Bayesian M-ary Hypothesis Testing: The Meta-Converse and Verdú-Han Bounds are Tight

Gonzalo Vazquez-Vilar, *Member, IEEE,* Adrià Tauste Campo, *Member, IEEE,*
Albert Guillén i Fàbregas, *Senior Member, IEEE,* Alfonso Martinez, *Senior Member, IEEE*

Abstract

Two alternative exact characterizations of the minimum error probability of Bayesian M-ary hypothesis testing are derived. The first expression corresponds to the error probability of an induced binary hypothesis test and implies the tightness of the meta-converse bound by Polyanskiy, Poor and Verdú; the second expression is function of an information-spectrum measure and implies the tightness of a generalized Verdú-Han lower bound. The formulas characterize the minimum error probability of several problems in information theory and help to identify the steps where existing converse bounds are loose.

Index Terms

Hypothesis testing, meta-converse, information spectrum, channel coding, Shannon theory.

I. Introduction

Statistical hypothesis testing appears in areas as diverse as information theory, image processing, signal processing, social sciences or biology. Depending on the field, this problem can be referred to as classification, discrimination, signal detection or model selection. The goal of M-ary hypothesis testing is to decide among M possible hypotheses based on the observation of a certain random variable. In a Bayesian formulation, a prior distribution over the hypotheses is assumed, and the problem is translated into a minimization of the average error probability or its

G. Vazquez-Vilar was with the Department of Information and Communica- tion Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain. He is now with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, Leganés 28911, Spain (e-mail: gvazquez@ieee.org).

A. Tauste Campo and A. Martinez are with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (e-mail: adria.tauste@upf.edu; alfonso.martinez@ ieee.org). A. Tauste Campo is also with the Hospital del Mar Medical Research Institute, Barcelona 08003, Spain.

A. Guillén i Fàbregas is with the Institució Catalana de Recerca i Estudis Avançats (ICREA), with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme under grant agreements 303633 and 329837 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150, TEC2012-38800-C03-03 and FPDI-2013-18602.

This work was presented in part at the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, July 7–12, 2013.

generalization, the Bayes risk. When the number of hypotheses is $M = 2$, the problem is referred to as binary hypothesis testing. While a Bayesian approach in this case is still possible, the binary setting allows a simple formulation in terms of the two types of pairwise errors with no prior distribution over the hypotheses. The work of Neyman and Pearson [1] established the optimum binary test in this setting. Thanks to its simplicity and robustness, this has been the most popular approach in the literature.

In the context of reliable communication, binary hypothesis testing has been instrumental in the derivation of converse bounds to the error probability. In [2, Sec. III] Shannon, Gallager and Berlekamp derived lower bounds to the error probability in the transmission of M messages, including the sphere-packing bound, by analyzing an instance of binary hypothesis testing [2], [3]. In [4], Forney used a binary hypothesis test to determine the optimum decision regions in decoding with erasures. In [5], Blahut emphasized the fundamental role of binary hypothesis testing in information theory and provided an alternative derivation of the sphere-packing exponent. Inspired by this result, Omura presented in [6] a general method for lower-bounding the error probability of channel coding and source coding. More recently, Polyanskiy, Poor and Verdú [7] applied the Neyman-Pearson lemma to a particular binary hypothesis test to derive the meta-converse bound, a fundamental finite-length lower bound to the channel-coding error probability from which several converse bounds can be recovered. The meta-converse bound was extended to joint source-channel coding in [8], [9].

The information-spectrum method expresses the error probability as the tail probability of a certain random variable, often referred to as information density, entropy density or information random variable [10]. This idea was initially used by Shannon in [11] to obtain bounds to the channel coding error probability. Verdú and Han capitalized on this analysis to provide error bounds and capacity expressions that hold for general channels, including arbitrary memory, input and output alphabets [12]–[14] (see also [10]).

In this work, we further develop the connection between hypothesis testing, information-spectrum and converse bounds in information theory by providing a number of alternative expressions for the error probability of Bayesian M-ary hypothesis testing. We show that this probability can be equivalently described by the error probability of a binary hypothesis test with certain parameters. In particular, this result implies that the meta-converse bound by Polyanskiy, Poor and Verdú gives the minimum error probability when it is optimized over its free parameters. We also provide an explicit alternative expression using information-spectrum measures and illustrate the connection with existing information-spectrum bounds. This result implies that a suitably optimized generalization of the Verdú-Han bound also gives the minimum error probability. We discuss in some detail examples and extensions.

The rest of this paper is organized as follows. In Section II of this paper we formalize the binary hypothesis testing problem and introduce notation. In Section III we present M-ary hypothesis testing and propose a number of alternative expressions to the average error probability. The hypothesis-testing framework is related to several previous converse results in Section IV. Proofs of several results are included in the appendices.

II. Binary Hypothesis Testing

Let Y be a random variable taking values over a discrete alphabet $\mathcal{Y}$. We define two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, such that Y is distributed according to a given distribution P under $\mathcal{H}_0$, and according to a distribution Q under $\mathcal{H}_1$. A binary hypothesis test is a mapping $\mathcal{Y} \to \{0,1\}$, where 0 and 1 correspond respectively to $\mathcal{H}_0$ and $\mathcal{H}_1$. Denoting by $\hat{H} \in \{0,1\}$ the random variable associated with the test output, we may describe the (possibly randomized) test by a conditional distribution $T \triangleq P_{\hat{H}|Y}$.

The performance of a binary hypothesis test is characterized by two conditional error probabilities, namely $\epsilon_0(P,T)$ or type-0 probability, and $\epsilon_1(P,T)$ or type-1 probability, respectively given by

$$\epsilon_0(P,T) \triangleq \Pr\bigl[\hat{H} = 1 \,\big|\, \mathcal{H}_0\bigr] = \sum_y P(y) T(1|y), \tag{1}$$

$$\epsilon_1(Q,T) \triangleq \Pr\bigl[\hat{H} = 0 \,\big|\, \mathcal{H}_1\bigr] = \sum_y Q(y) T(0|y). \tag{2}$$

In the Bayesian setting, for $\mathcal{H}_i$ with prior probability $\Pr[\mathcal{H}_i]$, $i = 0,1$, the smallest average error probability is

$$\bar{\epsilon} \triangleq \min_T \Bigl\{ \Pr[\mathcal{H}_0]\, \epsilon_0(P,T) + \Pr[\mathcal{H}_1]\, \epsilon_1(Q,T) \Bigr\}. \tag{3}$$

In the non-Bayesian setting, the priors $\Pr[\mathcal{H}_i]$, $i = 0,1$, are unknown and the quantity $\bar{\epsilon}$ is not defined. Instead, one can characterize the optimal trade-off between $\epsilon_0(\cdot)$ and $\epsilon_1(\cdot)$. We define the smallest type-0 error $\epsilon_0(\cdot)$ among all tests T with a type-1 error $\epsilon_1(\cdot)$ at most β as

$$\alpha_\beta\bigl(P,Q\bigr) \triangleq \min_{T:\epsilon_1(Q,T)\le\beta} \Bigl\{ \epsilon_0(P,T) \Bigr\}. \tag{4}$$

The tests minimizing (3) and (4) have the same form. The minimum is attained by the Neyman-Pearson test [1],

$$T_{\text{NP}}(0|y) = \begin{cases} 1, & \text{if } \frac{P(y)}{Q(y)} > \gamma, \\ p, & \text{if } \frac{P(y)}{Q(y)} = \gamma, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where $\gamma \ge 0$ and $p \in [0,1]$ are parameters. When $\gamma = \frac{\Pr[\mathcal{H}_1]}{\Pr[\mathcal{H}_0]}$, the test T_{NP} minimizes (3) with the value of p being irrelevant since it does not affect the objective. When γ and p are chosen such that the type-1 error $\epsilon_1(Q, T_{\text{NP}})$ is equal to β, T_{NP} attains the minimum in (4). The test minimizing (3) and (4) is not unique in general, as the form of the test can vary for observations y satisfying $P(y) = Q(y)$. Any test achieving (4) is said to be optimal in the Neyman-Pearson sense.

III. M-ary Hypothesis Testing

Consider two random variables V and Y with joint distribution P_{VY}, where V takes values on a discrete alphabet $\mathcal{V}$ of cardinality $|\mathcal{V}| = M$, and Y takes values in a discrete alphabet $\mathcal{Y}$. We shall assume that the cardinality $|\mathcal{V}|$ is finite; see Remark 1 in Section III-B for an extension to infinite alphabets $\mathcal{V}$. While throughout the article we use discrete notation for clarity of exposition, the results directly generalize to continuous alphabets $\mathcal{Y}$; see Remark 2 in Section III-B.

