability of error P_{e} when θ grows unbounded.

Second, an alternative lower bound for P_e is the Verdú-Han bound established in [15]. This bound was recently generalized in [16, Thm. 1]. We remark that the Verdú-Han bound is not tight even if $P_{X|Y}(x|y)$ admits a unique maximizer for every $y\in\mathcal{Y}$. For example, we can obtain from the ternary hypothesis testing example in [2, Sec. III-A] and [16, Sec. III-A] that:

$$P_e = \frac{3}{5} > \max_{\gamma\ge 0}\left(\Pr\left[P_{X|Y}(X|Y)\le\gamma\right] - \gamma\right) = \frac{27}{47}, \tag{17}$$

where the maximizer in (17) is $\gamma^* = \frac{20}{47}$. Noting the sub-optimality of the Verdú-Han bound, the authors in [16] generalized it by varying the output statistics. They also proved the tightness of the resulting generalized Verdú-Han bound:

$$P_e = \max_{Q_Y}\max_{\gamma\ge 0}\left(\Pr\left[\frac{P_{X,Y}(X,Y)}{Q_Y(Y)}\le\gamma\right] - \gamma\right). \tag{18}$$

[1]Note that the set $\{(x,y)\in\mathcal{X}\times\mathcal{Y} : P_{X|Y}(x|y) = P_{X|Y}(e(y)|y)\}$ includes all ties. For example, for the 2-fold BSC (i.e., the BSC used twice to transmit 2-tuple inputs) with uniform P_X over $\mathcal{C} = \{00, 11\}$, both $(00, 01)$ and $(11, 01)$ will be in this set, i.e.,

$$\{(x,y)\in\mathcal{X}\times\mathcal{Y} : P_{X|Y}(x|y) = P_{X|Y}(e(y)|y)\}$$
$$= \{(00,00),(00,01),(11,01),(00,10),(11,10),(11,11)\}.$$

It is pertinent to note that the maximizers of (18) are given by

$$\gamma^* = \int_{\mathcal{Y}} \max_{x\in\mathcal{X}} P_{X,Y}(x,y)\,\mathrm{d}P_Y(y) = 1 - P_e \tag{19}$$

and

$$Q_Y^*(y) = \frac{\max_{x\in\mathcal{X}} P_{X,Y}(x,y)}{\int_{\mathcal{Y}} \max_{x\in\mathcal{X}} P_{X,Y}(x,y)\,\mathrm{d}P_Y(y)} \tag{20}$$

$$= \frac{P_Y(y)P_{X|Y}(e(y)|y)}{1-P_e}. \tag{21}$$

Hence, the determination of the maximizers of the above generalized Verdú-Han bound is equivalent to determining the minimum error probability P_e itself.

Similar to the generalized Poor-Verdú bound with parameter θ, any Q_Y and γ adopted for the generalized Verdú-Han bound yields a lower bound on P_e. However, an interesting difference between the generalized Poor-Verdú bound and the generalized Verdú-Han bound is that when P_X is uniformly distributed over its support $\mathcal{C}$, the former bound can be transformed into a function of the *information density*

$$i_{XW}(x,y) \triangleq \frac{P_{Y|X}(y|x)}{P_Y(y)} \tag{22}$$

of the channel $W = P_{Y|X}$ with input X and output Y, while the latter bound cannot. This transformation may facilitate the interpretation of the error exponent via the information density (or equivalently, the Hamming distance) for memoryless symmetric channels such as the BSC.

III. Exponential Tightness of the Asymptotic Generalized Poor-Verdú Bound for the BSC at Zero Rate

In this section, we prove that the asymptotic expression of the generalized Poor-Verdú bound given in (14) exactly characterizes the zero-rate coding error exponent of the BSC with crossover probability $p < \frac{1}{2}$. Note that while the error exponent formula for the BSC at zero-rate, $E(0)$, is already known, $E(0) = -\frac{1}{4}\ln\big(4p(1-p)\big)$ [6], we do not explicitly calculate it. Rather, we demonstrate that the bound in (14) is exponentially tight for arbitrary sequences of zero-rate block codes used over the BSC, hence indirectly achieving $E(0)$. This approach may be beneficial for a larger class of channels.

