Codeword or Noise? Exact Random Coding Exponents for Joint Detection and Decoding

Nir Weinberger and Neri Merhav
Technion, Dept. of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
{nirwein@tx,merhav@ee}.technion.ac.il

Abstract

We consider the problem of coded communication, where in each time frame, 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 probabilities for the average code in the ensemble. It is shown that previously proposed decoders are in general strictly sub-optimal.

Index Terms

Joint detection/decoding, error exponent, false alarm, misdetection, random coding, synchronization.

I. Introduction

In the classical communication scenario studied in information theory, the decoder knows a-priori that a codeword is presently transmitted, and it is only required to decode it to one of the possible messages. However, in other scenarios, it might happen that for some of the time, the receiver does not observe channel output which corresponds to one of the possible codewords, but instead observes 'pure noise'. For example, the encoder may be 'silent' (non-transmitting) part of the time, because, e.g., it has no messages ready to be conveyed, or because, for some reason, it cannot currently transmit (it might be defective). In other cases, an interferer (or a jammer) may be present, which disrupts communication only part of the

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

time, but when this does occur, the channel output is completely dominated by the interferer. In all these examples, it might be of interest for the decoder to become aware of this special event. In continuation to our previous examples, the reception of pure noise may indicate that the transmitter should be repaired, or that other physical channel should be used, since the current one has detrimental interference. In other applications, it might be required from security motivations to be certain, with high probability, that one of the legal codewords in the codebook is transmitted. An additional central motivation, which we discuss below in more detail, is asynchronous communication.

In the above scenarios, the receiver has to be able to reliably detect the existence of the message, and only then decode it. The traditional approach has been to separate the problems of detection and coding/decoding, and to use a special pattern of letters to mark the beginning of a message transmission. The transmission of this pattern is, however, an undesired overhead.

This problem of joint detection/decoding belongs to a larger class of hypothesis testing problems, in which after performing the test, another task should be performed, depending on the chosen hypothesis. For example, in [17], [18], the problem of joint hypothesis testing and Bayesian estimation was considered, and in [15] the subsequent task is lossless source coding. A common theme for all the problems in this class, is that separately optimizing the detection and the task is sub-optimal, and so, joint optimization is beneficial.

In addition, joint detection/decoding is related to the *message-wise unequal error protection* (UEP) problem [2], in which one of the messages requires special attention, and should be more protected than all the other messages. In our case, the rejection region plays a role that is analogous to the role of the decision region of the preferred codeword in the UEP problem, but the difference is that in the UEP problem, the codeword of the protected message is under the control of the system designer. However, as we discuss in Section VII, a considerable part of our results is also suitable for the UEP problem.

As mentioned earlier, one of the motivations for the problem described above is the intimately related, long-standing, problem of *synchronization*, which has been studied extensively in the communications community throughout several decades (see, e.g., [1], [9], [10], [12], [20], and references therein, for a non-exhaustive sample of earlier works). In the synchronization problem, the receiver has to perform three tasks[1]: (i) to decide on the existence of a codeword, (ii) to locate the starting time instant of the message, and (iii) to decode it. Here too, the traditional approach is to separate the problems of synchronization and coding/decoding.

Recently, a model was proposed to study the fundamental limits of asynchronous communication [4],

[1]In practical communication systems, the receiver usually also needs to acquire symbol and carrier synchronization [9]. In our simplified discrete time, discrete input/output alphabet model, we assume, as in [26], [27] that such synchronization is not needed, or previously obtained by some other mechanisms.

[5], [24], [25], which includes tasks (ii) and (iii) above. In this model, it is known that a codeword of block length n will be transmitted, but the starting time of the codeword is only known to lie within a time window of $A = e^{n\tau}$ symbols, where $\tau \geq 0$, is termed the asynchronism exponent. In all other time points, the transmitter is idle, and the receiver observes pure noise. A sequential decoder was proposed in order to locate the codeword and then decode it. In this line of research, the goal is to characterize the region of achievable rates and asynchronism exponents. [2]

In an other line of research, [26], [27], performing tasks (i) and (iii) above is considered, as a simplified model for the synchronization problem, namely, assuming that task (ii), of properly locating the codeword, is possible without affecting other figures of merit of the system. According to this model, which is termed *slotted asynchronism*, 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. It may be easily observed that this model for asynchronism and the problem of detecting the existence of a codeword discussed above are essentially the same. Thus, henceforth we treat the problems on the same footing.