The estimation of V given Y is an M-ary hypothesis-testing problem. Since the joint distribution P_{VY} defines a prior distribution P_V over the alternatives, the problem is naturally cast within the Bayesian framework.

An M-ary hypothesis test is defined by a (possibly random) transformation $P_{\hat{V}|Y}: \mathcal{Y} \to \mathcal{V}$, where $\hat{V}$ denotes the random variable associated to the test output.[1] We denote the average error probability of a test $P_{\hat{V}|Y}$ by $\bar{\epsilon}(P_{\hat{V}|Y})$. This probability is given by

$$\bar{\epsilon}(P_{\hat{V}|Y}) \triangleq \Pr\left[\hat{V} \neq V\right] \tag{6}$$

$$= 1 - \sum_{v,y} P_{VY}(v,y) P_{\hat{V}|Y}(v|y). \tag{7}$$

Minimizing over all possible conditional distributions $P_{\hat{V}|Y}$ gives the smallest average error probability, namely

$$\bar{\epsilon} \triangleq \min_{P_{\hat{V}|Y}} \bar{\epsilon}(P_{\hat{V}|Y}). \tag{8}$$

An optimum test chooses the hypothesis v with largest posterior probability $P_{V|Y}(v|y)$ given the observation y, that is the Maximum a Posteriori (MAP) test. The MAP test that breaks ties randomly with equal probability is given by

$$P^{\text{MAP}}_{\hat{V}|Y}(v|y) = \begin{cases} \frac{1}{|\mathcal{S}(y)|}, & \text{if } v \in \mathcal{S}(y), \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

where the set $\mathcal{S}(y)$ is defined as

$$\mathcal{S}(y) \triangleq \left\{ v \in \mathcal{V} \;\middle|\; P_{V|Y}(v|y) = \max_{v' \in \mathcal{V}} P_{V|Y}(v'|y) \right\}. \tag{10}$$

Substituting (9) in (7) gives

$$\bar{\epsilon} = 1 - \sum_{v,y} P_{VY}(v,y) P^{\text{MAP}}_{\hat{V}|Y}(v|y) \tag{11}$$

$$= 1 - \sum_{y} \max_{v'} P_{VY}(v',y). \tag{12}$$

The next theorem introduces two alternative equivalent expressions for the minimum error probability $\bar{\epsilon}$.

Theorem 1: The minimum error probability of an M-ary hypothesis test (with possibly non-equally likely hypotheses) can be expressed as

$$\bar{\epsilon} = \max_{Q_Y} \alpha_{\frac{1}{M}}\left(P_{VY}, Q_V \times Q_Y\right) \tag{13}$$

$$= \max_{Q_Y} \sup_{\gamma \geq 0} \left\{ \Pr\left[\frac{P_{VY}(V,Y)}{Q_Y(Y)} \leq \gamma \right] - \gamma \right\}, \tag{14}$$

where $Q_V(v) \triangleq \frac{1}{M}$ for all $v \in \mathcal{V}$, and the probability in (14) is computed with respect to P_{VY}. Moreover, a maximizing distribution Q_Y in both expressions is

$$Q^{\star}_Y(y) \triangleq \frac{1}{\mu} \max_{v'} P_{VY}(v',y), \tag{15}$$

[1] While both binary and M-ary hypothesis tests are defined by conditional distributions, to avoid confusion, we denote binary tests by T and M-ary tests by $P_{\hat{V}|Y}$.

where $\mu \triangleq \sum_y \max_{v'} P_{VY}(v', y)$ is a normalizing constant.

Proof: See Section III-B. ■

Eq. (13) in Theorem 1 shows that the error probability of Bayesian M-ary hypothesis testing can be expressed as the best type-0 error probability of an induced binary hypothesis test discriminating between the original distribution P_{VY} and an alternative product distribution $Q_V \times Q_Y^{\star}$ with type-1-error equal to $\frac{1}{M}$. Eq. (14) in Theorem 1 provides an alternative characterization based on information-spectrum measures, namely the generalized information density $\log \frac{P_{VY}(v,y)}{Q_Y(y)}$. By choosing $Q_Y = Q_Y^{\star}$ and $\gamma = \mu$, the term $\Pr\left[\frac{P_{VY}(V,Y)}{Q_Y(Y)} \leq \gamma\right] - \gamma$ can be interpreted as the error probability of an M-ary hypothesis test that, for each v, compares the posterior likelihood $P_{V|Y}(v|y)$ with a threshold equal to $\max_{v'} P_{V|Y}(v'|y)$ and decides accordingly, i. e., this test emulates the MAP test yielding the exact error probability. The two alternative expressions provided in Theorem 1 are not easier to compute than $\bar{\epsilon}$ in (12). To see this, note that the normalization factor μ in $Q_Y^{\star}$ is such that $\mu = 1 - \bar{\epsilon}$.

For any fixed test $P_{\hat{V}|Y}$, not necessarily MAP, using (8) it follows that $\bar{\epsilon}(P_{\hat{V}|Y}) \geq \bar{\epsilon}$. Therefore, Theorem 1 provides a lower bound to the error probability of any M-ary hypothesis test. This bound is expressed in (13) as a binary hypothesis test discriminating between P_{VY} and an auxiliary distribution $Q_{VY} = Q_V \times Q_Y$. Optimizing over general distributions Q_{VY} (not necessarily product) may yield tighter bounds for a fixed test $P_{\hat{V}|Y}$, as shown next.

Theorem 2: The error probability of an M-ary hypothesis test $P_{\hat{V}|Y}$ satisfies

$$\bar{\epsilon}(P_{\hat{V}|Y}) = \max_{Q_{VY}} \alpha_{\epsilon_1(Q_{VY}, P_{\hat{V}|Y})}\big(P_{VY}, Q_{VY}\big) \tag{16}$$

$$= \max_{Q_{VY}} \sup_{\gamma \geq 0} \left\{ \Pr\left[\frac{P_{VY}(V,Y)}{Q_{VY}(V,Y)} \leq \gamma\right] - \gamma \epsilon_1(Q_{VY}, P_{\hat{V}|Y}) \right\}, \tag{17}$$

where

$$\epsilon_1(Q_{VY}, P_{\hat{V}|Y}) \triangleq \sum_{v,y} Q_{VY}(v,y) P_{\hat{V}|Y}(v|y). \tag{18}$$

Proof: Let us consider the binary test $T(0|v,y) = P_{\hat{V}|Y}(v|y)$. The type-0 and type-1 error probabilities of this test are $\epsilon_0(P_{VY}, T) = \bar{\epsilon}(P_{\hat{V}|Y})$ and $\epsilon_1(Q_{VY}, T) = \epsilon_1(Q_{VY}, P_{\hat{V}|Y})$ defined in (18), respectively. Therefore, from the definition of $\alpha_{(\cdot)}(\cdot)$ in (4) we obtain that, for any Q_{VY},

$$\bar{\epsilon}(P_{\hat{V}|Y}) \geq \alpha_{\epsilon_1(Q_{VY}, P_{\hat{V}|Y})}\big(P_{VY}, Q_{VY}\big). \tag{19}$$

For $Q_{VY} = P_{VY}$, using that $\alpha_\beta(P_{VY}, P_{VY}) = 1 - \beta$, the right-hand side of (19) becomes $1 - \epsilon_1(P_{VY}, P_{\hat{V}|Y})$. As $1 - \epsilon_1(P_{VY}, P_{\hat{V}|Y}) = 1 - \epsilon_1(P_{VY}, T) = \epsilon_0(P_{VY}, T) = \bar{\epsilon}(P_{\hat{V}|Y})$, then (16) follows from optimizing (19) over Q_{VY}. To obtain (17) we apply the lower bound in Lemma 1 in Section III-B to (16) and note that, for $\gamma = 1$, $Q_{VY} = P_{VY}$, the bound holds with equality. ■

The proof of Theorem 2 shows that the auxiliary distribution $Q_{VY} = P_{VY}$ maximizes (16) and (17) for any M-ary hypothesis test $P_{\hat{V}|Y}$. Nevertheless, the auxiliary distribution optimizing (16) and (17) is is not unique in general, as seen in Theorem 1 for the MAP test and in the next result for arbitrary maximum-metric tests.

