Codeword or Noise? Exact Random Coding Exponents for Slotted Asynchronism*

Neri Merhav

August 21, 2013

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

We consider the problem of slotted asynchronous coded communication, where in each time frame (slot), the transmitter is either silent or transmits a codeword from a given (randomly selected) codebook. The task of the decoder is to decide whether transmission has taken place, and if so, to decode the message. We derive the optimum detection/decoding rule in the sense of the best trade-off among the probabilities of decoding error, false alarm, and misdetection. For this detection/decoding rule, we then derive single–letter characterizations of the exact exponential rates of these three probabilities for the average code in the ensemble.

Index Terms: Synchronization, error exponent, false alarm, misdetection, random coding.

*This research was supported by the Israel Science Foundation (ISF), grant no. 412/12.

1 Introduction

The problem of synchronization has been a long–standing, important issue in communication throughout several decades (see, e.g., [1], [2], [4], [5], [7], [11], [12], [13], [14] and references therein, for a non–exhaustive sample of earlier works).

The general problem setting under consideration allows the transmitter to send messages only part of the time, and to be 'silent' (non–transmitting) when it has no messages ready to be conveyed. The receiver then has to be able to reliably detect the existence of the message, locate its starting time instant, and decode it. The traditional approach has been to separate the problems of synchronization and coding/decoding, where in the former, a special pattern of symbols (synchronization word) is used to mark the beginning of a message transmission. This transmission of a synchronization word is, however, is an undesired overhead.

Following [13] and [14], in this work, we treat the synchronization and coding jointly and we adopt the simplified model of *slotted* communication. According to this model, a transmission can start only at time instants that are integer multiples of the slot length, which is also the block length. Thus, in each slot (or block), the transmitter is either entirely silent, or it transmits a codeword corresponding to one of M possible messages. In the silent mode, it is assumed that the transmitter repetitively feeds the channel by a special channel input symbol denoted by '0' (indeed, in the case of a continuous input alphabet, it is natural to assign a zero input signal), and then the channel output vector is thought of as "pure noise." The decoder in turn has to decide whether a message has been sent or the received channel output vector is pure noise. In case it decides in favor of the former, it then has to decode the message.

In [13] and [14], three figures of merit were defined in order to judge performance: (i) the probability of *false alarm* (FA) – i.e., deciding that a message has been sent when actually, the transmitter was silent and the channel output was pure noise, (ii) the probability of *misdetection* (MD) – that is, deciding that the transmitter was silent when it actually transmitted some message, and (iii) the probability of *decoding error* (DE) – namely, not deciding on the correct message sent. Wang [13] and Wang *et al.* [14] have posed the problem of characterizing the best achievable region of the error exponents associated with these three probabilities for a given discrete memoryless channel (DMC). It was stated in [14] that this general problem is open, and so, the focus both in

[13] and [14] was directed to the narrower problem of trading off the FA exponent and the MD exponent when the DE exponent constraint is completely relaxed, that is, there is no demand on exponential decay rate of the DE probability. Upper and lower bounds on the maximum achievable FA exponent for a given MD exponent were derived in these works. In the extreme case where the MD exponent constraint is omitted (set to zero), these bounds coincide, and so, the characterization of the best achievable MD exponent is exact.

In this paper, we adopt the same problem setting of slotted asynchronous communication as in [13] and [14]. We first derive, for a given code, the optimum detection–decoding rule that minimizes the DE probability subject to given constraints on the FA and the MD probabilities. This detection–decoding rule turns out to be completely different from the one in the achievability parts of [13] and [14]. In particular, denoting the codewords by $\{\boldsymbol{x}_m\}$, the channel output vector by $\boldsymbol{y}$ (all of length n), and the channel conditional probability by $W(\boldsymbol{y}|\boldsymbol{x}_m)$, then according to this rule, a transmission is detected iff

$$e^{n\alpha}\sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m) + \max_{1\le m\le M} W(\boldsymbol{y}|\boldsymbol{x}_m) \ge e^{n\beta} W(\boldsymbol{y}|0^n) \tag{1}$$

where α and β are chosen to meet the MD and FA constraints. Of course, whenever the received $\boldsymbol{y}$ passes this test, the maximum likelihood (ML) decoder is applied, assuming that all messages are equiprobable *a-priori*. The performance of this optimum detector/decoder is analyzed under the random coding regime of fixed composition codes, and the achievable trade-off between the three error exponents is given in full generality, that is, not merely in the margin where at least one of the exponents vanishes. It should be pointed out that our analysis technique, which is based on type class enumeration (see, e.g., [6], [10] and references therein), provides the *exact* random coding exponents, not just bounds. These relationships between the random coding exponents and the parameters α and β can, in principle, be inverted (in a certain domain) in order to find the assignments of α and β needed to satisfy given constraints on the exponents of the FA and the MD probabilities. For the sake of fairness, on the other hand, it should also be made clear that since we consider only the random coding regime, these are merely achievability results, with no converse bounds pertaining to optimal codes.

The outline of the paper is as follows. In Section 2, we establish some notation conventions, provide some preliminaries, and finally, formulate the problem. In Section 3, we derive the optimum

detector/decoder and discuss some of its properties. In Section 4, we present our main theorem, which is about single–letter formulas for the various error exponents. Finally, in Section 5, we prove this theorem.

2 Notation Conventions, Preliminaries and Problem Formulation

2.1 Notation Conventions and Preliminaries

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets, similarly as other sets, will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, ($n$ – positive integer) may take a specific vector value $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$, the n–th order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector.

For a given vector $\boldsymbol{x}$, let $\hat{Q}_X$ denote[1] the empirical distribution, that is, the vector $\{\hat{Q}_X(x),\ x \in \mathcal{X}\}$, where $\hat{Q}_X(x)$ is the relative frequency of the letter x in the vector $\boldsymbol{x}$. Let $\mathcal{T}_P$ denote the type class associated with P, that is, the set of all sequences $\{\boldsymbol{x}\}$ for which $\hat{Q}_X = P$. Similarly, for a pair of vectors $(\boldsymbol{x}, \boldsymbol{y})$, the empirical joint distribution will be denoted by $\hat{Q}_{XY}$ or simply $\hat{Q}$ for short. Conditional empirical distributions will be denoted by $\hat{Q}_{X|Y}$ and $\hat{Q}_{Y|X}$, the y–marginal by $\hat{Q}_Y$, etc. Accordingly, the empirical mutual information induced by $(\boldsymbol{x}, \boldsymbol{y})$ will be denoted by $I(\hat{Q}_{XY})$ or $I(\hat{Q})$, the divergence between $\hat{Q}_X$ and $P = \{P(x),\ x \in \mathcal{X}\}$ – by $\mathcal{D}(\hat{Q}_Y \| P)$, and the conditional divergence between the empirical conditional distribution $\hat{Q}_{Y|X}$ and the channel $W = \{W(y|x)\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$, will be denoted by $\mathcal{D}(\hat{Q}_{Y|X} \| W | \hat{Q}_X)$, that is,

$$\mathcal{D}(\hat{Q}_{Y|X} \| W | \hat{Q}_X) = \sum_{x \in \mathcal{X}} \hat{Q}_X(x) \sum_{y \in \mathcal{Y}} \hat{Q}_{Y|X}(y|x) \log \frac{\hat{Q}_{Y|X}(y|x)}{W(y|x)}, \tag{2}$$

and so on. The joint distribution induced by $\hat{Q}_X$ and $\hat{Q}_{Y|X}$ will be denoted by $\hat{Q}_X \times \hat{Q}_{Y|X}$, and a similar notation will be used when the roles of X and Y are switched. The marginal of X, induced by $\hat{Q}_Y$ and $\hat{Q}_{X|Y}$ will be denoted by $(\hat{Q}_Y \times \hat{Q}_{X|Y})_X$, and so on. Similar notation conventions will

[1] In our notation, we do not index $\hat{Q}_X$ by $\boldsymbol{x}$ because the underlying sequence $\boldsymbol{x}$ will be clear from the context.

apply, of course, to generic distributions Q_{XY}, Q_X, Q_Y, $Q_{Y|X}$, and $Q_{X|Y}$, which are not necessarily empirical distributions (without "hats").