In [26], [27], 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 [26] and Wang *et al.* [27] 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 [27] that this general problem is open, and so, the focus both in [26], [27] 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, only that it decreases to zero as the block length increases. 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 FA

[2]It should be remarked that the rate in these work is measured in a non-standard way, as the *reaction delay* of the decoder, until it decides on a codeword.

exponent is exact. As a side note, we remark the maximal FA exponent is related to the asynchronism exponent discussed above, since if a FA exponent of τ is achieved for a given rate R, the loosely speaking, a asynchronism level of $A = e^{n\tau}$ is possible to accommodate, with a vanishing error probability. In this sense, our results, as well as [26], [27] may be partially compared with [4], [5], [24], [25].[3]

In this paper, 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 significantly different from the one in the achievability parts of [26], [27] (cf. the discussion in Section III) In particular, denoting the codewords by $\{\mathbf{x}_m\}$, the channel output vector by $\mathbf{y}$ (all of length n), and the channel conditional probability by $W(\mathbf{y}|\mathbf{x}_m)$, then according to this rule, a transmission is detected iff

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

where $\mathbf{0}$ is the all-zero channel input vector, and α and β are chosen to meet the MD and FA constraints. Of course, whenever the received $\mathbf{y}$ 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., [14], [23] 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 II, we establish some notation conventions, provide some preliminaries, and also formulate the problem. In Section III, we derive the optimum detector/decoder and discuss some of its properties. In Section IV, we present our main theorem, which is about single-letter formulas for the various error exponents. In Section V, we prove this theorem, and in Section VI we provide numerical examples of the results obtained, as well as numerical comparison with the results of [26]. Finally, in Section VII, we provide directions for future research. In Appendix A, we discuss computational aspects of the various exponent functions derived in this paper.

[3]Nonetheless, we do not pursue this direction here, mainly because the models are different, and in [4], [5], [24], [25], no special attention is given for error exponents.

II. Notation Conventions, Preliminaries and Problem Formulation

A. 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 $\mathbf{X} = (X_1, \ldots, X_n)$, ($n$ - positive integer) may take a specific vector value $\mathbf{x} = (x_1, \ldots, x_n)$ 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 $\mathbf{x}$, let $\hat{Q}_X$ denote[4] 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 $\mathbf{x}$. Let $\mathcal{T}_P$ denote the type class associated with P, that is, the set of all sequences $\{\mathbf{x}\}$ for which $\hat{Q}_X = P$. Similarly, for a pair of vectors $(\mathbf{x}, \mathbf{y})$, 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 $(\mathbf{x}, \mathbf{y})$ 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 $\mathcal{D}(\hat{Q}_X \| P)$, 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,

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

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 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 probability of an event $\mathcal{A}$ will be denoted by $\Pr\{\mathcal{A}\}$, and the expectation operator will be denoted by $\mathbf{E}\{\cdot\}$. Whenever there is room for ambiguity, the underlying probability distribution will appear as a subscript, e.g., $\mathbf{E}_Q\{\cdot\}$. 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

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

$\limsup_{n\to\infty} \frac{1}{n}\log(\frac{a_n}{b_n}) \leq 0$, and so on. Throughout the sequel, we will make frequent use of the fact that $\sum_{i=1}^{k_n} a_n(i) \doteq \max_{1\leq i\leq k_n} a_n(i)$ as long as $\{a_n(i)\}$ are positive and $k_n \doteq 1$. Accordingly, for k_n sequences of positive random variables $\{A_n(i)\}$, all defined on a common probability space, and a deterministic sequence b_n,

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

provided that $b'_n \doteq b_n$ implies $\Pr\{A_n(i) \geq b'_n\} \doteq \Pr\{A_n(i) \geq b_n\}$.[5] 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,

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

Another fact that will be used extensively is that for a given set of M pairwise independent events $\{\mathcal{A}_i\}_{i=1}^M$,

$$\Pr\left\{\bigcup_{i=1}^{M}\mathcal{A}_i\right\} \doteq \min\left\{1, \sum_{i=1}^{M}\Pr\{\mathcal{A}_i\}\right\}. \qquad (5)$$

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 $\{\mathcal{A}_i\}$ are pairwise independent [21, Lemma A.2], [22, Lemma 1].

B. Problem Formulation

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

[5]Consider the case where $b_n \doteq e^{bn}$ (b being a constant, independent of n) and the exponent of $\Pr\{A_n(i) \geq e^{bn}\}$ is a continuous function of b.