Consider the maximum-metric test $P^{(q)}_{\hat{V}|Y}$ that chooses the hypothesis v with largest metric $q(v,y)$, where $q(v,y)$ is an arbitrary function of v and y. This test can be equivalently described as

$$P^{(q)}_{\hat{V}|Y}(v|y) = \begin{cases} \frac{1}{|\mathcal{S}_q(y)|}, & \text{if } v \in \mathcal{S}_q(y), \\ 0, & \text{otherwise,} \end{cases} \tag{20}$$

where the set $\mathcal{S}_q(y)$ is defined as

$$\mathcal{S}_q(y) \triangleq \left\{ v \in \mathcal{V} \;\middle|\; q(v,y) = \max_{v' \in \mathcal{V}} q(v',y) \right\}. \tag{21}$$

Corollary 1: For the maximum metric test $P_{\hat{V}|Y} = P^{(q)}_{\hat{V}|Y}$, a distribution Q_{VY} maximizing (16) and (17) is

$$Q^{(q)}_{VY}(v,y) \triangleq \frac{P_{VY}(v,y)}{\mu'} \frac{\max_{v'} q(v',y)}{q(v,y)}, \tag{22}$$

where μ' is a normalizing constant.

Proof: See Appendix A. ■

The expressions in Theorem 2 still depend on the specific test through $\epsilon_1(\cdot)$, cf. (18). For the optimal MAP test, i. e., a maximum metric test with metric $q(v,y) = P_{V|Y}(v|y)$, we obtain $Q^{(q)}_{VY} = Q_V \times Q^\star_Y$ with uniform Q_V and $Q^\star_Y$ defined in (15). For uniform Q_V it holds that

$$\epsilon_1(Q_V \times Q_Y, P_{\hat{V}|Y}) = \frac{1}{M}, \tag{23}$$

for any Q_Y, $P_{\hat{V}|Y}$. As a result, for the optimal MAP test, the expressions in Theorem 2 and the distribution defined in Corollary 1 recover those in Theorem 1.

A. Example

To show the computation of the various expressions in Theorem 1 let us consider the ternary hypothesis test examined in [14, Figs. 1 and 2] and revisited in [15, Sec. III.A]. Let $\mathcal{V} = \mathcal{Y} = \{0,1,2\}$, $P_V(v) = \frac{1}{3}$, $v = 0,1,2$, and

$$P_{Y|V}(y|v) = \begin{cases} 0.40, & (v,y) = (0,0), (1,1) \text{ and } (2,2), \\ 0.33, & (v,y) = (0,2), (1,2) \text{ and } (2,0), \\ 0.27, & \text{otherwise.} \end{cases} \tag{24}$$

Direct calculation shows that the MAP estimate is $\hat{v}(y) = y$, and from (12) we obtain $\bar{\epsilon} = 0.6$.

In order to evaluate the expressions in Theorem 1 we first compute $Q^\star_Y$ in (15), which yields $Q^\star_Y(y) = \frac{1}{3}$, $y = 0,1,2$. According to (13) a binary hypothesis test between P_{VY} and $Q^\star_{VY}$, where $Q^\star_{VY}(v,y) = \frac{1}{9}$, for all v,y, with type-1 error $\epsilon_1 = \frac{1}{3}$, yields the minimum error probability

$$\bar{\epsilon} = \alpha_{\frac{1}{3}}\big(P_{VY}, Q^\star_{VY}\big). \tag{25}$$

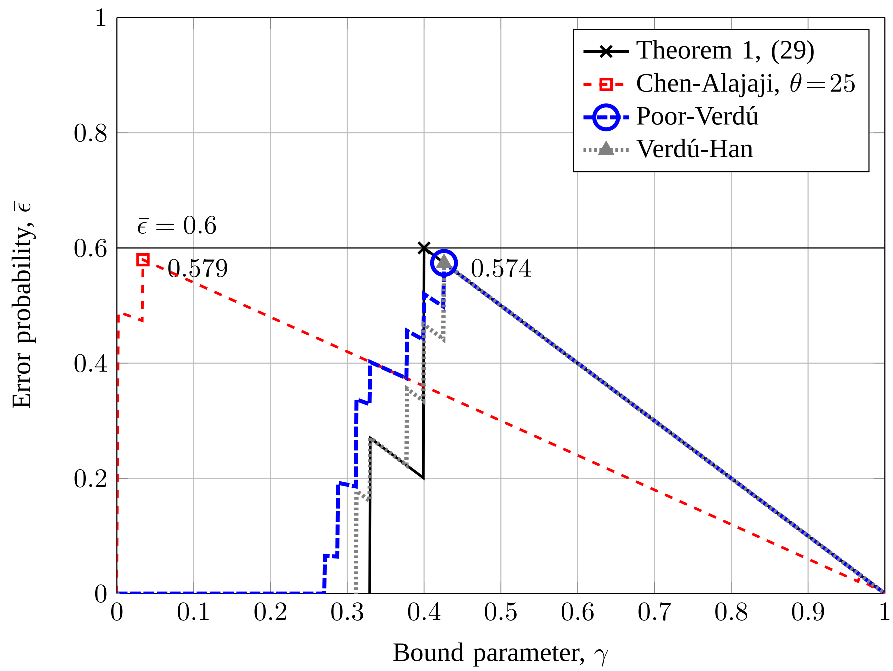

Figure 1. Information-spectrum lower bounds to the minimum error probability for the example in Section III-A, as a function of the bound parameter γ.

Solving the Neyman-Pearson test in (5) for the type-1 error $\epsilon_1 = \frac{1}{3}$, we obtain $\gamma = 1.2$ and $p = 1$ and therefore

$$T_{\text{NP}}(0|y) = \begin{cases} 1, & \text{if } P_{VY}(v,y) \geq \frac{2}{15}, \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

Hence, (25) yields

$$\bar{\epsilon} = \epsilon_0(P_{VY}, T_{\text{NP}}) \tag{27}$$

$$= 1 - \sum_{v,y} P_{VY}(v,y) T_{\text{NP}}(0|y) = 0.6. \tag{28}$$

Similarly, to evaluate (14) in Theorem 1, we substitute $Q_Y^{\star}$ to obtain

$$\bar{\epsilon} = \sup_{\gamma \geq 0} \left\{ \Pr\left[P_{VY}(V,Y) \leq \frac{\gamma}{3} \right] - \gamma \right\}. \tag{29}$$

Fig. 1 shows the argument of (29) with respect to $\gamma \in [0,1]$ compared to the exact error probability $\bar{\epsilon}$, shown in the plot with an horizontal line. For comparison, we also include the Verdú-Han lower bound [13, Th. 4], the Poor-Verdú lower bound [14, Th. 1] and the lower bound proposed by Chen and Alajaji in [15, Th. 1]. The Chen-Alajaji bound [15, Th. 1] is parametrized by $\theta \geq 0$ and, for $\theta = 1$, it reduces to the Poor-Verdú lower bound. We observe that (29) gives the exact error probability $\bar{\epsilon} = 0.6$ at $\gamma = 1 - \bar{\epsilon}$. The Verdú-Han and the Poor-Verdú lower bounds both coincide and yield $\bar{\epsilon} \geq 0.574$. For this example, as shown in [15], the Chen-Alajaji lower bound is tight for $\theta \to \infty$. For $\theta = 25$ the bound is still $\bar{\epsilon} \geq 0.579$.

As an application of Theorem 2 and Corollary 1 we study now a variation of the previous example. For a hypothesis $v \in \mathcal{V}$, let $(y_1, y_2) \in \mathcal{Y}^2$ denote two independent observations of the random variable Y distributed according to $P_{Y|V=v}$ in (24). We consider the suboptimal hypothesis test that decides on the source message v maximizing the metric $q(v, y_1, y_2) = P_{Y|V}(y_1|v)$. That is, for equiprobable hypotheses, this test applies the MAP rule based on the first observation, ignoring the second one. The expressions in Theorem 1 do not depend on the decoder and yield the MAP error probability $\bar{\epsilon} = 0.592$. Then, for $P^{(q)}_{\hat{V}|Y_1Y_2}$ in (20), it holds that $\bar{\epsilon}\big(P^{(q)}_{\hat{V}|Y_1Y_2}\big) \geq 0.592$.