The expectation operator will be denoted by $\boldsymbol{E}\{\cdot\}$. Whenever there is room for ambiguity, the underlying probability distribution will appears as a subscript, e.g., $\boldsymbol{E}_Q\{\cdot\}$. Logarithms and exponents will be understood to be taken to the natural base unless specified otherwise. The indicator function will be denoted by $\mathcal{I}(\cdot)$. Sets will normally be denoted by calligraphic letters. The complement of a set $\mathcal{A}$ will be denoted by $\overline{\mathcal{A}}$. The notation $[t]_+$ will stand for $\max\{t, 0\}$. For two positive sequences, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ will mean asymptotic equivalence in the exponential scale, that is, $\lim_{n\to\infty} \frac{1}{n} \log(\frac{a_n}{b_n}) = 0$. Similarly, $a_n \dot{\leq} b_n$ will mean $\limsup_{n\to\infty} \frac{1}{n} \log(\frac{a_n}{b_n}) \leq 0$, and so on. Throughout the sequel, we will make frequent use of the fact that $\sum_{i=1}^{k_n} a_i(n) \doteq \max_{1\leq i\leq k_n} a_i(n)$ as long as as $\{a_i(n)\}$ are positive and $k_n \doteq 1$. Accordingly, for k_n sequences of positive random variables $\{A_i(n)\}$, all defined on a common probability space, and a deterministic sequence B_n,

$$\begin{aligned}
\Pr\left\{\sum_{i=1}^{k_n} A_i(n) \geq B_n\right\} &\doteq \Pr\left\{\max_{1\leq i\leq k_n} A_i(n) \geq B_n\right\} \\
&= \Pr\bigcup_{i=1}^{k_n} \{A_i(n) \geq B_n\} \\
&\doteq \sum_{i=1}^{k_n} \Pr\{A_i(n) \geq B_n\} \\
&\doteq \max_{1\leq i\leq k_n} \Pr\{A_i(n) \geq B_n\}, \qquad (3)
\end{aligned}$$

provided that $B'_n \doteq B_n$ implies $\Pr\{A_i(n) \geq B'_n\} \doteq \Pr\{A_i(n) \geq B_n\}$.[2] In simple words, summations and maximizations are equivalent and can be both "pulled out outside" $\Pr\{\cdot\}$ without changing the exponential order, as long as $k_n \doteq 1$. By the same token,

$$\begin{aligned}
\Pr\left\{\sum_{i=1}^{k_n} A_i(n) \leq B_n\right\} &\doteq \Pr\left\{\max_{1\leq i\leq k_n} A_i(n) \leq B_n\right\} \\
&= \Pr\bigcap_{i=1}^{k_n} \{A_i(n) \leq B_n\}. \qquad (4)
\end{aligned}$$

Another fact that will be used extensively is that for a given set of M pairwise independent events

[2] Consider the case where $B_n \doteq e^{bn}$ (b being a constant independent of n) and the exponent of $\Pr\{A_i(n) \geq e^{bn}\}$ is a continuous function of b.

$\{\mathcal{A}_i\}_{i=1}^{M}$,

$$\Pr\left\{\bigcup_{i=1}^{M}\mathcal{A}_i\right\} \dot{=} \min\left\{1, \sum_{i=1}^{M}\Pr\{\mathcal{A}_i\}\right\}. \tag{5}$$

The right–hand side (r.h.s.) is obviously the union bound, which holds true even if the events are not pairwise independent. On the other hand, when multiplied by a factor of 1/2, the r.h.s. becomes a lower bound to $\Pr\{\bigcup_{i=1}^{M}\mathcal{A}_i\}$, provided that $\{A_i\}$ are pairwise independent [8, Lemma A.2], [9, Lemma 1].

2.2 Problem Formulation

Consider a discrete memoryless channel (DMC), characterized by a finite input alphabet $\mathcal{X}_0$, a finite out alphabet $\mathcal{Y}$ and a given matrix of single–letter transition probabilities $\{W(y|x),\ x \in \mathcal{X}_0,\ y \in \mathcal{Y}\}$. It is further assumed that $\mathcal{X}_0$ contains a special symbol denoted by '0', which designates the channel input in the absence of transmission. We shall denote $\mathcal{X} = \mathcal{X}_0 \setminus \{0\}$ and $Q_0(y) = W(y|x=0)$.

We assume an ensemble of random codes, where each codeword is selected independently at random, uniformly within a type class $\mathcal{T}_P$. Let $\mathcal{C} = \{\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots, \boldsymbol{x}_M\}$, $\boldsymbol{x}_m \in \mathcal{X}^n$, $m = 1, \ldots, M$, $M = e^{nR}$ (R being the coding rate in nats per channel use), denote the (randomly chosen) code, which is revealed to both the encoder and the decoder.

A detector/decoder, for a code operating in the setting of slotted asynchronous communication, is a partition of $\mathcal{Y}^n$ into $M+1$ regions, denoted $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_M$. If $\boldsymbol{y} \in \mathcal{R}_m$, $m = 1, 2, \ldots, M$, then the decoder decodes the message to be m. If $\boldsymbol{y} \in \mathcal{R}_0$, then the decoder declares that nothing has been transmitted, that is, $\boldsymbol{x} = 0^n$ and then $\boldsymbol{y}$ is "pure noise." The probability of decoding error (DE) is defined as

$$P_{\mathrm{DE}} = \frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_m}) = \frac{1}{M}\sum_{m=1}^{M}\sum_{k\neq m} W(\mathcal{R}_k|\boldsymbol{x}_m), \tag{6}$$

where the inner summation at the right–most side *includes* $k = 0$. The probability of false alarm (FA) is defined as

$$P_{\mathrm{FA}} = Q_0(\overline{\mathcal{R}_0}) = \sum_{m=1}^{M} Q_0(\mathcal{R}_m), \tag{7}$$

and the probability of misdetection (MD) is defined as

$$P_{\mathrm{MD}} = \frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0|\boldsymbol{x}_m). \tag{8}$$

For a given code $\mathcal{C}$, we are basically interested in achievable trade-offs between P_{DE}, P_{FA}, and P_{MD}. Consider the following problem:

$$\begin{aligned} \text{minimize} \quad & P_{\text{DE}} \\ \text{subject to} \quad & P_{\text{FA}} \leq \epsilon_{\text{FA}} \\ & P_{\text{MD}} \leq \epsilon_{\text{MD}} \end{aligned} \tag{9}$$

where ϵ_{FA} and ϵ_{MD} are given prescribed quantities, and it assumed that these two constraints are not contradictory.[3]

Our goal is to find the optimum detector/decoder and then analyze the random coding exponents associated with the resulting error probabilities.

3 The Optimum Detector/Decoder

Let us define the following detector/decoder:

$$\mathcal{R}_0^* = \left\{ \boldsymbol{y} : \ a \cdot \sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m) + \max_m W(\boldsymbol{y}|\boldsymbol{x}_m) \leq b \cdot Q_\star(\boldsymbol{y}) \right\} \tag{10}$$

$$\mathcal{R}_m^* = \overline{\mathcal{R}_0^*} \bigcap \left\{ \boldsymbol{y} : \ W(\boldsymbol{y}|\boldsymbol{x}_m) > \max_{k \neq m} W(\boldsymbol{y}|\boldsymbol{x}_k) \right\}, \quad m = 1, 2, \ldots, M, \tag{11}$$

where ties are broken arbitrarily, and where $a \geq 0$ and $b \geq 0$ are deterministic constants. The following lemma establishes the optimality of the decision rule $\mathcal{R}^* = \{\mathcal{R}_0^*, \mathcal{R}_1^*, \ldots, \mathcal{R}_M^*\}$ in the sense of the trade-off among the probabilities P_{MD}, P_{FA} and P_{DE}. It tells us that there is no other decision rule that simultaneously yields strictly smaller error probabilities of all three kinds.

Lemma 1 *Let $\mathcal{R}^* = \{\mathcal{R}_0^*, \mathcal{R}_1^*, \ldots, \mathcal{R}_M^*\}$ be as above and let $\mathcal{R} = \{\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_M\}$ be any another partition of $\mathcal{Y}^n$ into $M+1$ regions. If*

$$Q_0(\overline{\mathcal{R}_0}) \leq Q_0(\overline{\mathcal{R}_0^*}) \tag{12}$$

[3]Note that there is some tension between P_{MD} and P_{FA} as they are related via the Neyman–Pearson lemma. For a given ϵ_{FA}, the minimum achievable MD probability is positive, in general. It is assumed then that the prescribed value of ϵ_{MD} is not smaller than this minimum. In the problem under consideration, it makes sense to relax the tension between the two constraints to a certain extent, in order to allow some freedom to minimize P_{DE} under these constraints.

and

$$\frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0|\boldsymbol{x}_m) \leq \frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0^*|\boldsymbol{x}_m), \tag{13}$$

then

$$\frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_m^*}|\boldsymbol{x}_m) \leq \frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_m}|\boldsymbol{x}_m). \tag{14}$$

Proof. We begin from the obvious observation that for a given choice of $\mathcal{R}_0$, the optimum choice of the other decision regions is always:

$$\mathcal{R}_m = \overline{\mathcal{R}_0}\bigcap\left\{\boldsymbol{y}:\ W(\boldsymbol{y}|\boldsymbol{x}_m) > \max_{k\neq m} W(\boldsymbol{y}|\boldsymbol{x}_k)\right\}, \qquad m = 1,2,\ldots,M. \tag{15}$$

In other words, once a transmission has been detected, the best decoding rule is the ML decoding rule. Similarly as in classical hypothesis testing theory, this is true because the probability of correct decoding,

$$P_{\rm CD} = \frac{1}{M}\sum_{m=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{R}_m} W(\boldsymbol{y}|\boldsymbol{x}_m), \tag{16}$$

is upper bounded by

$$P_{\rm CD} \leq \frac{1}{M}\sum_{m=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{R}_m} \max_k W(\boldsymbol{y}|\boldsymbol{x}_k) = \frac{1}{M}\sum_{\boldsymbol{y}\in\overline{\mathcal{R}_0}} \max_m W(\boldsymbol{y}|\boldsymbol{x}_m) \tag{17}$$

and this bound is achieved by (15). Thus, upon adopting (15) for a given choice of $\mathcal{R}_0$, it remains to prove that the choice $\mathcal{R}_0^*$ satisfies the assertion of the lemma.