We assume an ensemble of random codes, where each codeword is selected independently at random, uniformly within a type class $\mathcal{T}_P$.[6] Therefore, all the joint types Q considered henceforth have $Q_X = P$. In addition, we define the set $\mathcal{Q}_P$ as the collection of all $\{Q_{X|Y}\}$ such that $(Q_Y \times Q_{X|Y})_X = P$, and again, all joint types Q considered in this paper will satisfy $Q_{X|Y} \in \mathcal{Q}_P$. Let $\mathcal{C} = \{\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_M\}$, $\mathbf{x}_m \in \mathcal{X}^n$, $m = 1, \ldots, M$, $M = e^{nR}$ (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, is a partition of $\mathcal{Y}^n$ into $M + 1$ regions, denoted $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_M$. If $\mathbf{y} \in \mathcal{R}_m$, $m = 1, 2, \ldots, M$, then the decoder decodes the message to be m. If $\mathbf{y} \in \mathcal{R}_0$, then the decoder declares that nothing has been transmitted, that is, $\mathbf{x} = \mathbf{0}$ and then $\mathbf{y}$ is "pure noise".

For the given detector/decoder, the probability of *false alarm* (FA) is defined as

$$P_{\text{FA}} \triangleq Q_0(\overline{\mathcal{R}_0}) = \sum_{m=1}^{M} Q_0(\mathcal{R}_m), \tag{6}$$

the probability of *misdetection* (MD) is defined as

$$P_{\text{MD}} \triangleq \frac{1}{M} \sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m), \tag{7}$$

and the probability of *inclusive error* (IE) is defined as

$$P_{\text{IE}} = \frac{1}{M} \sum_{m=1}^{M} W(\overline{\mathcal{R}_m}|\mathbf{x}_m) = \frac{1}{M} \sum_{m=1}^{M} \sum_{k \neq m} W(\mathcal{R}_k|\mathbf{x}_m), \tag{8}$$

where the inner summation at the right-most side *includes* $k = 0$. Thus, the IE event is the total error event, namely, of not deciding on the correct codeword transmitted.[7] The probability of decoding to an erroneous codeword, excluding the rejection region, is termed the *exclusive error* (EE) and is defined as

$$P_{\text{EE}} \triangleq P_{\text{IE}} - P_{\text{MD}}. \tag{9}$$

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

$$\begin{aligned} \text{minimize} \quad & P_{\text{IE}} \\ \text{subject to} \quad & P_{\text{FA}} \leq \epsilon_{\text{FA}} \\ & P_{\text{MD}} \leq \epsilon_{\text{MD}} \end{aligned} \tag{10}$$

[6] We do not restrict the input type P to assign zero probability to the silent symbol $'0'$, so this symbol can also be used for communication. The error exponents derived in the remaining of the paper assume a given input type P, and one can consider, as a special case, a type which assigns zero probability to $'0'$, like in the binary symmetric channel example in [26, Section 2.4.1] and [27].

[7] The inclusion of terms such as $k = 0$ is conventional in related problems. For example, in Forney's error/erasure setting [8], one of the events defined and analyzed is the total error event, which is comprised of a union of an undetected error event and an erasure event.

where ϵ_{FA} and ϵ_{MD} are given prescribed quantities, and it is assumed that these two constraints are not contradictory. Indeed, 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_{IE} under these constraints.

Our goal is to find the optimum detector/decoder for the problem (10), and then analyze the random coding exponents associated with the resulting error probabilities. The choice of P_{IE}, rather than P_{EE}, as the objective to minimize, enables the derivation of the optimal decoder in Lemma 1 to follow. However, in some cases, both problems lead to the same optimal decoder (cf. the discussion in Section 3). In any case, for the optimal detector/decoder, the random coding exponent of P_{EE} is also of interest.