Let us choose the auxiliary distribution

$$Q_{VY_1Y_2}(v, y_1, y_2) = \frac{1}{9} P_{Y|V}(y_2|v). \tag{30}$$

Using that $P^{(q)}_{\hat{V}|Y_1Y_2}(v|y_1, y_2) = \mathbb{1}\big\{v = y_1\big\}$ is independent of y_2, we obtain

$$\epsilon_1\big(Q_{VY_1Y_2}, P^{(q)}_{\hat{V}|Y_1Y_2}\big) = \frac{1}{9} \sum_{v,y_1,y_2} P_{Y|V}(y_2|v) P^{(q)}_{\hat{V}|Y_1Y_2}(v|y_1, y_2) \tag{31}$$

$$= \frac{1}{9} \sum_{v,y_1} \mathbb{1}\big\{v = y_1\big\} \tag{32}$$

$$= \frac{1}{3}. \tag{33}$$

Therefore, the bound implied in Theorem 2 for this specific choice of $Q_{VY_1Y_2}$ yields

$$\bar{\epsilon}\Big(P^{(q)}_{\hat{V}|Y_1Y_2}\Big) \geq \alpha_{\frac{1}{3}}\big(P_{VY_1Y_2}, Q_{VY_1Y_2}\big). \tag{34}$$

Since the marginal corresponding to Y_2 is the same for $P_{VY_1Y_2}$ and $Q_{VY_1Y_2}$ in (30), this component does not affect to the binary test and can be eliminated from (34). Therefore, the right-hand side in (34) coincides with that of (25), and yields the lower bound $\bar{\epsilon}\big(P^{(q)}_{\hat{V}|Y_1Y_2}\big) \geq 0.6$. It can be checked that an application of (17) in Theorem 2 yields the same result. We conclude that allowing joint distributions $Q_{VY_1Y_2}$ we obtain decoder-specific bounds.

B. *Proof of Theorem 1*

We first prove the equality between the left- and right-hand sides of (13) by showing the equivalence of the optimization problems (8) and (13). From (8) we have that

$$\bar{\epsilon} = \min_{P_{\hat{V}|Y} : \sum_v P_{\hat{V}|Y}(v|y) \leq 1, y \in \mathcal{Y}} \sum_{v,y} P_{VY}(v, y) \Big(1 - P_{\hat{V}|Y}(v|y)\Big) \tag{35}$$

$$= \max_{\lambda(\cdot) \geq 0} \min_{P_{\hat{V}|Y}} \left\{ \sum_{v,y} P_{VY}(v, y) \left(1 - P_{\hat{V}|Y}(v|y)\right) + \sum_y \lambda(y) \left(\sum_v P_{\hat{V}|Y}(v|y) - 1 \right) \right\}, \tag{36}$$

where in (35) we wrote explicitly the (active) constraints resulting from $P_{\hat{V}|Y}$ being a conditional distribution; and (36) follows from introducing the constraints into the objective via the Lagrange multipliers $\lambda(y) \geq 0$, $y \in \mathcal{Y}$.

Similarly, we write (13) as

$$\max_{Q_Y} \alpha_{\frac{1}{M}} \left(P_{VY}, Q_V \times Q_Y\right)$$
$$= \max_{Q_Y} \min_{T: \sum_{v,y} \frac{1}{M} Q_Y(y) T(0|v,y) \leq \frac{1}{M}} \left\{ \sum_{v,y} P_{VY}(v,y) T(1|v,y) \right\} \tag{37}$$
$$= \max_{\eta \geq 0} \max_{Q_Y} \min_{T} \left\{ \sum_{v,y} P_{VY}(v,y) \Big(1 - T(0|v,y)\Big) + \eta \left(\sum_{v,y} Q_Y(y) T(0|v,y) - 1 \right) \right\}, \tag{38}$$

where in (37) we used the definitions of Q_V and $\alpha_\beta(\cdot)$; and (38) follows from introducing the constraint into the objective via the Lagrange multiplier η.

Since η and Q_Y only appear in the objective function of (38) as $\eta Q_Y(y)$, $y \in \mathcal{Y}$, we may optimize (38) over $\bar{\lambda}(y) \triangleq \eta Q_Y(y)$ instead. Then, (38) becomes

$$\max_{\bar{\lambda}(\cdot) \geq 0} \min_{T} \left\{ \sum_{v,y} P_{VY}(v,y) \Big(1 - T(0|v,y)\Big) + \sum_{y} \bar{\lambda}(y) \left(\sum_{v} T(0|v,y) - 1 \right) \right\}. \tag{39}$$

Comparing (36) and (39), it is readily seen that the optimization problems (8) and (13) are equivalent. Hence, the first part of the theorem follows.

We need the following result to prove identity (14).

Lemma 1: For any pair of distributions $\{P, Q\}$ over $\mathcal{Y}$ and any $\gamma' \geq 0$, it holds

$$\alpha_\beta\big(P, Q\big) \geq \mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma'\right] - \gamma' \beta. \tag{40}$$

Proof: The bound (40) with the term $\mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma'\right]$ replaced by $\mathbb{P}\left[\frac{P(Y)}{Q(Y)} < \gamma'\right]$ corresponds to [7, Eq. (102)]. The proof of the lemma follows the steps in [16, Eq. (2.71)-(2.74)] and is included in Appendix B for completeness.

■

Applying (40) to (13) with $\gamma' = \gamma M$, $P \leftarrow P_{VY}$ and $Q \leftarrow Q_V \times Q_Y$ and optimizing over γ we obtain

$$\bar{\epsilon} \geq \max_{Q_Y} \sup_{\gamma \geq 0} \left\{ \Pr\left[\frac{P_{VY}(V,Y)}{Q_Y(Y)} \leq \gamma\right] - \gamma \right\}. \tag{41}$$

By using the distribution $Q_Y = Q_Y^\star$ in (15) and by choosing $\gamma = \mu$, the probability term in (41) becomes

$$\Pr\left[\frac{P_{VY}(V,Y)}{Q_Y^\star(Y)} \leq \mu\right] = \Pr\Big[P_{V|Y}(V|Y) \leq \max_{v'} P_{V|Y}(v'|Y)\Big] = 1. \tag{42}$$

Substituting $Q_Y = Q_Y^\star$, $\gamma = \mu$, and using (42) in (41) we obtain

$$\bar{\epsilon} \geq \max_{Q_Y} \sup_{\gamma \geq 0} \left\{ \Pr\left[\frac{P_{VY}(V,Y)}{Q_Y(Y)} \leq \gamma\right] - \gamma \right\} \tag{43}$$
$$\geq 1 - \mu \tag{44}$$
$$= 1 - \sum_{y} \max_{v'} P_{VY}(v', y) \tag{45}$$
$$= \bar{\epsilon}, \tag{46}$$

where in (45) we used the definition of μ and (46) follows from (12). The identity (14) in the theorem is due to (43)-(46), where it is readily seen that $Q_Y = Q_Y^\star$ is a maximizer of (14). Moreover, since $Q_Y^\star$ is a maximizer of

(14), and Lemma 1 applies for a fixed Q_Y, it follows that $Q_Y^\star$ is also an optimal solution to (13). The second part of the theorem thus follows from (43)-(46).

Remark 1: A simple modification of Theorem 1 generalizes the result to countably infinite alphabets $\mathcal{V}$. We define $\bar{Q}_V$ to be the counting measure, i. e., $\bar{Q}_V(v) = 1$ for all v. The function $\alpha_\beta(\cdot)$ in (4) is defined for arbitrary σ-finite measures, not necessarily probabilities. Then, by substituting Q_V by $\bar{Q}_V$, the type-1 error measure is $\epsilon_1(\bar{Q}_V \times Q_Y, T) = 1$ for any T, and (13) becomes

$$\bar{\epsilon} = \max_{Q_Y} \alpha_1 \left(P_{VY}, \bar{Q}_V \times Q_Y \right). \tag{47}$$

Since (14) directly applies to both finite or countably infinite $\mathcal{V}$, so does Theorem 1 with (13) replaced by (47).

Remark 2: For continuous observation alphabets $\mathcal{Y}$, the constraint of $P_{\hat{V}|Y}$ being a conditional distribution

$$\sum_v P_{\hat{V}|Y}(v|y) \le 1, \; y \in \mathcal{Y}, \tag{48}$$

can be equivalently described as

$$\max_{Q_Y} \int \sum_v P_{\hat{V}|Y}(v|y) \, \mathrm{d}Q_Y(y) \le 1. \tag{49}$$

The fact that (48) implies (49) trivially follows by averaging both sides of (48) over an arbitrary Q_Y, and in particular, for the one maximizing (49). To prove that (49) implies (48), let us assume that (48) does not hold, i. e., $\sum_v P_{\hat{V}|Y}(v|\bar{y}) > 1$ for some $\bar{y} \in \mathcal{Y}$. Let $\bar{Q}_Y$ be the distribution that concentrates all the mass at $\bar{y}$. Since for $Q_Y = \bar{Q}_Y$ the condition (49) is violated, so happens for the maximizing Q_Y. As a result, (49) implies (48), as desired, and the equivalence between both expressions follows.