Fix a sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ of blocklength n, with $\mathcal{C}_n \subseteq \{0,1\}^n$, and let P_{X^n} be the uniform distribution over $\mathcal{C}_n$, where X^n denotes the n-tuple $(X_1,\ldots,X_n)$. Denote by

$$a_n \triangleq P_{\mathrm{e}}(\mathcal{C}_n) \tag{23}$$

the minimum probability of decoding error for transmitting code $\mathcal{C}_n$ over the BSC with crossover probability $p < \frac{1}{2}$, and let b_n denote the RHS of (14) in this channel coding context:

$$b_n \triangleq P_{X^n,Y^n}\Big\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : P_{X^n|Y^n}(x^n|y^n) < \max_{u^n\in\mathcal{C}_n} P_{X^n|Y^n}(u^n|y^n)\Big\}. \tag{24}$$

Since the BSC has $p < \frac{1}{2}$, the inequality condition in (24) can be equivalently characterized via the Hamming distance $d(\cdot,\cdot)$. Hence,

$$b_n = P_{X^n,Y^n}(\mathtt{N}(\mathcal{C}_n)), \tag{25}$$

where

$$\mathtt{N}(\mathcal{C}_n) \triangleq \Big\{(x^n,y^n)\in\mathcal{C}_n\times\mathcal{Y}^n : d(x^n,y^n) > \min_{u^n\in\mathcal{C}_n\setminus\{x^n\}} d(u^n,y^n)\Big\}. \tag{26}$$

Define the *set of ties* with respect to code $\mathcal{C}_n$ as

$$\mathtt{T}(\mathcal{C}_n) \triangleq \Big\{(x^n,y^n)\in\mathcal{C}_n\times\mathcal{Y}^n : d(x^n,y^n) = \min_{u^n\in\mathcal{C}_n\setminus\{x^n\}} d(u^n,y^n)\Big\}, \tag{27}$$

and let

$$\delta_n = P_{X^n,Y^n}(\mathtt{T}(\mathcal{C}_n)). \tag{28}$$

Then the minimum probability of error $a_n = P_{\mathrm{e}}(\mathcal{C}_n)$ satisfies

$$b_n \le a_n \le b_n + \delta_n, \tag{29}$$

where the upper bound is achieved when we have decoding errors for $(x^n,y^n)\in\mathtt{T}(\mathcal{C}_n)$. Note that (29) implies that

$$0 \le \frac{1}{n}\log\frac{a_n}{b_n} \le \frac{1}{n}\log\left(1+\frac{\delta_n}{b_n}\right). \tag{30}$$

As a result, in order to prove that a_n and b_n have the same error exponent, it suffices to prove that

$$\limsup_{n\to\infty}\frac{1}{n}\log\left(1+\frac{\delta_n}{b_n}\right) = 0. \tag{31}$$

We next establish the following main theorem, which confirms (31) at zero rates (in Corollary 2 below).

Theorem 1: For any sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$, we have

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{a_n}{b_n} \le \limsup_{n\to\infty}\frac{1}{n}\log M_n, \tag{32}$$

where $M_n = |\mathcal{C}_n|$.

Before giving the proof, we elucidate the underlying idea behind it. We first introduce the following necessary notation. For block code $\mathcal{C}_n = \{x^n_{(1)}, x^n_{(2)}, \ldots, x^n_{(M_n)}\}$ consisting of M_n distinct codewords, define the sets

$$\mathcal{T}_i \triangleq \{y^n\in\mathcal{Y}^n : (x^n_{(i)},y^n)\in\mathtt{T}(\mathcal{C}_n)\} \tag{33}$$

for $i = 1,\ldots,M_n$. We can then write

$$\delta_n = P_{X^n,Y^n}(\mathtt{T}(\mathcal{C}_n)) \tag{34}$$

$$= \sum_{i=1}^{M_n} P_{X^n}(x^n_{(i)})\Pr\left(Y^n\in\mathcal{T}_i\,\middle|\,X^n = x^n_{(i)}\right). \tag{35}$$