III. The Optimum Detector/Decoder

Let us define the following detector/decoder:

$$\mathcal{R}_0^* = \left\{\mathbf{y}: \ a\cdot\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) + \max_m W(\mathbf{y}|\mathbf{x}_m) \le b\cdot Q_0(\mathbf{y})\right\} \tag{11}$$

$$\mathcal{R}_m^* = \overline{\mathcal{R}_0^*}\bigcap\left\{\mathbf{y}: \ W(\mathbf{y}|\mathbf{x}_m) > \max_{k\neq m} W(\mathbf{y}|\mathbf{x}_k)\right\}, \qquad m=1,2,\ldots,M, \tag{12}$$

where ties are broken arbitrarily, and where $a\ge 0$ and $b\ge 0$ 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_{IE}. 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*

$$Q_0(\overline{\mathcal{R}_0}) \le Q_0(\overline{\mathcal{R}_0^*}) \tag{13}$$

and

$$\frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m) \le \frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0^*|\mathbf{x}_m), \tag{14}$$

then

$$\frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_m^*}|\mathbf{x}_m) \le \frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_m}|\mathbf{x}_m). \tag{15}$$

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:

$$\mathcal{R}_m = \overline{\mathcal{R}_0} \bigcap \left\{ \mathbf{y} : \ W(\mathbf{y}|\mathbf{x}_m) > \max_{k \neq m} W(\mathbf{y}|\mathbf{x}_k) \right\}, \qquad m = 1, 2, \ldots, M. \tag{16}$$

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,

$$P_{\text{CD}} = \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y} \in \mathcal{R}_m} W(\mathbf{y}|\mathbf{x}_m), \tag{17}$$

is upper bounded by

$$P_{\text{CD}} \leq \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y} \in \mathcal{R}_m} \max_{k} W(\mathbf{y}|\mathbf{x}_k) = \frac{1}{M} \sum_{\mathbf{y} \in \overline{\mathcal{R}_0}} \max_{m} W(\mathbf{y}|\mathbf{x}_m) \tag{18}$$

and this bound is achieved by (16). Thus, upon adopting (16) 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 $\mathbf{y} \in \mathcal{Y}^n$

$$[\mathcal{I}\{\mathbf{y} \in \mathcal{R}_0^*\} - \mathcal{I}\{\mathbf{y} \in \mathcal{R}_0\}] \cdot \left[b \cdot Q_0(\mathbf{y}) - a \cdot \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) - \max_{m} W(\mathbf{y}|\mathbf{x}_m) \right] \geq 0. \tag{19}$$

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 $\mathbf{y} \in \mathcal{Y}^n$, we have:

$$\begin{aligned} 0 \ &\leq \ \sum_{\mathbf{y} \in \mathcal{Y}^n} [\mathcal{I}\{\mathbf{y} \in \mathcal{R}_0^*\} - \mathcal{I}\{\mathbf{y} \in \mathcal{R}_0\}] \cdot \left[b \cdot Q_0(\mathbf{y}) - a \cdot \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) - \max_{m} W(\mathbf{y}|\mathbf{x}_m) \right] \\ &= \ b \cdot [Q_0(\mathcal{R}_0^*) - Q_0(\mathcal{R}_0)] - a \cdot \left[\sum_{m=1}^{M} W(\mathcal{R}_0^*|\mathbf{x}_m) - \sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m) \right] - \\ &\quad \left[\sum_{\mathbf{y} \in \mathcal{R}_0^*} \max_{m} W(\mathbf{y}|\mathbf{x}_m) - \sum_{\mathbf{y} \in \mathcal{R}_0} \max_{m} W(\mathbf{y}|\mathbf{x}_m) \right] \end{aligned} \tag{20}$$

which yields

$$\begin{aligned} &\sum_{\mathbf{y} \in \mathcal{R}_0^*} \max_{m} W(\mathbf{y}|\mathbf{x}_m) - \sum_{\mathbf{y} \in \mathcal{R}_0} \max_{m} W(\mathbf{y}|\mathbf{x}_m) \\ \leq \ &b \cdot [Q_0(\mathcal{R}_0^*) - Q_0(\mathcal{R}_0)] - a \cdot \left[\sum_{m=1}^{M} W(\mathcal{R}_0^*|\mathbf{x}_m) - \sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m) \right] \\ = \ &b \cdot [Q_0(\overline{\mathcal{R}_0}) - Q_0(\overline{\mathcal{R}_0^*})] + a \cdot \left[\sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m) - \sum_{m=1}^{M} W(\mathcal{R}_0^*|\mathbf{x}_m) \right] \end{aligned} \tag{21}$$