By using (49) instead of (48) in (35)-(36), and after replacing the sums by integrals where needed, we obtain

$$\bar{\epsilon} = \max_{\eta \ge 0} \min_{P_{\hat{V}|Y}} \left\{ \int \sum_v P_{V|Y}(v|y) \left(1 - P_{\hat{V}|Y}(v|y) \right) \, \mathrm{d}P_Y(y) + \eta \left(\max_{Q_Y} \int \sum_v P_{\hat{V}|Y}(v|y) \, \mathrm{d}Q_Y(y) - 1 \right) \right\}. \tag{50}$$

For fixed Q_Y the argument in (50) is linear with respect to $P_{\hat{V}|Y}$, and for fixed $P_{\hat{V}|Y}$ is linear with respect to Q_Y. Therefore, applying Sion's minimax theorem [17, Cor. 3.5] to interchange $\min_{P_{\hat{V}|Y}}$ and $\max_{Q_Y}$, (50) becomes (38). The first part of the theorem thus holds for continuous alphabets $\mathcal{Y}$. Since Lemma 1 applies to arbitrary probability spaces, so does (41). Therefore, for continuous alphabets $\mathcal{Y}$, the second part of the theorem follows from (41), (42) and (43)-(46) after replacing the sum by an integral in (45).

Remark 3: The optimality of $Q_Y^\star$ in (13) can also be proved constructively. Consider the binary hypothesis testing problem between P_{VY} and $Q_V \times Q_Y^\star$. We define a test

$$T_{\text{MAP}}(0|v,y) \triangleq \begin{cases} \frac{1}{|\mathcal{S}(y)|}, & \text{if } v \in \mathcal{S}(y), \\ 0, & \text{otherwise.} \end{cases} \tag{51}$$

For Q_V uniform, the type-1 error probability of this test is $\epsilon_1(Q_V \times Q_Y^\star, T_{\text{MAP}}) = \frac{1}{M}$. Using that the MAP test is a maximum metric test with $q(v,y) = P_{VY}(v,y)$, according to the proof of Corollary 1 in Appendix A, the type-0 error probability of T_{MAP} is precisely $\alpha_{\frac{1}{M}} \left(P_{VY}, Q_V \times Q_Y^\star \right)$. Moreover, since $\bar{\epsilon} = \epsilon_0(P_{VY}, T_{\text{MAP}})$ we conclude that $Q_Y = Q_Y^\star$ is an optimizer of (13). While both T_{MAP} and T_{NP} attain the Neyman-Pearson performance, in general

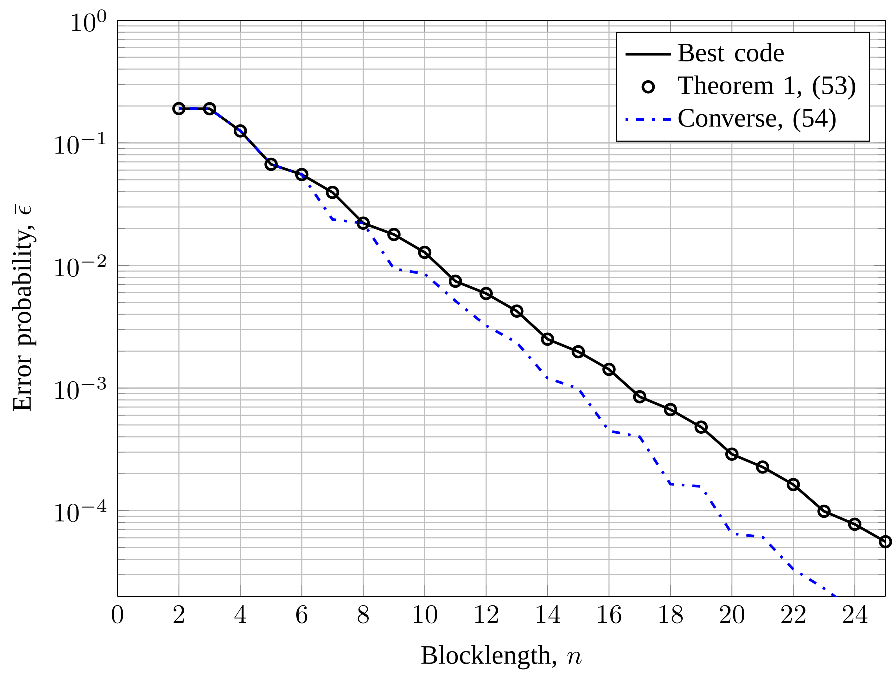

Figure 2. Channel coding error probability bounds for a BSC with cross-over probability 0.1 and $M = 4$ codewords.

they are not the same test, as they may differ in the set of points that lead to a MAP test tie, i.e., the values of y such that $|\mathcal{S}(y)| > 1$.

IV. Connection to Previous Converse Results

We next study the connection between Theorem 1 and previous converse results in the literature:

1) The meta-converse bound: In channel coding, one of M equiprobable messages is to be sent over a channel with one-shot law $P_{Y|X}$. The encoder maps the source message $v \in \{1, \ldots, M\}$ to a codeword $x(v)$ using a specific codebook $\mathcal{C}$. Since there is a codeword for each message, the distribution P_V induces a distribution $P_X^{\mathcal{C}}$ over the channel input. At the decoder, the decision among the M possible transmitted codewords based on the channel output y is equivalent to an M-ary hypothesis test with equiprobable hypotheses. The smallest error probability of this test for a codebook $\mathcal{C}$ is denoted as $\bar{\epsilon}(\mathcal{C})$.

Fixing an arbitrary Q_Y in (13) and considering the codeword set instead of the message set, we obtain

$$\bar{\epsilon}(\mathcal{C}) \geq \alpha_{\frac{1}{M}}\left(P_X^{\mathcal{C}} \times P_{Y|X}, P_X^{\mathcal{C}} \times Q_Y\right), \tag{52}$$

namely the meta-converse bound of [7, Th. 26] for a given codebook and the choice $Q_{XY} = P_X^{\mathcal{C}} \times Q_Y$. Theorem 1 thus shows that the meta-converse bound is tight for a fixed codebook after optimization over the auxiliary distribution Q_Y.

Upon optimization over Q_Y and minimization over codebooks we obtain

$$\min_{\mathcal{C}} \bar{\epsilon}(\mathcal{C}) = \min_{P_X^{\mathcal{C}}} \max_{Q_Y} \left\{ \alpha_{\frac{1}{M}} \left(P_X^{\mathcal{C}} \times P_{Y|X}, P_X^{\mathcal{C}} \times Q_Y \right) \right\} \tag{53}$$

$$\geq \min_{P_X} \max_{Q_Y} \left\{ \alpha_{\frac{1}{M}} \left(P_X \times P_{Y|X}, P_X \times Q_Y \right) \right\}. \tag{54}$$

The minimization in (53) is done over the set of distributions induced by all possible codes, while the minimization in (54) is done over the larger set of all possible distributions over the channel inputs. The bound in (54) coincides with [7, Th. 27].

Fig. 2 depicts the minimum error probability for the transmission of $M = 4$ messages over n independent, identically distributed channel uses of a memoryless binary symmetric channel (BSC) with single-letter cross-over probability 0.1. We also include the meta-converse (53), computed for the best code [18, Th. 37] and $Q_Y = Q_Y^{\star}$, and the lower bound in (54). Here, we exploited the fact that for the BSC the saddlepoint in (54) is attained for uniform P_X, Q_Y [19, Th. 22]. The computation of (53) and (54) follows similar steps to those presented in Section III-A for a different example. It is interesting to observe that while (53) characterizes the exact error probability, the weakening (54) yields a much looser bound.

2) Lower bound based on a bank of M binary tests: Eq. (13) relates the error probability $\bar{\epsilon}$ to the type-0 error probability of a binary test between distributions P_{VY} and $Q_V^{\star} \times Q_Y$. Instead of a single binary test, it is also possible to consider a bank of M binary hypothesis tests between distributions $P_{Y|V=v}$ and Q_Y [8]. In this case, we can also express the average error probability of M-ary hypothesis testing as

$$\bar{\epsilon} = \max_{Q_Y} \left\{ \sum_v P_V(v)\, \alpha_{Q_{\hat{V}}^{\star}(v)} \left(P_{Y|V=v}, Q_Y \right) \right\} \tag{55}$$

where $Q_{\hat{V}}^{\star}(v) \triangleq \sum_y Q_Y(y) P_{\hat{V}|Y}^{\text{MAP}}(v|y)$; see Appendix C.