The proof of this fact is similar to the proof of the Neyman–Pearson lemma. Let $\mathcal{R}_0^*$ be as above and let $\mathcal{R}_0$ be another, competing rejection region. First, observe that for every $\boldsymbol{y}\in\mathcal{Y}^n$

$$[\mathcal{I}\{\boldsymbol{y}\in\mathcal{R}_0^*\} - \mathcal{I}\{\boldsymbol{y}\in\mathcal{R}_0\}]\cdot\left[b\cdot Q_0(\boldsymbol{y}) - a\cdot\sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m) - \max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\right] \geq 0. \tag{18}$$

This is true because, by definition of $\mathcal{R}_0^*$, the two factors of the product at the left–hand side (l.h.s.) are either both non–positive or both non–negative. Thus, taking the summation over all $\boldsymbol{y}\in\mathcal{Y}^n$, we have:

$$\begin{aligned} 0 &\leq \sum_{\boldsymbol{y}\in\mathcal{Y}^n}[\mathcal{I}\{\boldsymbol{y}\in\mathcal{R}_0^*\} - \mathcal{I}\{\boldsymbol{y}\in\mathcal{R}_0\}]\cdot\left[b\cdot Q_0(\boldsymbol{y}) - a\cdot\sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m) - \max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\right] \\ &= b\cdot[Q_0(\mathcal{R}_0^*) - Q_0(\mathcal{R}_0)] - a\cdot\left[\sum_{m=1}^{M} W(\mathcal{R}_0^*|\boldsymbol{x}_m) - \sum_{m=1}^{M} W(\mathcal{R}_0|\boldsymbol{x}_m)\right] - \end{aligned}$$

$$\left[\sum_{\boldsymbol{y}\in\mathcal{R}_0^*}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)-\sum_{\boldsymbol{y}\in\mathcal{R}_0}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\right] \tag{19}$$

which yields

$$\begin{aligned}
&\sum_{\boldsymbol{y}\in\mathcal{R}_0^*}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)-\sum_{\boldsymbol{y}\in\mathcal{R}_0}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\\
\leq\;& b\cdot[Q_0(\mathcal{R}_0^*)-Q_0(\mathcal{R}_0)]-a\cdot\left[\sum_{m=1}^M W(\mathcal{R}_0^*|\boldsymbol{x}_m)-\sum_{m=1}^M W(\mathcal{R}_0|\boldsymbol{x}_m)\right]\\
=\;& b\cdot[Q_0(\overline{\mathcal{R}_0})-Q_0(\overline{\mathcal{R}_0^*})]+a\cdot\left[\sum_{m=1}^M W(\mathcal{R}_0|\boldsymbol{x}_m)-\sum_{m=1}^M W(\mathcal{R}_0^*|\boldsymbol{x}_m)\right]
\end{aligned} \tag{20}$$

Since $a\geq 0$ and $b\geq 0$, it follows that

$$Q_0(\overline{\mathcal{R}_0})\leq Q_0(\overline{\mathcal{R}_0^*}) \tag{21}$$

and

$$\frac{1}{M}\sum_{m=1}^M W(\mathcal{R}_0|\boldsymbol{x}_m)\leq\frac{1}{M}\sum_{m=1}^M W(\mathcal{R}_0^*|\boldsymbol{x}_m) \tag{22}$$

together imply that

$$\sum_{\boldsymbol{y}\in\mathcal{R}_0^*}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\leq\sum_{\boldsymbol{y}\in\mathcal{R}_0}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m) \tag{23}$$

or equivalently,

$$\sum_{\boldsymbol{y}\in\overline{\mathcal{R}_0^*}}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\geq\sum_{\boldsymbol{y}\in\overline{\mathcal{R}_0}}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m), \tag{24}$$

which in turn yields

$$\begin{aligned}
\frac{1}{M}\sum_{m=1}^M W(\overline{\mathcal{R}_m^*}|\boldsymbol{x}_m)\;&\equiv\;1-\frac{1}{M}\sum_{\boldsymbol{y}\in\overline{\mathcal{R}_0^*}}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\\
&\leq\;1-\frac{1}{M}\sum_{\boldsymbol{y}\in\overline{\mathcal{R}_0}}\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)\\
&\equiv\;\frac{1}{M}\sum_{m=1}^M W(\overline{\mathcal{R}_m}|\boldsymbol{x}_m).
\end{aligned} \tag{25}$$

This completes the proof of Lemma 1. □

Discussion. At this point, two comments are in order.

1. The results thus far hold for any given code $\mathcal{C}$. As mentioned earlier, in this work, we analyze the ensemble performance. Specifically, let $\bar{P}_{\mathrm{DE}}$, $\bar{P}_{\mathrm{FA}}$, and $\bar{P}_{\mathrm{MD}}$ denote the corresponding ensemble averages of P_{DE}, P_{FA}, and P_{MD}, respectively. We will assess the random coding exponents of these three probabilities. The constants a and b can be thought of as Lagrange multipliers that are tuned to meet the given FA and MD constraints. For these Lagrange multipliers to have an impact on error exponents, we let them be exponential functions of n, that is, $a = e^{n\alpha}$ and $b = e^{n\beta}$, where α and β are real numbers, independent of n. The rejection region is then of the form

$$\mathcal{R}_0^* = \left\{ \boldsymbol{y} : \ e^{n\alpha} \sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m) + \max_m W(\boldsymbol{y}|\boldsymbol{x}_m) \le e^{n\beta} Q_0(\boldsymbol{y}) \right\}. \tag{26}$$

By the same token, we impose exponential constraints on the FA and MD probabilities, that is, $\epsilon_{\mathrm{FA}} = e^{-nE_{\mathrm{FA}}}$ and $\epsilon_{\mathrm{MD}} = e^{-nE_{\mathrm{MD}}}$, where $E_{\mathrm{FA}} \ge 0$ and $E_{\mathrm{MD}} \ge 0$ are given numbers, independent of n.

2. The detection/rejection rule defined by (26) involves a linear combination of $\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)$ and $\sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m)$, or equivalently, the overall output distribution induced by the code

$$Q_{\mathcal{C}}(\boldsymbol{y}) \stackrel{\Delta}{=} \frac{1}{M} \sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m). \tag{27}$$

In this context, the intuition behind the optimality of this detection rule is not trivial (at least for the author of this article), and as mentioned earlier, it is very different from that of [13] and [14]. It is instructive, nonetheless, to examine some special cases. The first observation is that for $\alpha \ge 0$, the term $e^{n\alpha} \sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m)$ dominates the term $\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)$, and so, the rejection region is essentially equivalent to

$$\mathcal{R}_0' = \left\{ \boldsymbol{y} : \ e^{n\alpha} \sum_{m=1}^{M} W(\boldsymbol{y}|\boldsymbol{x}_m) \le e^{n\beta} Q_0(\boldsymbol{y}) \right\} = \left\{ \boldsymbol{y} : \ e^{n(\alpha+R)} Q_{\mathcal{C}}(\boldsymbol{y}) \le e^{n\beta} Q_0(\boldsymbol{y}) \right\}, \tag{28}$$

which is exactly the Neyman–Pearson test between $Q_{\mathcal{C}}(\boldsymbol{y})$ and $Q_0(\boldsymbol{y})$. This means that $\alpha \ge 0$ corresponds to a regime of full tension between the FA and the MD constraints (see footnote no. 2). In this case, E_{FA} and E_{MD} are related via the Neyman–Pearson lemma, and there are no degrees of freedom left for minimizing $\bar{P}_{\mathrm{DE}}$ (or equivalently, maximizing its exponent). Indeed, the detection–rejection rule (28) depends only on one degree of freedom, which is the difference $\alpha - \beta$, and hence so are the FA and MD error exponents associated with it. At the other extreme, where

$e^{n\alpha} \ll 1$, and the term $\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)$ dominates, the detection rule becomes equivalent to

$$\mathcal{R}_0'' = \left\{\boldsymbol{y}: \ \max_m W(\boldsymbol{y}|\boldsymbol{x}_m) \le e^{n\beta} Q_0(\boldsymbol{y})\right\}. \tag{29}$$

In this case, the silent mode is essentially treated as corresponding to yet another codeword – $\boldsymbol{x}_0 = 0^n$, although it still has a special stature due to the factor $e^{n\beta}$. But for $\beta = 0$, this "silent codeword" is just an additional codeword with no special standing, and the decoding is completely ordinary. The interesting range is therefore the range where α is negative, but not too small, where both $Q_{\mathcal{C}}(\boldsymbol{y})$ and $\max_m W(\boldsymbol{y}|\boldsymbol{x}_m)$ play a considerable role.