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

$$Q_0(\overline{\mathcal{R}_0}) \leq Q_0(\overline{\mathcal{R}_0^*}) \tag{22}$$

and

$$\frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m) \leq \frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0^*|\mathbf{x}_m) \tag{23}$$

together imply that

$$\sum_{\mathbf{y}\in\mathcal{R}_0^*} \max_m W(\mathbf{y}|\mathbf{x}_m) \leq \sum_{\mathbf{y}\in\mathcal{R}_0} \max_m W(\mathbf{y}|\mathbf{x}_m) \tag{24}$$

or equivalently,

$$\sum_{\mathbf{y}\in\overline{\mathcal{R}_0^*}} \max_m W(\mathbf{y}|\mathbf{x}_m) \geq \sum_{\mathbf{y}\in\overline{\mathcal{R}_0}} \max_m W(\mathbf{y}|\mathbf{x}_m), \tag{25}$$

which in turn yields

$$\begin{aligned}
\frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_m^*}|\mathbf{x}_m) &= 1 - \frac{1}{M}\sum_{\mathbf{y}\in\overline{\mathcal{R}_0^*}} \max_m W(\mathbf{y}|\mathbf{x}_m) \\
&\leq 1 - \frac{1}{M}\sum_{\mathbf{y}\in\overline{\mathcal{R}_0}} \max_m W(\mathbf{y}|\mathbf{x}_m) \\
&= \frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_m}|\mathbf{x}_m).
\end{aligned} \tag{26}$$

This completes the proof of Lemma 1. ■

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

1) The detector/decoder derived in Lemma 1 is optimal for any *given* code $\mathcal{C}$, namely, the optimality is exact for all n, and not merely asymptotic. As mentioned earlier, in this work, we analyze the ensemble performance. Specifically, let $\bar{P}_{\text{IE}}$, $\bar{P}_{\text{EE}}$, $\bar{P}_{\text{FA}}$, and $\bar{P}_{\text{MD}}$ denote the corresponding ensemble averages of P_{IE}, P_{EE}, 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

$$\mathcal{R}_0^* = \left\{ \mathbf{y} : \ e^{n\alpha} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) + \max_m W(\mathbf{y}|\mathbf{x}_m) \leq e^{n\beta} Q_0(\mathbf{y}) \right\}. \tag{27}$$

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

2) The detection/rejection rule defined by (27) involves a linear combination of $\max_m W(\mathbf{y}|\mathbf{x}_m)$ and $\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)$, or equivalently, the overall output distribution induced by the code

$$Q_{\mathcal{C}}(\mathbf{y}) \triangleq \frac{1}{M}\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m). \tag{28}$$

In this context, the intuition behind the optimality of this detection rule is not trivial (at least for the authors of this article). It is instructive, nonetheless, to examine some special cases. The first observation is that for $\alpha \geq 0$, the term $e^{n\alpha}\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)$ dominates the term $\max_m W(\mathbf{y}|\mathbf{x}_m)$, and so, the rejection region is essentially equivalent to

$$\mathcal{R}_0' = \left\{\mathbf{y}:\ e^{n\alpha}\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) \leq e^{n\beta}Q_0(\mathbf{y})\right\} = \left\{\mathbf{y}:\ e^{n(\alpha+R)}Q_{\mathcal{C}}(\mathbf{y}) \leq e^{n\beta}Q_0(\mathbf{y})\right\}, \tag{29}$$

which is exactly the Neyman-Pearson test between $Q_{\mathcal{C}}(\mathbf{y})$ and $Q_0(\mathbf{y})$. This means that asymptotically, $\alpha \geq 0$ corresponds to a regime of full tension between the FA and the MD constraints (see Section II-B). In this case, E_{FA} and E_{MD} are related via the Neyman-Pearson lemma. Indeed, the detection-rejection rule (29) 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. Notice, however, that for finite block length, the error probabilities depend on the value of α and not just on the difference $\alpha - \beta$. At the other extreme, where $e^{n\alpha} \ll 1$, and the term $\max_m W(\mathbf{y}|\mathbf{x}_m)$ dominates, the detection rule becomes equivalent to

$$\mathcal{R}_0'' = \left\{\mathbf{y}:\ \max_m W(\mathbf{y}|\mathbf{x}_m) \leq e^{n\beta}Q_0(\mathbf{y})\right\}. \tag{30}$$

In this case, the silent mode is essentially treated as corresponding to yet another codeword - $\mathbf{x}_0 = \mathbf{0}$, although it still has a special stature due to the factor $e^{n\beta}$. The factor $e^{n\beta}$ may thought of as providing an unequal protection for the special 'silent codeword', compared to all other codewords in the codebook. However, 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 $Q_{\mathcal{C}}(\mathbf{y})$ and $\max_m W(\mathbf{y}|\mathbf{x}_m)$ play a considerable role.