If instead of fixing $Q_{\hat{V}}^{\star}$, we minimize (55) with respect to an arbitrary $Q_{\hat{V}}$, (55) then recovers the converse bound [8, Lem. 2] for almost-lossless joint source-channel coding. This lower bound is not tight in general as the minimizing distribution $Q_{\hat{V}}$ need not coincide with the distribution induced by the MAP decoder.

3) Verdú-Han lower bound: Weakening the identity in (14) for an arbitrary Q_Y we obtain

$$\bar{\epsilon} \geq \sup_{\gamma \geq 0} \left\{ \Pr\left[\frac{P_{VY}(V,Y)}{Q_Y(Y)} \leq \gamma \right] - \gamma \right\}. \tag{56}$$

By choosing $Q_Y = P_Y$ in (56) we recover the Verdú-Han lower bound in the channel [13, Th. 4] and joint source-channel coding settings [20, Lem. 3.2]. The bound (56) with arbitrary Q_Y coincides with the Hayashi-Nagaoka lemma for classical-quantum channels [21, Lem. 4], with its proof steps following exactly those of [13, Th. 4]. Theorem 1 shows that, by properly choosing Q_Y, this bound is tight in the classical setting.

4) Wolfowitz's strong converse: If we consider the hypothesis v with smallest error probability in (14), i. e.,

$$\bar{\epsilon} = \max_{Q_Y} \sup_{\gamma \geq 0} \left\{ \sum_v P_V(v) \Pr\left[\frac{P_{Y|V}(Y|v) P_V(v)}{Q_Y(Y)} \leq \gamma \right] - \gamma \right\} \tag{57}$$

$$\geq \max_{Q_Y} \sup_{\gamma \geq 0} \inf_v \left\{ \Pr\left[\frac{P_{Y|V}(Y|v) P_V(v)}{Q_Y(Y)} \leq \gamma \right] - \gamma \right\}, \tag{58}$$

we recover Wolfowitz's channel coding strong converse [22]. Hence, this converse bound is tight as long as the bracketed term in (58) does not depend on v for the pair $\{Q_Y, \gamma\}$ optimizing (57).

5) Poor-Verdú lower bound: By applying the following lemma, we recover the Poor-Verdú lower bound [14] from Theorem 1. Let us denote by $\mathbb{P}[\mathcal{E}]$ (resp. $\mathbb{Q}[\mathcal{E}]$) the probability of the event $\mathcal{E}$ with respect to the underlying distribution P (resp. Q).

Lemma 2: For a pair of discrete distributions $\{P, Q\}$ defined over $\mathcal{Y}$ and any $\gamma' \geq 0$, such that

$$0 \leq \beta \leq \frac{\mathbb{Q}\left[\frac{P(Y)}{Q(Y)} > \gamma'\right]}{\mathbb{P}\left[\frac{P(Y)}{Q(Y)} > \gamma'\right]}, \tag{59}$$

the following result holds,

$$\alpha_{\beta}\big(P, Q\big) \geq (1 - \gamma'\beta)\mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma'\right]. \tag{60}$$

Proof: See Appendix B. ■

Using Lemma 2 with $\gamma' = \gamma M$, $P \leftarrow P_{VY}$ and $Q \leftarrow Q_V \times Q_Y$ where Q_V is uniform, via (13), we obtain

$$\bar{\epsilon} \geq (1 - \gamma) \Pr\left[\frac{P_{VY}(V, Y)}{Q_Y(Y)} \leq \gamma\right], \tag{61}$$

provided that Q_Y and $\gamma \geq 0$ satisfy

$$\sum_{v,y} P_{VY}(v, y) \mathbb{1}\left\{\frac{P_{VY}(v, y)}{Q_Y(y)} > \gamma\right\} \leq \sum_{v,y} Q_Y(y) \mathbb{1}\left\{\frac{P_{VY}(v, y)}{Q_Y(y)} > \gamma\right\}. \tag{62}$$

This condition is fulfilled for any $\gamma \geq 0$ if $Q_Y = P_Y$ or $Q_Y = Q_Y^{\star}$ as defined in (15). However, there exist pairs $\{\gamma, Q_Y\}$ for which (62) does not hold. For $Q_Y = P_Y$, and optimizing over $\gamma \geq 0$, (61) recovers the Poor-Verdú bound [14, Th. 1]. For $Q_Y = Q_Y^{\star}$ in (15), optimizing over $\gamma \geq 0$, (61) provides an expression similar to those in Theorem 1:

$$\bar{\epsilon} = \max_{\gamma \geq 0} \left\{ (1 - \gamma) \Pr\left[\frac{P_{Y|V}(Y|V) P_V(V)}{Q_Y^{\star}(Y)} \leq \gamma\right]\right\}. \tag{63}$$

6) Lossy source coding: Finally, we consider a fixed-length lossy compression scenario, for which a converse based on hypothesis testing was recently obtained in [23, Th. 8]. The output of a general source v with distribution P_V is mapped to a codeword w in a codebook $\mathcal{C} = \{w_1, w_2, \dots, w_M\}$ with $w_1, w_2, \dots, w_M$ belonging to the reconstruction alphabet $\mathcal{W}$. We define a non-negative real-valued distortion measure $d(v, w)$ and a maximum allowed distortion D. The excess distortion probability is thus defined as $\epsilon_d(\mathcal{C}, D) \triangleq \Pr\big[d(V, W) > D\big]$. Consider an encoder that maps the source message v to codeword w with smallest pairwise distortion. The distortion associated to the source message v is then

$$d(v, \mathcal{C}) \triangleq \min_{w \in \mathcal{C}} d(v, w). \tag{64}$$

Consequently, the excess distortion probability is given by

$$\epsilon_d(\mathcal{C}, D) = \sum_{v} P_V(v) \mathbb{1}\big\{d(v, \mathcal{C}) > D\big\}. \tag{65}$$

Given the possible overlap between covering regions, there is no straightforward equivalence between the excess distortion probability and the error probability of an M-ary hypothesis test. We may yet define an alternative binary hypothesis test as follows. Given an observation v, we choose $\mathcal{H}_0$ if the encoder meets the maximum allowed distortion and $\mathcal{H}_1$ otherwise, i.e. the test is defined as

$$T_{\text{LSC}}(0|v) = \mathbb{1}\big\{d(v,\mathcal{C}) \leq D\big\}. \tag{66}$$

Particularizing (1) and (2) with this test, yields

$$\epsilon_0(P_V, T_{\text{LSC}}) = \sum_v P_V(v)\mathbb{1}\big\{d(v,\mathcal{C}) > D\big\}, \tag{67}$$

$$\epsilon_1(Q_V, T_{\text{LSC}}) = \sum_v Q_V(v)\mathbb{1}\big\{d(v,\mathcal{C}) \leq D\big\} \tag{68}$$

$$= \mathbb{Q}[d(V,\mathcal{C}) \leq D], \tag{69}$$

where $\mathbb{Q}[\mathcal{E}]$ denotes the probability of the event $\mathcal{E}$ with respect to the underlying distribution Q_V.

As (65) and (67) coincide, $\epsilon_d(\mathcal{C}, D)$ can be lower-bounded by the type-0 error of a Neyman-Pearson test, i.e.,

$$\epsilon_d(\mathcal{C}, D) \geq \max_{Q_V}\Big\{\alpha_{\mathbb{Q}[d(V,\mathcal{C})\leq D]}\big(P_V, Q_V\big)\Big\}. \tag{70}$$

Moreover, (70) holds with equality, as the next result shows.

Theorem 3: The excess distortion probability of lossy source coding with codebook $\mathcal{C}$ and maximum distortion D satisfies

$$\epsilon_d(\mathcal{C}, D) = \max_{Q_V}\Big\{\alpha_{\mathbb{Q}[d(V,\mathcal{C})\leq D]}\big(P_V, Q_V\big)\Big\} \tag{71}$$

$$\geq \max_{Q_V}\Big\{\alpha_{M \sup_{w\in\mathcal{W}} \mathbb{Q}[d(V,w)\leq D]}\big(P_V, Q_V\big)\Big\}. \tag{72}$$

Proof: See Appendix D. ∎

The right-hand-side of (71) still depends on the codebook $\mathcal{C}$ through $\mathbb{Q}[d(V,\mathcal{C}) \leq D]$. This dependence disappears in the relaxation (72), recovering the converse bound in [23, Th. 8]. The weakness of (72) comes from relaxing the type-1 error in the bound to M times the type-1-error contribution of the best possible codeword belonging to the reconstruction alphabet.