4 Performance

In this section, we present our main theorem, which provides exact single–letter characterizations for all three exponents as functions of α and β. We first need some definitions. Let

$$d(x,y) \stackrel{\Delta}{=} \ln\left[\frac{Q_0(y)}{W(y|x)}\right], \quad x \in \mathcal{X}, \ y \in \mathcal{Y} \tag{30}$$

and denote $D(Q) = \boldsymbol{E}_Q d(X,Y)$. For a given output distribution $Q_Y = \{Q_Y(y), \ y \in \mathcal{Y}\}$, define[4]

$$\boldsymbol{R}(\Delta; Q_Y) \stackrel{\Delta}{=} \inf_{\{Q_{Y|X}: \ D(Q)\le\Delta, \ (P\times Q_{Y|X})_Y = Q_Y\}} I(Q). \tag{31}$$

Next, define

$$\mu(Q_Y, R) \stackrel{\Delta}{=} \min_{Q_{X|Y}\in\mathcal{Q}_P, \ I(Q)\le R} \{I(Q) + D(Q)\}, \tag{32}$$

$$\tilde{\boldsymbol{R}}(\Delta, R; Q_Y) \stackrel{\Delta}{=} \begin{cases} \boldsymbol{R}(\Delta; Q_Y) - R & \Delta \le \mu(Q_Y, R) - R \\ 0 & \Delta > \mu(Q_Y, R) - R \end{cases} \tag{33}$$

$$E_A \stackrel{\Delta}{=} \inf_{Q_Y} [\mathcal{D}(Q_Y\|Q_0) + \tilde{\boldsymbol{R}}(\alpha - \beta, R; Q_Y)], \tag{34}$$

$$E_B \stackrel{\Delta}{=} \inf_{Q_Y} \left\{\mathcal{D}(Q_Y\|Q_0) + [\boldsymbol{R}(-\beta; Q_Y) - R]_+\right\}, \tag{35}$$

and

$$E_{\rm FA} \stackrel{\Delta}{=} \min\{E_A, E_B\}. \tag{36}$$

[4]Conceptually, $\boldsymbol{R}(D, Q_Y)$ can be thought of as the rate–distortion function of the "source" P subject to a constrained reproduction distribution Q_Y (or vice versa), but note that the "distortion measure" $d(x,y)$ here is not necessarily non–negative for all (x,y).

The inverse function of $\boldsymbol{R}(D;Q_Y)$, will be denoted by $\boldsymbol{D}(R;Q_Y)$, i.e.,

$$\boldsymbol{D}(R;Q_Y) = \inf_{\{Q_{Y|X}\colon\ I(Q)\le R,\ (P\times Q_{Y|X})_Y=Q_Y\}} D(Q). \tag{37}$$

Also, let $R_1(Q_Y)$ and $D_1(Q_Y)$ denote $I(Q^*)$ and $D(Q^*)$, where Q^* minimizes $I(Q)+D(Q)$. Now, let

$$E_{\rm MD} \stackrel{\Delta}{=} \inf \mathcal{D}(Q_{Y|X}\|W|P) \tag{38}$$

where the infimum is subject to the constraints:

1. $\boldsymbol{D}(R;Q_Y) \le [\alpha]_+ - \beta \le D(P\times Q_{Y|X})$
2. $D_1(Q_Y) \le [\alpha]_+ - \beta$ implies $\boldsymbol{R}([\alpha]_+ - \beta; Q_Y) \ge R - [-\alpha]_+$
3. $D_1(Q_Y) > [\alpha]_+ - \beta$ implies $R_1(Q_Y) + D_1(Q_Y) \ge R + \alpha - \beta$

with $Q_Y = (P\times Q_{Y|X})_Y$. Next define

$$E_1 = \inf_{\{Q_{Y|X}\colon\ D(P\times Q_{Y|X})\le[\alpha]_+-\beta\}} \Big\{\mathcal{D}(Q_{Y|X}\|W|P) + \big[\boldsymbol{R}(D(P\times Q_{Y|X}); (P\times Q_{Y|X})_Y) - R\big]_+\Big\}, \tag{39}$$

$$E_2 = \inf_{Q_{Y|X}} \Big\{\mathcal{D}(Q_{Y|X}\|W|P) + \big[\boldsymbol{R}(\alpha-\beta; (P\times Q_{Y|X})_Y) - R\big]_+\Big\}, \tag{40}$$

and finally,

$$E_{\rm DE} \stackrel{\Delta}{=} \min\{E_1, E_2, E_{\rm MD}\}. \tag{41}$$

Theorem 1 *Let W be a DMC and let $\mathcal{R}^*$ be both defined as in Section 2.2. Let the codewords of $\mathcal{C} = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M\}$, $M = e^{nR}$, be selected independently at random under the uniform distribution across a given type class $\mathcal{T}_P$. Then, the asymptotic exponents associated with $\bar{P}_{FA}$, $\bar{P}_{MD}$, and $\bar{P}_{DE}$ are given, respectively, by E_{FA}, E_{MD}, and E_{DE}, as defined in eqs. (36), (38), and (41).*

Discussion. As discussed in Section 3, we observe that for $\alpha \ge 0$, all three exponents depend on α and β only via the difference $\alpha-\beta$. It is also seen that there is nothing to lose by replacing a positive value of α by $\alpha = 0$, as long as the difference $\alpha-\beta$ is kept. For $\alpha < 0$, the various exponents depend on α and β individually, so there are two degrees of freedom to adjust both the FA and the MD exponents to pre–specified values in a certain range.

It is instructive to find out the maximum achievable information rate for which the average probability of decoding error still tends to zero, that is, the smallest rate R for which $E_{\rm DE}=0$, for given $E_{\rm MD}$ and $E_{\rm FA}$. This happens as soon as either $E_1=0$ or $E_2=0$. The exponent E_1 vanishes for $R=\boldsymbol{R}(D(P\times W);(P\times W)_Y)$. But

$$\begin{aligned}\boldsymbol{R}(P\times W;(P\times W)_Y) &= \min\{I(Q):\ D(Q)\le D(P\times W),\ (P\times Q_{Y|X})_Y=(P\times W)_Y\}\\ &\le I(P\times W).\end{aligned}\tag{42}$$

On the other hand, since $\mathcal{D}(Q_{Y|X}\|W|P)\ge 0$, it is easy to see that the constraint set $\{Q:\ D(Q)\le D(P\times W),\ (P\times Q_{Y|X})_Y=(P\times W)_Y\}$ is a subset of $\{Q:\ I(Q)\ge I(P\times W)\}$, and so,

$$\boldsymbol{R}(P\times W;(P\times W)_Y)\ge \min\{I(Q):\ I(Q)\ge I(P\times W)\}=I(P\times W),\tag{43}$$

therefore, $\boldsymbol{R}(P\times W;(P\times W)_Y)=I(P\times W)$, which is the ordinary achievable rate one would expect from a constant composition code of type class $\mathcal{T}_P$. The exponent E_2 vanishes at the rate $\boldsymbol{R}(\alpha-\beta;(P\times W)_Y)$ Therefore, there is no rate loss, compared to ordinary decoding, as long as

$$\alpha-\beta\le D(P\times W).\tag{44}$$

5 Proof of Theorem 1

This section is divided into three subsections, each one devoted to the analysis of one of the three error exponents.

5.1 The False Alarm Error Exponent

Let $\boldsymbol{y}$ be given and consider $\{\boldsymbol{X}_m\}$ as random. Then,

$$\bar{P}_{\rm FA}(\boldsymbol{y}) = Q_0\left\{e^{n\alpha}\cdot\sum_{m=1}^{M}W(\boldsymbol{y}|\boldsymbol{X}_m)+\max_m W(\boldsymbol{y}|\boldsymbol{X}_m)>e^{n\beta}Q_0(\boldsymbol{y})\right\}\tag{45}$$

$$\dot{=}\ Q_0\left\{e^{n\alpha}\cdot\sum_{m=1}^{M}W(\boldsymbol{y}|\boldsymbol{X}_m)>e^{n\beta}Q_\star(\boldsymbol{y})\right\}+Q_0\left\{\max_m W(\boldsymbol{y}|\boldsymbol{X}_m)>e^{n\beta}Q_\star(\boldsymbol{y})\right\}\tag{46}$$

$$=\ Q_0\left\{\sum_{m=1}^{M}W(\boldsymbol{y}|\boldsymbol{X}_m)>e^{n(\beta-\alpha)}Q_0(\boldsymbol{y})\right\}+Q_0\left\{\max_m W(\boldsymbol{y}|\boldsymbol{X}_m)>e^{n\beta}Q_\star(\boldsymbol{y})\right\}\tag{47}$$

$$\stackrel{\Delta}{=}\ A(\boldsymbol{y})+B(\boldsymbol{y}),\tag{48}$$

where we have used (3). It is sufficient now to show that $A = \boldsymbol{E}\{A(\boldsymbol{Y})\} \doteq e^{-nE_A}$ and $B = \boldsymbol{E}\{B(\boldsymbol{Y})\} \doteq e^{-nE_B}$. Now, for a given $\boldsymbol{y}$, let $N(\hat{Q}|\boldsymbol{y})$ be the number of codewords in $\mathcal{C}$ whose joint empirical distribution with $\boldsymbol{y}$ is $\hat{Q} = \{\hat{Q}(x,y),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$. Next, define

$$f(\hat{Q}) = \sum_{x,y} \hat{Q}(x,y) \ln W(y|x) \tag{49}$$

and

$$g(\hat{Q}_Y) = \sum_{y} \hat{Q}_Y(y) \ln Q_\star(y) + \beta - \alpha. \tag{50}$$