Similarly, defining the sets

$$\mathcal{N}_i \triangleq \{y^n\in\mathcal{Y}^n : (x^n_{(i)},y^n)\in\mathtt{N}(\mathcal{C}_n)\} \tag{36}$$

for $i = 1,\ldots,M_n$, we have

$$b_n = P_{X^n,Y^n}(\mathtt{N}(\mathcal{C}_n)) \tag{37}$$

$$= \sum_{i=1}^{M_n} P_{X^n}(x^n_{(i)})\Pr\left(Y^n\in\mathcal{N}_i\,\middle|\,X^n = x^n_{(i)}\right). \tag{38}$$

Finally for $i,j = 1,\ldots,M_n$ with $i\ne j$, define

$$\mathcal{B}_{i,j} \triangleq \left\{y^n\in\mathcal{Y}^n : d(x^n_{(i)},y^n) = d(x^n_{(j)},y^n)\right\}, \tag{39}$$

and

$$\Omega_{i,j} \triangleq \left\{y^n\in\mathcal{Y}^n : d(x^n_{(i)},y^n) > d(x^n_{(j)},y^n)\right\}. \tag{40}$$

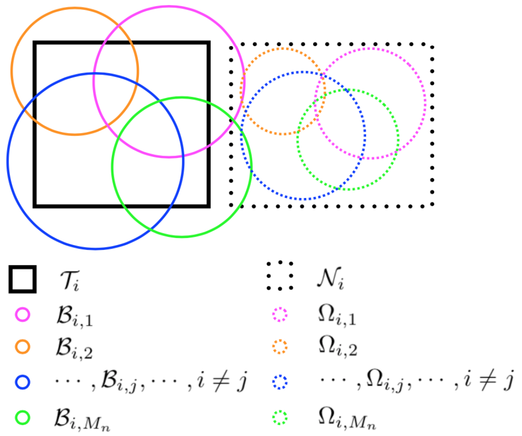

Fig. 1. Illustration of the idea behind the proof of Theorem 1.

Then as shown in Fig. 1, we have that $\cup_{j=1,j\neq i}^{M_n}\mathcal{B}_{i,j}$ is a finite cover of $\mathcal{T}_i$, i.e.,

$$\mathcal{T}_i \subseteq \cup_{j=1,j\neq i}^{M_n}\mathcal{B}_{i,j}. \tag{41}$$

Hence,

$$\begin{aligned}
&\Pr\left(Y^n \in \mathcal{T}_i | X^n = x_{(i)}^n\right) \\
&\leq \Pr\left(Y^n \in \bigcup_{j=1,j\neq i}^{M_n} \mathcal{B}_{i,j} \,\middle|\, X^n = x_{(i)}^n\right) && (42)\\
&\leq \sum_{j=1,j\neq i}^{M_n} \Pr\left(Y^n \in \mathcal{B}_{i,j} | X^n = x_{(i)}^n\right) && (43)\\
&\leq (M_n-1) \max_{1\leq j\leq M_n, j\neq i} \Pr\left(Y^n \in \mathcal{B}_{i,j} | X^n = x_{(i)}^n\right) && (44)\\
&= (M_n-1) \Pr\left(Y^n \in \mathcal{B}_{i,j_i^*} | X^n = x_{(i)}^n\right), && (45)
\end{aligned}$$

where the second inequality follows from the union bound and j_i^* is the maximizer of (44). Next, noting that

$$\Omega_{i,j} \subseteq \mathcal{N}_i \tag{46}$$

for all $1 \leq j \leq M_n$ and $j \neq i$, we have

$$\Pr\left(Y^n \in \Omega_{i,j_i^*} | X^n = x_{(i)}^n\right) \leq \Pr\left(Y^n \in \mathcal{N}_i | X^n = x_{(i)}^n\right). \tag{47}$$