In almost-lossless coding, $D = 0$, the error events for different codewords no longer overlap, and the problem naturally fits into the hypothesis testing paradigm. Moreover, when Q_V is assumed uniform we have that $\mathbb{Q}\left[d(V,w) \leq 0\right] = \mathbb{Q}\left[V = w\right] = \frac{1}{|\mathcal{V}|}$ for any w and, therefore, (72) is an equality.

Acknowledgement

The authors would like to thank Sergio Verdú for multiple discussions. We would also thank Te Sun Han for providing the classical version of the Hayashi-Nagaoka's lemma.

APPENDIX A
PROOF OF COROLLARY 1

For a binary hypothesis testing problem between the distributions P_{VY} and $Q_{VY}^{(q)}$ in (22) we define the test $T_q(0|v,y) \triangleq P_{\hat{V}|Y}^{(q)}(v|y)$. We now show that the test T_q achieves the same type-I and type-II error probability as a NP test T_{NP} in (5). To this end, let us fix $\gamma = \mu'$ and

$$p = \frac{\sum_y \sum_{v\in\mathcal{S}_q(y)} \frac{1}{|\mathcal{S}_q(y)|} P_{VY}(v,y)}{\sum_y \sum_{v\in\mathcal{S}_q(y)} P_{VY}(v,y)} \tag{73}$$

$$= \frac{\sum_y \sum_{v\in\mathcal{S}_q(y)} \frac{1}{|\mathcal{S}_q(y)|} Q_{VY}^{(q)}(v,y)}{\sum_y \sum_{v\in\mathcal{S}_q(y)} Q_{VY}^{(q)}(v,y)}, \tag{74}$$

where equality between (73) and (74) holds since $P_{VY}(v,y) = \mu' Q_{VY}^{(q)}(v,y)$ for all y, $v \in \mathcal{S}_q(y)$.

The type-0 error probability of the NP test (5) with these values of γ and p is given by

$$\epsilon_0(P_{VY}, T_{\text{NP}}) = 1 - \sum_{v,y} P_{VY}(v,y) T_{\text{NP}}(0|v,y) \tag{75}$$

$$= 1 - \sum_y \sum_{v\in\mathcal{S}_q(y)} p P_{VY}(v,y) \tag{76}$$

$$= 1 - \sum_y \sum_{v\in\mathcal{S}_q(y)} \frac{1}{|\mathcal{S}_q(y)|} P_{VY}(v,y) \tag{77}$$

$$= 1 - \sum_{v,y} P_{VY}(v,y) T_q(0|v,y) \tag{78}$$

$$= \epsilon_0(P_{VY}, T_q), \tag{79}$$

where in (76) we used the definition of T_{NP} in (5) with $P \leftarrow P_{VY}$ and $Q \leftarrow Q_{VY}^{(q)}$ and the definition of $\mathcal{S}_q(y)$ in (22); (77) follows from (73), and (78) follows from the definition of T_q. Analogously, the type-1 error probability of the NP test is

$$\epsilon_1(Q_{VY}^{(q)}, T_{\text{NP}}) = \sum_y \sum_{v\in\mathcal{S}_q(y)} p Q_{VY}^{(q)}(v,y) \tag{80}$$

$$= \sum_y \sum_{v\in\mathcal{S}_q(y)} \frac{1}{|\mathcal{S}_q(y)|} Q_{VY}^{(q)}(v,y) \tag{81}$$

$$= \sum_{v,y} Q_{VY}^{(q)}(v,y) T_q(0|v,y) \tag{82}$$

$$= \epsilon_1(Q_{VY}^{(q)}, T_q), \tag{83}$$

where (81) follows from (74); and (82) follows from the definition of T_q.

Then, using (75)-(79) and (80)-(83), we obtain

$$\alpha_{\epsilon_1\left(Q_{VY}^{(q)}, T_q\right)}\left(P_{VY}, Q_{VY}^{(q)}\right) = \epsilon_0(P_{VY}, T_{\text{NP}}) \tag{84}$$

$$= \epsilon_0(P_{VY}, T_q). \tag{85}$$

Noting that $\bar{\epsilon}\big(P^{(q)}_{\hat{V}|Y}\big)$ and $\epsilon_0(P_{VY}, T_q)$ coincide by definition, then (16) holds with equality for $Q_{VY} = Q^{(q)}_{VY}$.

Applying Lemma 1 to (16) and fixing $Q_{VY} = Q^{(q)}_{VY}$ yields

$$\bar{\epsilon}\Big(P^{(q)}_{\hat{V}|Y}\Big) \geq \sup_{\gamma' \geq 0} \left\{ \Pr\left[\frac{P_{VY}(V,Y)}{Q^{(q)}_{VY}(V,Y)} \leq \gamma' \right] - \gamma' \epsilon_1\Big(Q^{(q)}_{VY}, P^{(q)}_{\hat{V}|Y}\Big) \right\}. \tag{86}$$

Choosing $\gamma' = \mu'$ in (86) direct computation shows that

$$\Pr\left[\frac{P_{VY}(V,Y)}{Q^{(q)}_{VY}(V,Y)} \leq \mu' \right] = \Pr\Big[q(V,Y) \leq \max_{v'} q(v', Y) \Big] \tag{87}$$

$$= 1 \tag{88}$$

and

$$\mu' \epsilon_1\Big(Q^{(q)}_{VY}, P^{(q)}_{\hat{V}|Y}\Big) = \sum_{v,y} P_{VY}(v,y) \frac{\max_{v'} q(v', y)}{q(v,y)} P^{(q)}_{\hat{V}|Y}(v|y) \tag{89}$$

$$= \sum_{v,y} P_{VY}(v,y) P^{(q)}_{\hat{V}|Y}(v|y), \tag{90}$$

where in (90) we have used that $P^{(q)}_{\hat{V}|Y}(v|y) \neq 0$ implies $q(v,y) = \max_{v'} q(v', y)$. Therefore, substituting (87)-(88) and (89)-(90) in (86), and using the definition of $\bar{\epsilon}(P_{\hat{V}|Y})$ in (7), we conclude that (86) holds with equality, and so does (17) with $Q_{VY} = Q^{(q)}_{VY}$.

Appendix B
Proof of Lemmas 1 and 2

Consider a binary hypothesis test between distributions P and Q defined over the alphabet $\mathcal{Y}$. Let us denote by $\mathbb{P}[\mathcal{E}]$ the probability of the event $\mathcal{E}$ with respect to the underlying distribution P, and $\mathbb{Q}[\mathcal{E}]$ that with respect to Q.

For the sake of clarity we assume that, for a given type-1 error β, the term p in (5) is equal to zero. The proof easily extends to arbitrary p, although with more complicated notation. Then, there exists $\gamma^\star$ such that

$$\beta = \mathbb{Q}\left[\frac{P(Y)}{Q(Y)} > \gamma^\star \right], \tag{91}$$

and the NP lemma yields

$$\alpha_\beta(P, Q) = \mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma^\star \right]. \tag{92}$$

For $0 \leq \gamma' < \gamma^\star$, $\mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma'\right] \leq \mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma^\star\right] = \alpha_\beta(P,Q)$. Then both Lemmas 1 and 2 hold trivially.

For $\gamma' \geq \gamma^\star$ it follows that

$$\alpha_\beta(P,Q) = \mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma' \right] - \mathbb{P}\left[\gamma^\star < \frac{P(Y)}{Q(Y)} \leq \gamma' \right] \tag{93}$$

$$\geq \mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma' \right] - \gamma' \mathbb{Q}\left[\gamma^\star < \frac{P(Y)}{Q(Y)} \leq \gamma' \right] \tag{94}$$

$$= \mathbb{P}\left[\frac{P(Y)}{Q(Y)} \leq \gamma' \right] - \gamma' \left(\mathbb{Q}\left[\frac{P(Y)}{Q(Y)} > \gamma^\star \right] - \mathbb{Q}\left[\frac{P(Y)}{Q(Y)} > \gamma' \right] \right), \tag{95}$$

where (94) follows by noting that in the interval considered $P(y) < \gamma' Q(y)$. Lemma 1 follows from (95) by lower bounding $\mathbb{Q}\left[\frac{P(Y)}{Q(Y)} > \gamma'\right] \geq 0$ and using (91). In order to prove Lemma 2, we shall use in (95) the tighter lower bound

$$\mathbb{Q}\left[\frac{P(Y)}{Q(Y)} > \gamma'\right] \geq \beta \mathbb{P}\left[\frac{P(Y)}{Q(Y)} > \gamma'\right], \tag{96}$$

which holds by the assumption in (59).