Then,

$$\begin{aligned}
A(\boldsymbol{y}) &= Q_0 \left\{ \sum_{\hat{Q}_{X|Y}} N(\hat{Q}|\boldsymbol{y}) e^{nf(\hat{Q})} > e^{ng(\hat{Q}_Y)} \right\} \\
&\doteq Q_0 \left\{ \max_{\hat{Q}_{X|Y}} N(\hat{Q}|\boldsymbol{y}) e^{nf(\hat{Q})} > e^{ng(\hat{Q}_Y)} \right\} && (51) \\
&= Q_0 \bigcup_{\hat{Q}_{X|Y}} \left\{ N(\hat{Q}|\boldsymbol{y}) e^{nf(\hat{Q})} > e^{ng(\hat{Q}_Y)} \right\} && (52) \\
&\doteq \sum_{\hat{Q}_{X|Y}} Q_0 \left\{ N(\hat{Q}|\boldsymbol{y}) > e^{n[g(\hat{Q}_Y) - f(\hat{Q})} \right\} && (53) \\
&\doteq \max_{\hat{Q}_{X|Y}} Q_0 \left\{ N(\hat{Q}|\boldsymbol{y}) > e^{nu(\hat{Q})} \right\}, && (54)
\end{aligned}$$

where we have used again eq. (3) and where we have defined

$$u(\hat{Q}) \stackrel{\Delta}{=} g(\hat{Q}_Y) - f(\hat{Q}) = \sum_{x,y \in \mathcal{X} \times \mathcal{Y}} \hat{Q}(x,y) \ln \frac{Q_0(y)}{W(y|x)} + \beta - \alpha = D(\hat{Q}) + \beta - \alpha. \tag{55}$$

Now, since $N(\hat{Q}|\boldsymbol{y})$ is a Bernoulli random variable pertaining to e^{nR} trials and probability of success of the exponential order of $e^{-nI(\hat{Q})}$, we have, similarly as in [6, Subsection 6.3]

$$\Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{nu(\hat{Q})}\} \doteq \exp\left\{ -e^{n[u(\hat{Q})]_+} (n[I(\hat{Q}) - R + [u(Q)]_+] - 1) \right\}, \tag{56}$$

provided that for $u(\hat{Q}) > 0$, $I(\hat{Q}) - R + u(\hat{Q}) > 0$ (otherwise, $\Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{nu(\hat{Q})}\} \to 1$).[5] Therefore, the exponential rate $E(\hat{Q})$ of $\Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{nu(\hat{Q})}\}$ is as follows:

$$E(\hat{Q}) = \begin{cases} [I(\hat{Q}) - R]_+ & u(\hat{Q}) \leq 0 \\ \infty & u(\hat{Q}) > 0,\ u(\hat{Q}) > R - I(\hat{Q}) \\ 0 & u(\hat{Q}) > 0,\ u(\hat{Q}) < R - I(\hat{Q}) \end{cases} \tag{57}$$

[5]Note also that $\Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{nu(\hat{Q})}\} = \Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{n[u(\hat{Q})]_+}\}$ since $N(\hat{Q}|\boldsymbol{y})$ is an integer valued random variable.

For a given $\hat{Q}_Y$, let $\mathcal{Q}_P$ be the set of $\{\hat{Q}_{X|Y}\}$ such that $(\hat{Q}_Y \times \hat{Q}_{X|Y})_X = P$. Then,

$$\begin{aligned}
\min_{\hat{Q}_{X|Y}\in\mathcal{Q}_P} E(\hat{Q}) &= \begin{cases} \infty & \forall \hat{Q}_{X|Y} \in \mathcal{Q}_P:\ u(\hat{Q}) > 0,\ u(\hat{Q}) > R - I(\hat{Q}) \\ 0 & \exists \hat{Q}_{X|Y} \in \mathcal{Q}_P:\ 0 \le u(\hat{Q}) \le R - I(\hat{Q}) \\ 0 & \exists \hat{Q}_{X|Y} \in \mathcal{Q}_P:\ u(\hat{Q}) \le 0,\ I(\hat{Q}) \le R \\ \min_{\{\hat{Q}_{X|Y}\in\mathcal{Q}_P:\ u(\hat{Q})\le 0\}} [I(\hat{Q}) - R]_+ & \text{otherwise} \end{cases} \\
&= \begin{cases} \infty & \forall \hat{Q}_{X|Y} \in \mathcal{Q}_P:\ u(\hat{Q}) > [R - I(\hat{Q})]_+ \\ 0 & \exists \hat{Q}_{X|Y} \in \mathcal{Q}_P:\ I(\hat{Q}) \le \min\{R, R - u(\hat{Q})\} \\ \min_{\{\hat{Q}_{X|Y}\in\mathcal{Q}_P:\ u(\hat{Q})\le 0\}} [I(\hat{Q}) - R]_+ & \text{otherwise} \end{cases}
\end{aligned}$$

The condition for $\min_{\hat{Q}_{X|Y}\in\mathcal{Q}_P} E(\hat{Q})$ to vanish becomes

$$\begin{aligned}
\alpha - \beta + R &\ge \mu(\hat{Q}_Y, R) = \min_{\{\hat{Q}_{X|Y}\in\mathcal{Q}_P:\ I(\hat{Q})\le R\}} [I(\hat{Q}) + D(\hat{Q})] \\
&= \begin{cases} R + \boldsymbol{D}(R; \hat{Q}_Y) & R < R_1(\hat{Q}_Y) \\ R_1(\hat{Q}_Y) + D_1(\hat{Q}_Y) & R \ge R_1(\hat{Q}_Y) \end{cases} \qquad (58)
\end{aligned}$$

The condition for an infinite exponent is as follows: For $u(\hat{Q})$ to be non-negative for all $\hat{Q}_{X|Y}$, we need

$$\alpha - \beta \le D_{\min}(\hat{Q}_Y) \stackrel{\Delta}{=} \min_{\hat{Q}_{X|Y}\in\mathcal{Q}_P} D(\hat{Q}). \qquad (59)$$

For $u(\hat{Q}) \ge R - I(\hat{Q})$ for all $\hat{Q}_{X|Y} \in \mathcal{Q}_P$, we need $\alpha - \beta + R < \mu(\hat{Q}_Y, \infty)$. Thus, in summary,

$$\begin{aligned}
\min_{\hat{Q}_{X|Y}\in\mathcal{Q}_P} E(\hat{Q}) &= \begin{cases} 0 & \alpha - \beta \ge \mu(\hat{Q}_Y, R) - R \\ \infty & \alpha - \beta < \min\{\mu(\hat{Q}_Y, \infty) - R, D_{\min}(\hat{Q}_Y)\} \\ \min_{\{\hat{Q}_{X|Y}\in\mathcal{Q}_P:\ u(\hat{Q})\le 0\}} [I(\hat{Q}) - R]_+ & \text{elsewhere} \end{cases} \\
&= \begin{cases} 0 & \alpha - \beta \ge \mu(\hat{Q}_Y, R) - R \\ \infty & \alpha - \beta < \min\{\mu(\hat{Q}_Y, \infty) - R, D_{\min}(\hat{Q}_Y)\} \\ \Big[\boldsymbol{R}(\alpha - \beta; \hat{Q}_Y) - R\Big]_+ & \text{elsewhere} \end{cases} \\
&= \begin{cases} \boldsymbol{R}(\alpha - \beta; \hat{Q}_Y) - R & \alpha - \beta < \mu(\hat{Q}_Y, R) - R \\ 0 & \alpha - \beta \ge \mu(\hat{Q}_Y, R) - R \end{cases} \\
&= \tilde{\boldsymbol{R}}(\alpha - \beta, R; \hat{Q}_Y), \qquad (60)
\end{aligned}$$

where we have used the convention that the minimum over an empty set is infinity and the fact that $\boldsymbol{D}(R; \hat{Q}_Y) \ge \mu(Q_Y, R) - R$. For the overall exponent associated with A, we need to average over $\boldsymbol{Y}$, which gives $A \doteq e^{-nE_A}$ with

$$E_A = \min_{Q_Y} \{\mathcal{D}(Q_Y \| Q_0) + \tilde{\boldsymbol{R}}(\alpha - \beta, R; Q_Y)\}. \qquad (61)$$