Thus, if $\Pr\left(Y^n \in \mathcal{B}_{i,j_i^*} | X^n = x_{(i)}^n\right)$ and $\Pr\left(Y^n \in \Omega_{i,j_i^*} | X^n = x_{(i)}^n\right)$ are of comparable order in the sense that

$$\begin{aligned}
&\Pr\left(Y^n \in \Omega_{i,j_i^*} \,\middle|\, X^n = x_{(i)}^n\right) \\
&\geq c \cdot \Pr\left(Y^n \in \mathcal{B}_{i,j_i^*} \,\middle|\, X^n = x_{(i)}^n\right) && (48)
\end{aligned}$$

for some constant c independent of n and i, then we have

$$\Pr\left(Y^n \in \mathcal{N}_i \,\middle|\, X^n = x_{(i)}^n\right) \geq \frac{c}{M_n} \Pr\left(Y^n \in \mathcal{T}_i \,\middle|\, X^n = x_{(i)}^n\right) \tag{49}$$

which immediately gives

$$b_n \geq \frac{c}{M_n}\delta_n \tag{50}$$

and confirms (32). With this idea in mind, we next provide the detailed proof.

Proof of Theorem 1: We present the proof in four steps.

Step 1: First, we calculate $\Pr\left(Y^n \in \mathcal{B}_{i,j} | x_{(i)}^n\right)$. For each $x_{(i)}^n$ and $x_{(j)}^n$, if $d(x_{(i)}^n, x_{(j)}^n) = 2\ell \geq 2$ is even, then there are $\binom{2\ell}{\ell}\binom{n-2\ell}{m}$ of y^n's such that $d(x_{(i)}^n, y^n) = d(x_{(j)}^n, y^n) = \ell + m$ for $0 \leq m \leq n-2\ell$; else if $d(x_{(i)}^n, x_{(j)}^n) = 2\ell - 1$ is odd, then there exist no y^n such that $d(x_{(i)}^n, y^n) = d(x_{(j)}^n, y^n)$. As a result, we have that

$$\begin{aligned}
&\Pr\left(Y^n \in \mathcal{B}_{i,j} \,\middle|\, x_{(i)}^n\right) \\
&= \begin{cases} \sum_{m=0}^{n-2\ell} \binom{2\ell}{\ell}\binom{n-2\ell}{m}(1-p)^{n-\ell-m}p^{\ell+m}, \\ \qquad\qquad \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell; \\ 0, \qquad\qquad \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell - 1 \end{cases} && (51)\\
&= \begin{cases} \binom{2\ell}{\ell}p^\ell(1-p)^\ell, & \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell; \\ 0, & \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell - 1. \end{cases} && (52)
\end{aligned}$$

Step 2: We next lower-bound $\Pr\left(Y^n \in \Omega_{i,j} | x_{(i)}^n\right)$ in terms of $\Pr\left(Y^n \in \mathcal{B}_{i,j} | x_{(i)}^n\right)$.

If $d(x_{(i)}^n, x_{(j)}^n) = 2\ell$ is even, there are

$$\sum_{\ell'=0}^{\min\{m,\ell-1\}} \binom{2\ell}{\ell+\ell'+1}\binom{n-2\ell}{m-\ell'} \tag{53}$$

of y^n's satisfying $d(x_{(i)}^n, y^n) = \ell + 1 + m$ and $d(x_{(i)}^n, y^n) > d(x_{(j)}^n, y^n)$ for $0 \leq m \leq n - 2\ell$; else if $d(x_{(i)}^n, x_{(j)}^n) = 2\ell - 1$ is odd, then there are

$$\sum_{\ell'=0}^{\min\{m,\ell\}} \binom{2\ell-1}{\ell+\ell'+1}\binom{n-2\ell+1}{m-\ell'} \tag{54}$$

of y^n's satisfying $d(x_{(i)}^n, y^n) = \ell + 1 + m$ and $d(x_{(i)}^n, y^n) > d(x_{(j)}^n, y^n)$ for $0 \leq m \leq n - 2\ell + 1$.