APPENDIX C

ONE TEST VERSUS MULTIPLE TESTS

In this appendix, we prove the equivalence between the optimization problems in (13) and (55). First, note that the argument of the maximization in (55) can be written in terms of tests T_v for fixed v as

$$\sum_v P_V(v) \alpha_{Q_{\hat{V}}(v)}\big(P_{Y|V=v}, Q_Y\big)$$

$$= \sum_v P_V(v) \min_{T_v : \epsilon_1(Q_Y, T_v) \leq Q_{\hat{V}}(v)} \Big\{ \epsilon_0(P_{Y|V=v}, T_v) \Big\} \tag{97}$$

$$= \sum_v P_V(v) \max_{\lambda(v) \geq 0} \min_{T_v} \left\{ \sum_y P_{Y|V}(y|v) T_v(1|y) - \lambda(v) \left(\sum_{y'} Q_Y(y') T_v(0|y') - Q_{\hat{V}}(v) \right) \right\}, \tag{98}$$

where (97) follows from the definition of $\alpha_{(\cdot)}(\cdot)$, and in (98) we used the definitions of the type-0 and type-1 errors and introduced the constraints into the objective by means of the Lagrange multipliers $\lambda(v)$.

Similarly, from (13) we have that

$$\max_{Q_Y} \alpha_{\frac{1}{M}} \left(P_{VY}, Q_V \times Q_Y\right)$$

$$= \max_{Q_V \times Q_Y} \alpha_{\epsilon_1(Q_V \times Q_Y, T_{\text{MAP}})} \left(P_{VY}, Q_V \times Q_Y\right) \tag{99}$$

$$= \max_{Q_Y} \max_{\eta \geq 0} \max_{Q_V} \min_{T} \left\{ \sum_{v,y} P_{VY}(v,y) T(1|v,y) + \eta \left(\sum_{v',y'} Q_V(v') Q_Y(y') \left(T(0|v',y') - P^{\text{MAP}}_{\hat{V}|Y}(v'|y') \right) \right) \right\} \tag{100}$$

$$= \max_{Q_Y} \sum_v P_V(v) \max_{\bar{\lambda}(v) \geq 0} \min_{T} \left\{ \sum_y P_{Y|V}(y|v) T(1|v,y) + \bar{\lambda}(v) \left(\sum_{y'} Q_Y(y') T(0|v,y') - Q_{\hat{V}}(v) \right) \right\}, \tag{101}$$

where (99) follows as Q_V uniform is a maximizer of the RHS of (99); in (100) used the definition of $\alpha_{(\cdot)}(\cdot)$, and introduced the constraint into the objective by means of the Lagrange multiplier η; and in (101) we rearranged terms and defined

$$\bar{\lambda}(v) \triangleq \frac{\eta Q_V(v)}{P_V(v)}. \tag{102}$$

The result follows from (98) and (101) by optimizing (98) over Q_Y and identifying $T(i|v,y) \equiv T_v(i|y)$, $i = 0, 1$.

APPENDIX D

PROOF OF THEOREM 3

We define

$$Q_V^{\mathcal{C}}(v) \triangleq \frac{1}{\mu''} \mathbb{1}\big\{ d(v, \mathcal{C}) > D \big\}, \tag{103}$$

with μ'' a normalization constant.

The NP test (5) with $P \leftarrow P_V$, $Q \leftarrow Q_V^{\mathcal{C}}$, $\gamma = \mu''$, $p = 1$, particularizes to

$$T_{\text{NP}}(0|v) = \begin{cases} 1, & \text{if } P_V(v) \geq \mathbb{1}\big\{d(v, \mathcal{C}) > D\big\}, \\ 0, & \text{otherwise.} \end{cases} \tag{104}$$

Assuming that $P_V(v) < 1$ for all v, eq. (104) reduces to

$$T_{\text{NP}}(0|v) = \mathbb{1}\big\{d(v, \mathcal{C}) \leq D\big\} \tag{105}$$

$$= T_{\text{LSC}}(0|v). \tag{106}$$

That is, for $Q_V = Q_V^{\mathcal{C}}$, the test T_{LSC} defined in (66) is optimal in the Newman-Pearson sense. Then it holds that

$$\max_{Q_V} \big\{\alpha_{\epsilon_1(Q_V, T_{\text{LSC}})}\big(P_V, Q_V\big)\big\} \geq \alpha_{\epsilon_1(Q_V^{\mathcal{C}}, T_{\text{LSC}})}\big(P_V, Q_V^{\mathcal{C}}\big) \tag{107}$$

$$= \epsilon_0\big(P_V, T_{\text{LSC}}\big) \tag{108}$$

$$= \epsilon_d(\mathcal{C}, D), \tag{109}$$

where the last step follows since (65) and (67) coincide.

From (70) and (107)-(108), the equality (71) follows by noting that $\epsilon_1(Q_V, T_{\text{LSC}}) = \mathbb{Q}[d(V, \mathcal{C}) \leq D]$.

Let $P_{W|V}$ denote the encoder that maps the source message v to the codeword $w \in \mathcal{C}$ with smallest pairwise distortion. The lower bound (72) follows from the fact that

$$\epsilon_1(Q_V, T_{\text{LSC}}) = \sum_v Q_V(v) \mathbb{1}\left\{d(v, \mathcal{C}) \leq D\right\} \tag{110}$$

$$= \sum_v Q_V(v) \sum_w P_{W|V}(w|v) \mathbb{1}\left\{d(v, w) \leq D\right\} \tag{111}$$

$$\leq \sum_{w \in \mathcal{C}} \sum_v Q_V(v) \mathbb{1}\left\{d(v, w) \leq D\right\} \tag{112}$$

$$\leq M \sup_{w \in \mathcal{C}} \sum_v Q_V(v) \mathbb{1}\left\{d(v, w) \leq D\right\} \tag{113}$$

$$\leq M \sup_{w \in \mathcal{W}} \sum_v Q_V(v) \mathbb{1}\left\{d(v, w) \leq D\right\}, \tag{114}$$

where in (112) we used that $P_{W|V}(w|v) = 0$ for $w \notin \mathcal{C}$ and that $P_{W|V}(w|v) \leq 1$ for $w \in \mathcal{C}$; (113) follows from considering the largest term in the sum, and in (114) we relaxed the set over which the maximization is performed.

References

[1] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc. Lond. A*, vol. 231, no. 694-706, p. 289, 1933.

[2] 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, no. 1, pp. 65–103, Jan. 1967.

[3] ——, "Lower bounds to error probability for coding on discrete memoryless channels. II," *Inf. Contr.*, vol. 10, no. 5, pp. 522–552, May 1967.

[4] G. D. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[5] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, July 1974.

[6] J. K. Omura, "A lower bounding method for channel and source coding probabilities," *Inf. Contr.*, vol. 27, no. 2, pp. 148–177, Feb. 1975.

[7] 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.

[8] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, and A. Martinez, "Converse bounds for finite-length joint source-channel coding," in *50th Allerton Conf. on Comm., Cont. and Comp.*, Monticello, Illinois, USA, Oct. 2012.

[9] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[10] T. S. Han, *Information-spectrum methods in information theory*. Berlin, Germany: Springer-Verlag, 2003.

[11] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inf. Contr.*, vol. 1, no. 1, pp. 6–25, 1957.

[12] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[13] 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.

[14] 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.

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

[16] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, Sep. 2010.

[17] M. Sion, "On general minimax theorems." *Pacific J. Math.*, vol. 8, no. 1, pp. 171–176, 1958.

[18] P.-N. Chen, H.-Y. Lin, and S. Moser, "Optimal ultrasmall block-codes for binary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7346–7378, Nov. 2013.

[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] T. S. Han, "Joint source-channel coding revisited: information-spectrum approach," arXiv:0712.2959v1 [cs.IT], Dec. 2007.

[21] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1753–1768, July 2003.

[22] J. Wolfowitz, "Notes on a general strong converse," *Inf. Contr.*, vol. 12, no. 1, pp. 1–4, Jan. 1968.

[23] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.