Moving on to the analysis of $B(\boldsymbol{y})$,

$$\begin{aligned} B(\boldsymbol{y}) &= Q_0\left\{\max_m W(\boldsymbol{y}|\boldsymbol{X}_m) > e^{n\beta}Q_0(\boldsymbol{y})\right\} && (62)\\ &= Q_0\bigcup_{m=1}^{M}\left\{W(\boldsymbol{y}|\boldsymbol{X}_m) > e^{n\beta}Q_0(\boldsymbol{y})\right\} && (63)\\ &\doteq \min\left\{1, M\cdot Q_0\{W(\boldsymbol{y}|\boldsymbol{X}_1) > e^{n\beta}Q_0(\boldsymbol{y})\}\right\}, && (64)\end{aligned}$$

where in the last line, we have used (5). Now,

$$Q_0\{W(\boldsymbol{y}|\boldsymbol{X}_1) > e^{n\beta}Q_0(\boldsymbol{y})\} \doteq e^{-nI_0(\hat{Q}_Y)}, \tag{65}$$

where

$$\begin{aligned} I_0(\hat{Q}_Y) &= \min_{\hat{Q}_{X|Y}}\left\{I(\hat{Q}):\ D(\hat{Q})\le -\beta,\ \ \hat{Q}_{X|Y}\in\mathcal{Q}_P\right\}\\ &= \boldsymbol{R}(-\beta;\hat{Q}_Y). && (66)\end{aligned}$$

Thus, $B\doteq e^{-nE_B}$ with

$$E_B = \min_{Q_Y}\{\mathcal{D}(Q_Y\|Q_0) + [\boldsymbol{R}(-\beta;Q_Y) - R]_+\}. \tag{67}$$

5.2 The Misdetection Error Exponent

Without loss of generality, we will assume that $\boldsymbol{X}_1 = \boldsymbol{x}_1$ was transmitted. We first condition on $\boldsymbol{x}_1$ and $\boldsymbol{y}$.

$$\begin{aligned} \bar{P}_{\rm MD}(\boldsymbol{x}_1,\boldsymbol{y}) &= \Pr\left\{e^{n\alpha}\sum_{m=1}^{M}W(\boldsymbol{y}|\boldsymbol{X}_m) + \max_m W(\boldsymbol{y}|\boldsymbol{X}_m) \le e^{n\beta}Q_\star(\boldsymbol{y})\Big|\boldsymbol{X}_1=\boldsymbol{x}_1,\boldsymbol{y}\right\}\\ &= \Pr\left\{e^{n\alpha}\sum_{m=1}^{M}W(\boldsymbol{y}|\boldsymbol{X}_m)+\right.\\ &\qquad \left.\max\{W(\boldsymbol{y}|\boldsymbol{x}_1), \max_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m)\}\le e^{n\beta}Q_\star(\boldsymbol{y})\Big|\boldsymbol{X}_1=\boldsymbol{x}_1,\boldsymbol{y}\right\}\\ &\doteq \Pr\left\{e^{n\alpha}\left[W(\boldsymbol{y}|\boldsymbol{x}_1)+\sum_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m)\right] + W(\boldsymbol{y}|\boldsymbol{x}_1) + \max_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m)\le e^{n\beta}Q_\star(\boldsymbol{y})\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}\\ &\doteq \Pr\left\{e^{n[\alpha]_+}W(\boldsymbol{y}|\boldsymbol{x}_1) + e^{n\alpha}\sum_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m) + \max_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m)\le e^{n\beta}Q_\star(\boldsymbol{y})\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}\\ &\doteq \Pr\left\{e^{n[\alpha]_+}W(\boldsymbol{y}|\boldsymbol{x}_1) < e^{n\beta}Q_\star(\boldsymbol{y}), e^{n\alpha}\sum_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m) + \max_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m)\le e^{n\beta}Q_\star(\boldsymbol{y})\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}\end{aligned}$$

$$
\begin{aligned}
&= \mathcal{I}\left\{e^{n[\alpha]^+}W(\boldsymbol{y}|\boldsymbol{x}_1) < e^{n\beta}Q_\star(\boldsymbol{y})\right\} \times \\
&\quad \Pr\left\{e^{n\alpha}\sum_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m) + \max_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m) \le e^{n\beta}Q_\star(\boldsymbol{y})\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} \\
&\stackrel{\Delta}{=} C\cdot D. \qquad (68)
\end{aligned}
$$

Using the identity

$$\max_{m>1}W(\boldsymbol{y}|\boldsymbol{x}_m) \equiv \max_{\hat{Q}_{X|Y}}\mathcal{I}\{N(\hat{Q}|\boldsymbol{y})\ge 1\}\cdot e^{nf(\hat{Q})} \qquad (69)$$

(where now $N(\hat{Q}|\boldsymbol{y})$ does not count $\boldsymbol{x}_1$), we now have

$$
\begin{aligned}
D &= \Pr\left\{e^{n\alpha}\sum_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m) + \max_{m>1}W(\boldsymbol{y}|\boldsymbol{X}_m) \le e^{n\beta}Q_0(\boldsymbol{y})\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} && (70)\\
&= \Pr\left\{e^{n\alpha}\sum_{\hat{Q}_{X|Y}}N(\hat{Q}|\boldsymbol{y})e^{nf(\hat{Q})} + \max_{\hat{Q}_{X|Y}}\mathcal{I}\{N(\hat{Q}|\boldsymbol{y})\ge 1\}\cdot e^{nf(\hat{Q})} \le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} && (71)\\
&\dot{=} \Pr\left\{e^{n\alpha}\sum_{\hat{Q}_{X|Y}}N(\hat{Q}|\boldsymbol{y})e^{nf(\hat{Q})} + \sum_{\hat{Q}_{X|Y}}\mathcal{I}\{N(\hat{Q}|\boldsymbol{y})\ge 1\}e^{nf(\hat{Q})} \le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} && (72)\\
&= \Pr\left\{\sum_{\hat{Q}_{X|Y}}[e^{n\alpha}N(\hat{Q}|\boldsymbol{y}) + \mathcal{I}\{N(\hat{Q}|\boldsymbol{y})\ge 1\}]e^{nf(\hat{Q})} \le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} && (73)\\
&\dot{=} \Pr\left\{\max_{\hat{Q}_{X|Y}}[e^{n\alpha}N(\hat{Q}|\boldsymbol{y}) + \mathcal{I}\{N(\hat{Q}|\boldsymbol{y})\ge 1\}]e^{nf(\hat{Q})} \le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} && (74)\\
&= \Pr\bigcap_{\hat{Q}_{X|Y}}\left\{e^{n\alpha}N(\hat{Q}|\boldsymbol{y}) + \mathcal{I}\{N(\hat{Q}|\boldsymbol{y})\ge 1\} \le e^{n[u(\hat{Q})+\alpha]}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} && (75)\\
&= \Pr\bigcap_{\hat{Q}_{X|Y}}\left\{N(\hat{Q}|\boldsymbol{y}) \le e^{nv(\hat{Q})}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}, && (76)
\end{aligned}
$$

where

$$v(\hat{Q}) = \begin{cases} u(\hat{Q}) & u(\hat{Q})+\alpha > 0 \\ -\infty & u(\hat{Q})+\alpha \le 0 \end{cases} \qquad (77)$$

Now, if there exists at least one $\hat{Q}_{X|Y}\in\mathcal{Q}_P$ for which $I(\hat{Q}) < R$ and $R - I(\hat{Q}) > v(\hat{Q})$, then this $\hat{Q}_{X|Y}$ alone is responsible for a double exponential decay of D (because then the event in question would be a large deviations event whose probability decays exponentially with $M = e^{nR}$, thus double–exponentially with n), let alone the intersection over all $\{\hat{Q}_{X|Y}\}$. The condition for this to happen is $R > R_0(\hat{Q}_Y) \stackrel{\Delta}{=} \min_{Q_{X|Y}\in\mathcal{Q}_P}\max\{I(\hat{Q}), I(\hat{Q}) + v(\hat{Q})\}$. Conversely, if for every $\hat{Q}$ with

$\hat{Q}_{X|Y} \in \mathcal{Q}_P$, we have $I(\hat{Q}) > R$ or $R - I(\hat{Q}) < v(\hat{Q})$, that is, $R < R_0(\hat{Q}_Y)$, then D is close to 1 since the intersection is over a sub–exponential number of events with very high probability. It follows that D behaves like $\mathcal{I}\{R_0(\hat{Q}_Y) > R\}$, Thus,

$$\begin{aligned} P_{\rm MD} &\doteq \boldsymbol{E}\mathcal{I}\Big\{R_0(\hat{Q}_Y) > R,\ W(\boldsymbol{Y}|\boldsymbol{X}_1) \le e^{n(\beta-[\alpha]_+)}Q_0(\boldsymbol{Y})\Big\} \\ &= \exp\left[-n \inf_{Q_{Y|X}\in\mathcal{Q}_P}\big\{\mathcal{D}(Q_{Y|X}\|W|P):\ R_0(Q_Y) > R,\ D(Q) > [\alpha]_+ - \beta\big\}\right]. \end{aligned} \tag{78}$$

Now, let us take a closer look at $R_0(Q_Y)$:

$$\begin{aligned} \max\{I(Q), I(Q)+v(Q)\} &= \begin{cases} \max\{I(Q), I(Q)+u(Q)\} & u(Q) > -\alpha \\ I(Q) & u(Q) \le -\alpha \end{cases} && (79) \\ &= I(Q) + u(Q)\cdot\mathcal{I}\{u(Q) > [-\alpha]_+\}. && (80) \end{aligned}$$