Taking $\ell' = 0$ in (53) and (54) gives a lower bound on $\Pr\left(Y^n \in \Omega_{i,j} | x_{(i)}^n\right)$ as follows:

$$\begin{aligned}
&\Pr\left(Y^n \in \Omega_{i,j} \,\middle|\, X^n = x_{(i)}^n\right) \\
&\geq \begin{cases} \sum_{m=0}^{n-2\ell} \binom{2\ell}{\ell+1}\binom{n-2\ell}{m}(1-p)^{n-\ell-1-m}p^{\ell+1+m}, \\ \qquad\qquad \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell \\ \sum_{m=0}^{n-2\ell+1} \binom{2\ell-1}{\ell+1}\binom{n-2\ell+1}{m}(1-p)^{n-\ell-1-m}p^{\ell+1+m}, \\ \qquad\qquad \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell - 1 \end{cases} && (55)\\
&= \begin{cases} \binom{2\ell}{\ell+1}p^{\ell+1}(1-p)^{\ell-1}, & \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell \\ \binom{2\ell-1}{\ell+1}p^{\ell+1}(1-p)^{\ell-2}, & \text{if } d(x_{(i)}^n, x_{(j)}^n) = 2\ell - 1 \end{cases} && (56)\\
&\geq \frac{\ell}{(\ell+1)}\frac{p}{(1-p)} \Pr\left(Y^n \in \mathcal{B}_{i,j} \,\middle|\, X^n = x_{(i)}^n\right) && (57)\\
&\geq \frac{p}{2(1-p)} \Pr\left(Y^n \in \mathcal{B}_{i,j} \,\middle|\, X^n = x_{(i)}^n\right), && (58)
\end{aligned}$$

where (57) follows from (52) and (58) holds since $\ell \geq 1$.

Step 3: We next can write

$$\mathcal{T}_i = \left\{ y^n \in \mathcal{Y}^n : d(x^n_{(i)}, y^n) = \min_{u^n \in \mathcal{C}_n \setminus \{x^n_{(i)}\}} d(u^n, y^n) \right\} \tag{59}$$

$$\subseteq \bigcup_{j=1, j\neq i}^{M_n} \left\{ y^n \in \mathcal{Y}^n : d(x^n_{(i)}, y^n) = d(x^n_{(j)}, y^n) \right\} \tag{60}$$

$$= \bigcup_{j=1, j\neq i}^{M_n} \mathcal{B}_{i,j}, \tag{61}$$

which implies, as already shown in (44), that

$$\Pr\left(Y^n \in \mathcal{T}_i \middle| X^n = x^n_{(i)}\right) \leq (M_n - 1) \Pr\left(Y^n \in \mathcal{B}_{i,j_i^*} \middle| X^n = x^n_{(i)}\right), \tag{62}$$

where j_i^* is the maximizer in (44). Therefore with this j_i^*, we have that

$$\Pr\left(Y^n \in \mathcal{N}_i \middle| X^n = x^n_{(i)}\right)$$

$$\geq \Pr\left(Y^n \in \Omega_{i,j_i^*} \middle| X^n = x^n_{(i)}\right) \tag{63}$$

$$\geq \frac{p}{2(1-p)} \Pr\left(Y^n \in \mathcal{B}_{i,j_i^*} \middle| X^n = x^n_{(i)}\right) \tag{64}$$

$$\geq \frac{p}{2(1-p)} \frac{1}{(M_n - 1)} \Pr\left(Y^n \in \mathcal{T}_i \middle| X^n = x^n_{(i)}\right), \tag{65}$$

where (64) follows from (58), and (65) is based on (62).

Step 4: We conclude from (65) that

$$b_n = \sum_{i=1}^{M_n} P_{X^n}(x^n_{(i)}) \Pr\left(Y^n \in \mathcal{N}_i \middle| X^n = x^n_{(i)}\right) \tag{66}$$

$$\geq \frac{p}{2(1-p)} \frac{1}{(M_n - 1)} \cdot \sum_{i=1}^{M_n} P_{X^n}(x^n_{(i)}) \Pr\left(Y^n \in \mathcal{T}_i \middle| X^n = x^n_{(i)}\right) \tag{67}$$

$$= \frac{p}{2(1-p)} \frac{1}{(M_n - 1)} \delta_n, \tag{68}$$

which implies that

$$\limsup_{n\to\infty} \frac{1}{n} \log\left(1 + \frac{\delta_n}{b_n}\right)$$

$$\leq \limsup_{n\to\infty} \frac{1}{n} \log\left(1 + \frac{2(1-p)}{p}(M_n - 1)\right) \tag{69}$$

$$= \limsup_{n\to\infty} \frac{1}{n} \log(M_n), \tag{70}$$

where the last step holds whether either M_n is bounded or unbounded. ∎

Finally, we directly obtain that (31) holds when the (asymptotic) rate of the code sequence considered in Theorem 1 is zero, hence confirming the exponential tightness of the asymptotic generalized Poor-Verdú bound for the BSC at rate zero.