Thus,

$$\begin{aligned} R_0(Q) &= \min_{Q_{X|Y}\in\mathcal{Q}_P}[I(Q) + u(Q)\cdot\mathcal{I}\{u(Q) > [-\alpha]_+\}] && (81) \\ &= \min\left\{\min_{Q_{X|Y}\in\mathcal{Q}_P:\ u(Q)\le[-\alpha]_+} I(Q), \min_{Q_{X|Y}\in\mathcal{Q}_P:\ u(Q)>[-\alpha]_+}[I(Q)+u(Q)]\right\}. && (82) \end{aligned}$$

Now,

$$\begin{aligned} \min_{Q_{X|Y}\in\mathcal{Q}_P:\ u(Q)\le[-\alpha]_+} I(Q) &= \boldsymbol{R}(\alpha + [-\alpha]_+ - \beta; Q_Y) && (83) \\ &= \boldsymbol{R}([\alpha]_+ - \beta; Q_Y) && (84) \end{aligned}$$

and

$$\begin{aligned} &\min_{Q_{X|Y}\in\mathcal{Q}_P:\ u(Q)>[-\alpha]_+}[I(Q)+u(Q)] && (85) \\ &= \beta - \alpha + \min_{Q_{X|Y}\in\mathcal{Q}_P:\ D(Q)>[\alpha]_+-\beta}[I(Q)+D(Q)] && (86) \\ &= \beta - \alpha + \begin{cases} R_1(Q_Y) + D_1(Q_Y) & [\alpha]_+ - \beta < D_1(Q_Y) \\ \boldsymbol{R}([\alpha]_+ - \beta; Q_Y) + [\alpha]_+ - \beta & \text{otherwise} \end{cases} && (87) \\ &= \begin{cases} R_1(Q_Y) + D_1(Q_Y) + \beta - \alpha & [\alpha]_+ - \beta < D_1(Q_Y) \\ \boldsymbol{R}([\alpha]_+ - \beta; Q_Y) + [\alpha]_+ - \alpha & \text{otherwise} \end{cases} && (88) \\ &= \begin{cases} R_1(Q_Y) + D_1(Q_Y) + \beta - \alpha & [\alpha]_+ - \beta < D_1(Q_Y) \\ \boldsymbol{R}([\alpha]_+ - \beta; Q_Y) + [-\alpha]_+ & \text{otherwise} \end{cases} && (89) \end{aligned}$$

Thus,

$$E_{\rm MD} = \inf \mathcal{D}(Q_{Y|X}\|W|P), \tag{90}$$

where the infimum is over all $\{Q_{Y|X}\}$ that satisfies the following conditions:

1. $\boldsymbol{D}(R;Q_Y) \leq [\alpha]_+ - \beta \leq D(P \times Q_{Y|X})$

2. $D_1(Q_Y) \leq [\alpha]_+ - \beta$ implies $\boldsymbol{R}([\alpha]_+ - \beta; Q_Y) \geq R - [-\alpha]_+$

3. $D_1(Q_Y) > [\alpha]_+ - \beta$ implies $R_1(Q_Y) + D_1(Q_Y) \geq R + \alpha - \beta$

where $Q_Y = (P \times Q_{Y|X})_Y$.

5.3 The Decoding Error Exponent

Let us denote

$$\Omega_m \stackrel{\Delta}{=} \left\{ \boldsymbol{y} : \; W(\boldsymbol{y}|\boldsymbol{x}_m) > \max_{k \neq m} W(\boldsymbol{y}|\boldsymbol{x}_k) \right\}. \tag{91}$$

Then, for $m \geq 1$, $\mathcal{R}_m^* = \overline{\mathcal{R}_0^*} \cap \Omega_m$. For a given code, the probability of decoding error is given by

$$\begin{aligned} P_{\text{DE}} &= \frac{1}{M} \sum_{m=1}^{M} W(\overline{\mathcal{R}_m^*}|\boldsymbol{x}_m) && (92) \\ &= \frac{1}{M} \sum_{m=1}^{M} W(\mathcal{R}_0^* \cup \overline{\Omega_m}|\boldsymbol{x}_m) && (93) \\ &= \frac{1}{M} \sum_{m=1}^{M} W(\overline{\mathcal{R}_0^*} \cap \overline{\Omega_m}|\boldsymbol{x}_m) + \frac{1}{M} \sum_{m=1}^{M} W(\mathcal{R}_0^*|\boldsymbol{x}_m). && (94) \end{aligned}$$

Upon taking the ensemble average, the second term becomes $\bar{P}_{\text{MD}}$, which we have already analyzed in the previous subsection. Its error exponent, E_{MD}, indeed appears as one of the arguments of the $\min\{\cdot\}$ operator in eq. (41), and so, it remains to show that the exponent of the ensemble average of the first term is $\min\{E_1, E_2\}$. Let $\boldsymbol{X}_1 = \boldsymbol{x}_1$ be transmitted and let $\boldsymbol{Y} = \boldsymbol{y}$ be received. As before, we first condition on $(\boldsymbol{x}_1, \boldsymbol{y})$.

$$\begin{aligned} \Pr\{\overline{\mathcal{R}_0^*} \cap \overline{\Omega_1}|\boldsymbol{x}_1, \boldsymbol{y}\} &= \Pr\left\{ e^{n\alpha} \sum_m W(\boldsymbol{y}|\boldsymbol{X}_m) + \max_m W(\boldsymbol{y}|\boldsymbol{X}_m) > e^{n\beta} Q_\star(\boldsymbol{y}), \right. \\ &\qquad \left. \max_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) \geq W(\boldsymbol{y}|\boldsymbol{x}_1) \middle| \boldsymbol{x}_1, \boldsymbol{y} \right\} \\ &\doteq \Pr\left\{ e^{n[\alpha]_+} W(\boldsymbol{y}|\boldsymbol{x}_1) + e^{n\alpha} \sum_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) + \max_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) > e^{n\beta} Q_\star(\boldsymbol{y}), \right. \\ &\qquad \left. \max_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) \geq W(\boldsymbol{y}|\boldsymbol{x}_1) \middle| \boldsymbol{x}_1, \boldsymbol{y} \right\} \\ &\doteq A(\boldsymbol{x}_1, \boldsymbol{y}) + B(\boldsymbol{x}_1, \boldsymbol{y}) + C(\boldsymbol{x}_1, \boldsymbol{y}) \end{aligned} \tag{95}$$

where

$$A(\boldsymbol{x}_1,\boldsymbol{y}) = \mathcal{I}\left\{W(\boldsymbol{y}|\boldsymbol{x}_1) \geq e^{n(\beta-[\alpha]_+)}Q_\star(\boldsymbol{y})\right\} \cdot \Pr\left\{\max_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) \geq W(\boldsymbol{y}|\boldsymbol{x}_1)\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}, \tag{96}$$

$$B(\boldsymbol{x}_1,\boldsymbol{y}) = \Pr\left\{\sum_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) \geq e^{n(\beta-\alpha)}Q_\star(\boldsymbol{y}),\ \max_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) \geq W(\boldsymbol{y}|\boldsymbol{x}_1)\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}, \tag{97}$$

and

$$C(\boldsymbol{x}_1,\boldsymbol{y}) = \Pr\left\{\max_{m>1} W(\boldsymbol{y}|\boldsymbol{X}_m) \geq \max\{e^{n\beta}Q_\star(\boldsymbol{y}),\ W(\boldsymbol{y}|\boldsymbol{x}_1)\}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}. \tag{98}$$

We next analyze each one of these terms. First, observe that for a given constant S (which may depend on the given $\boldsymbol{x}_1$ and $\boldsymbol{y}$), we have

$$\begin{aligned}
\Pr\left\{\max_{m>1}\frac{W(\boldsymbol{y}|\boldsymbol{X}_m)}{Q_\star(\boldsymbol{y})} \geq e^{-nS}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\} &\doteq \min\left\{1, e^{nR}\cdot\Pr\left\{\frac{W(\boldsymbol{y}|\boldsymbol{X}_2)}{Q_\star(\boldsymbol{y})} > e^{-nS}|\boldsymbol{x}_1,\boldsymbol{y}\right\}\right\} &(99)\\
&\doteq \min\left\{1, e^{nR}\cdot\Pr\left\{\frac{W(\boldsymbol{y}|\boldsymbol{X}_2)}{Q_\star(\boldsymbol{y})} > e^{-nS}|\boldsymbol{x}_1,\boldsymbol{y}\right\}\right\} &(100)\\
&\doteq \exp\{-n[\boldsymbol{R}(S,\hat{Q}_Y)-R]_+\}. &(101)
\end{aligned}$$