Corollary 2: For any sequence of zero-rate codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ used over the BSC, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} \leq \limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{C}_n| = 0.$$

Remark 1: It is worth emphasizing that Corollary 2 does *not* hold for the memoryless binary erasure channel (BEC); i.e., the asymptotic generalized Poor-Verdú bound is not exponentially tight for this channel. Indeed for the BEC, the bound in (3) is unchanged for every $\theta \geq 1$ (including when $\theta \to \infty$) and is hence identical to the original Poor-Verdú bound. The latter bound was shown in [17] not to achieve the BEC's error exponent at low rates.

IV. Conclusion

We derived a closed-form formula for the asymptotic generalized Poor-Verdú error bound to the multihypothesis testing error probability and proved that, unlike the case for the BEC [17], it achieves the zero-rate error coding exponent of the BSC.

The proof of Theorem 1 relies largely on the analysis of $\delta_n = \Pr(\mathtt{T}(\mathcal{C}_n))$, which is the probability of ties among competing pairs of codewords and received words. In particular, we used the union bound for estimating the probability of ties given $X^n = x^n_{(i)}$ in (62), which may be loose when the sequence of codes is no longer of zero rate. Thus, if a sharper bound can be employed, the multiplicative factor $\frac{p}{2(1-p)} \frac{1}{(M_n-1)}$ in (68) may be improved. We conjecture that Corollary 2 holds not just for zero-rate codes but that it can be indeed extended to arbitrary code sequences of positive rate. Proving this conjecture is an interesting future direction.

Other future work includes studying the relationship between the generalized Poor-Verdú bound and the meta-converse channel coding bound [18], [19][2] and the further examination of tight bounds for codes with small blocklength (e.g., see [16], [18], [20]) used over channels with and without memory.

References

[1] H. V. Poor and S. Verdú, "A lower bound on the probability of error in multihypothesis testing," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1992-1994, Nov. 1995.

[2] P.-N. Chen and F. Alajaji, "A generalized Poor-Verdú error bound for multihypothesis testings," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 311-316, Jan. 2012.

[3] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3-18, Jan. 1965.

[4] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels - I," *Inf. Contr.*, vol. 10, pp. 65-103, Jan. 1967.

[5] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels - II," *Inf. Contr.*, vol. 10, pp. 522-552, May 1967.

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

[7] R. J. McEliece and J. K. Omura, "An improved upper bound on the block coding error exponent for binary-input discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 23, no. 5, pp. 611-613, Sep. 1977.

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

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

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

[11] S. Litsyn, "New upper bounds on error exponents," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 385-398, March 1999.

[2] Note that it is shown in [19] that the original Poor-Verdú bound [1] actually coincides with the sphere-packing bound for discrete memoryless channels.

[12] A. Barg and A. McGregor, "Distance distribution of binary codes and the error probability of decoding," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4237-4246, Dec. 2005.
[13] M. Dalai, "Lower bounds on the probability of error for classical and classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8027-8056, Dec. 2013.
[14] M. V. Burnashev, "On the BSC reliability function: Expanding the region where it is known exactly," *Probl. Inf. Trans.*, vol. 51, no. 4, pp. 307-325, 2015.
[15] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147-1157, July 1994.
[16] G. Vazquez-Vilar, A. T. Campo, A. G. i Fábregas, and A. Martinez, "Bayesian M-ary hypothesis testing: The meta-converse and Verdú-Han bounds are tight," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2324-2333, May 2016.
[17] F. Alajaji, P.-N. Chen, and Z. Rached, "A note on the Poor-Verdú conjecture for the channel reliability function," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 309-313, Jan. 2002.
[18] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307-2359, May 2010.
[19] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576-2595, May 2013.
[20] P.-N. Chen, H.-Y. Lin, and S. M. Moser, "Optimal ultrasmall block-codes for binary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7346-7378, Nov. 2013.