In our case, $S = D(\tilde{Q})$, where $\tilde{Q}$ is the empirical joint distribution of $\boldsymbol{x}_1$ and $\boldsymbol{y}$. Thus,

$$\begin{aligned}
A &\stackrel{\Delta}{=} \boldsymbol{E}\{A(\boldsymbol{X}_1,\boldsymbol{Y})\}\\
&\doteq \exp\left[-n\min_{\{Q:\ Q_{X|Y}\in\mathcal{Q}_P:\ D(Q)\leq[\alpha]_+-\beta\}}\{\mathcal{D}(Q_{Y|X}\|W|P)+[\boldsymbol{R}(D(Q),Q_Y)-R]_+\}\right]\\
&= e^{-nE_1}. &(102)
\end{aligned}$$

Concerning $C(\boldsymbol{x}_1,\boldsymbol{y})$, we similarly have:

$$\begin{aligned}
C(\boldsymbol{x}_1,\boldsymbol{y}) &= \Pr\left\{\max_{m>1}\frac{W(\boldsymbol{y}|\boldsymbol{X}_m)}{Q_\star(\boldsymbol{y})} \geq \max\left\{e^{n\beta},\ \frac{W(\boldsymbol{y}|\boldsymbol{x}_1)}{Q_\star(\boldsymbol{y})}\right\}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}\\
&\doteq \min\left\{1, e^{nR}\cdot\Pr\left\{\frac{W(\boldsymbol{y}|\boldsymbol{X}_2)}{Q_\star(\boldsymbol{y})} \geq \max\left\{e^{n\beta},\ \frac{W(\boldsymbol{y}|\boldsymbol{x}_1)}{Q_\star(\boldsymbol{y})}\right\}\Big|\boldsymbol{x}_1,\boldsymbol{y}\right\}\right\}\\
&\doteq \exp\left\{-n[\boldsymbol{R}(\min\{-\beta,D(\tilde{Q})\};\hat{Q}_y)-R]_+\right\}, &(103)
\end{aligned}$$

and so,

$$\begin{aligned}
C &\stackrel{\Delta}{=} \boldsymbol{E}\{C(\boldsymbol{X}_1,\boldsymbol{Y})\}\\
&\doteq \exp\left\{-n\min_{Q:\ Q_{X|Y}\in\mathcal{Q}_P}\{\mathcal{D}(Q_{Y|X}\|W|P)+[\boldsymbol{R}(\min\{-\beta,D(Q)\};Q_Y)-R]_+\}\right\}\\
&\dot{\leq} e^{-nE_1}, &(104)
\end{aligned}$$

therefore, C is always dominated by A. It remains then to show that $B = \boldsymbol{E}\{B(\boldsymbol{X}_1, \boldsymbol{Y})\} \doteq e^{-nE_2}$. First, for given $(\boldsymbol{x}_1, \boldsymbol{y})$,

$$
\begin{aligned}
B(\boldsymbol{x}_1, \boldsymbol{y}) &\doteq \Pr\left\{ \sum_{\hat{Q}_{X|Y}} N(\hat{Q}|\boldsymbol{y}) e^{nf(\hat{Q})} \geq e^{ng(\hat{Q})},\ \sum_{\hat{Q}_{X|Y}} \mathcal{I}\{N(\hat{Q}|\boldsymbol{y}) \geq 1\} \cdot e^{nf(\hat{Q})} \geq e^{nf(\tilde{Q})} \right\} \\
&\doteq \Pr\left\{ \max_{\hat{Q}_{X|Y}} N(\hat{Q}|\boldsymbol{y}) e^{nf(\hat{Q})} \geq e^{ng(\hat{Q})},\ \max_{\hat{Q}_{X|Y}} \mathcal{I}\{N(\hat{Q}|\boldsymbol{y}) \geq 1\} \cdot e^{nf(\hat{Q})} \geq e^{nf(\tilde{Q})} \right\} \\
&\doteq \Pr\left[\bigcup_{\hat{Q}_{X|Y}} \{N(\hat{Q}|\boldsymbol{y}) \geq e^{nu(\hat{Q})}\} \right] \bigcap \left[\bigcup_{\hat{Q}_{X|Y}} \{\mathcal{I}\{N(\hat{Q}|\boldsymbol{y}) \geq 1\} \geq e^{n[f(\tilde{Q}) - f(\hat{Q})]}\} \right] \\
&= \Pr\left[\bigcup_{\hat{Q}_{X|Y}} \{N(\hat{Q}|\boldsymbol{y}) \geq e^{nu(\hat{Q})}\} \right] \bigcap \left[\bigcup_{\hat{Q}_{X|Y}:\ f(\tilde{Q}) \leq f(\hat{Q})} \{\mathcal{I}\{N(\hat{Q}|\boldsymbol{y}) \geq 1\} \geq e^{n[f(\tilde{Q}) - f(\hat{Q})]}\} \right] \\
&= \Pr \bigcup_{\{\hat{Q}_{X|Y}, Q'_{X|Y}:\ f(\tilde{Q}) \leq f(Q')\}} \Big\{ N(\hat{Q}|\boldsymbol{y}) \geq e^{nu(\hat{Q})},\ N(Q'|\boldsymbol{y}) \geq 1 \Big\} \\
&\doteq \Pr \bigcup_{\hat{Q}} \Big\{ N(\hat{Q}|\boldsymbol{y}) \geq e^{n[u(\hat{Q})]_+} \Big\} + \\
&\qquad \sum_{\hat{Q}_{X|Y} \neq Q'_{X|Y}:\ f(\tilde{Q}) \leq f(Q')} \Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{n[u(\hat{Q})]_+},\ N(Q'|\boldsymbol{y}) \geq 1\} \\
&\doteq \max_{\hat{Q}_{X|Y}} \Pr \Big\{ N(\hat{Q}|\boldsymbol{y}) \geq e^{n[u(\hat{Q})]_+} \Big\} + \\
&\qquad \max_{\hat{Q}_{X|Y} \neq Q'_{X|Y}: f(\tilde{Q}) \leq f(Q')} \Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{n[u(\hat{Q})]_+},\ N(Q'|\boldsymbol{y}) \geq 1\} \\
&\doteq \max_{\hat{Q}_{X|Y}} \Pr\{N(\hat{Q}|\boldsymbol{y}) \geq e^{n[u(\hat{Q})]_+}\} \\
&= \exp\{-n[\boldsymbol{R}(\alpha - \beta; \hat{Q}_Y) - R]_+\}. \qquad (105)
\end{aligned}
$$

where the last passage follows from an analysis almost identical to that of E_A in Subsection 5.1. Thus,

$$
B = \boldsymbol{E}\{B(\boldsymbol{X}_1, \boldsymbol{Y})\} = \exp\{-n \min_{Q_{Y|X}} \{\mathcal{D}(Q_{Y|X} \| W | P) + [\boldsymbol{R}(\alpha - \beta; Q_Y) - R]_+\} = e^{-nE_2}. \qquad (106)
$$

References

[1] R. H. Barker, "Group synchronization of binary digital systems," *Communication Theory*. London: Butterworth. pp. 273–287, 1953.

[2] S. Golomb, J. Davey, I. Reed, H. van Trees, and J. Stiffer, "Syncronization," *IEEE Trans. on Communication Systems*, vol. 11, no. 4, pp. 481–491, 1963.

[3] G. D. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback systems," *IEEE Trans. Inform. Theory*, vol. IT–14, no. 2, pp. 206–220, March 1968.

[4] L. Franks, "Carrier and bit synchronization in data communication – a tutorial review," *IEEE Trans. on Communication Systems*, vol. 28, no. 8, pp. 1107–1121, 1980.

[5] J. Massey, "Optimum frame synchronization," *IEEE Trans. on Communications*, vol. 20, no. 2, pp. 115–119, 1972.

[6] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, nos. 1–2, pp. 1–212, 2009.

[7] R. Scholtz, "Frame synchronization techniques," *IEEE Trans. on Communication Systems*, vol. 28, no. 8, pp. 1204–1213, 1980.

[8] N. Shulman, *Communication over an Unknown Channel via Common Broadcasting*, Ph.D. dissertation, Department of Electrical Engineering – Systems, Tel Aviv University, July 2003. http://www.eng.tau.ac.il/∼shulman/papers/Nadav_PhD.pdf

[9] A. Somekh–Baruch and N. Merhav, "Achievable error exponents for the private fingerprinting game," *IEEE Trans. Inform. Theory*, vol. 53, no. 5, pp. 1827–1838, May 2007.

[10] A. Somekh–Baruch and N. Merhav, "Exact random coding exponents for erasure decoding," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, October 2011.

[11] A. Tchankerten, V. Chandar, and G. Wornell, "On the capacity region for asynchronous channels," *Proc. 2008 IEEE International Symposium on Information Theory*, pp. 1213–1217, 2008.

[12] A. Tchankerten, A. Khisti, and G. Wornell, "Information theoretic perspectives of synchronization," *Proc. 2006 IEEE International Symposium on Information Theory*, pp. 371–375, 2006.

[13] D. Wang, *Distinguishing Codes From Noise: Fundamental Limits and Applications to Sparse Communication*, Master thesis, Massachusetts Institute of Technology, Department of EECS, June 2010.

[14] D. Wang, V. Chandar, S.-Y. Chung, and G. Wornell, "Error exponents in asynchronous communication," *Proc. 2011 IEEE International Symposium on Information Theory*, pp. 1071–1075, 2011.