3) It is interesting to compare the resulting optimal detector/decoder for $\alpha = 0$ with detector/decoder pairs proposed in [26] (and [27]). In [26, Chapter 4] both the ensemble of independent identically distributed (i.i.d.) codebooks over a distribution P, and the ensemble of fixed composition codebooks over a distribution P are considered. For i.i.d. codebooks, the detector proposed in [26] is simply a Neyman-Pearson test between Q_0 and $(P \times W)_Y$. The weakness of this approach is that for a *given* codebook, the single-letter output marginal (of each Y_i, for $1 \leq i \leq n$) might be different from $(P \times W)_Y$ (especially at low rates), and this leads to a mismatched Neyman-Pearson test. When averaging over all

codebooks in the ensemble, this degrades the achievable exponents. Indeed, the optimal detector/decoder $\mathcal{R}^*$ with $\alpha \geq 0$ proposed here for fixed composition ensemble of codes, is similar in form to the optimal detector/decoder of [26] for i.i.d. codebooks. However, the substantial difference is that it corresponds to a Neyman-Pearson test between the overall output distribution induced by the code $Q_{\mathcal{C}}$ (cf. (28)) and Q_0. In contrast, the decoder proposed in [26] for fixed composition codes (which were introduced in order to improve performance) is a non-optimal typicality detector, which is based on unions of conditional type classes centered around the codewords in the codebook. The non-optimal decoder may clearly degrade the resulting exponents compared to the optimal detector/decoder, as we shall see in a numerical example in Section VI.

4) Consider the following related problem

$$\begin{aligned} \text{minimize} \quad & P_{\text{EE}} \\ \text{subject to} \quad & P_{\text{FA}} \leq \epsilon_{\text{FA}} \\ & P_{\text{MD}} \leq \epsilon_{\text{MD}} \end{aligned} \tag{31}$$

and let $\mathcal{R}^{**}$ be the optimal detector/decoder for the problem (31). Now, as $P_{\text{IE}} = P_{\text{EE}} + P_{\text{MD}}$, it may be easily verified that when $P_{\text{MD}} = \epsilon_{\text{MD}}$ for the optimal detector/decoder $\mathcal{R}^*$ (of the problem (10)), then $\mathcal{R}^*$ is also the optimal detector/decoder for the problem (31). However, when $P_{\text{MD}} < \epsilon_{\text{MD}}$ for $\mathcal{R}^*$, then $\mathcal{R}^{**}$ is different, since it easy to check that for the problem 31, the constraint $P_{\text{MD}} \leq \epsilon_{\text{MD}}$ for $\mathcal{R}^{**}$ must be achieved with equality. To gain some intuition why (31) is more complicated than (10), notice that in (10) the two probabilities P_{MD} and P_{IE} (which are both conditioned on the event that one of the codewords was sent) increase with the expansion of $\mathcal{R}_0$. On the other hand, P_{FA}, which is conditioned on the all-zero sequence, decreases with the expansion of $\mathcal{R}_0$. The fact that all probabilities which are conditioned on the same event have a similar trend with respect to an expansion of $\mathcal{R}_0$ is crucial to the proof of Lemma 1. Indeed, in contrast, in (10), both P_{MD} and P_{EE} behave oppositely as $\mathcal{R}_0$ is expanded (the former increases, while the later decreases). Therefore, asymptotic analysis of P_{EE} for $\mathcal{R}^*$ is interesting, especially in the range where $\mathcal{R}^{**}$ is equivalent to $\mathcal{R}^*$ and hence optimal for P_{EE} as well.

IV. Performance

In this section, we present our main theorem, which provides exact single-letter characterizations for all exponents as functions of the coding rate R, and given α and β. Following comment no. 2 in the discussion at the end Section III, we assume throughout that $\alpha \leq 0$.

We first need some definitions. Let

$$d(x,y) \triangleq \ln\left[\frac{Q_0(y)}{W(y|x)}\right], \qquad x \in \mathcal{X},\ y \in \mathcal{Y} \tag{32}$$

and denote $D(Q) \triangleq \mathbf{E}_Q d(X,Y)$.

For a given output distribution $Q_Y = \{Q_Y(y),\ y \in \mathcal{Y}\}$ define

$$\mathbf{R}(\Delta; Q_Y) \triangleq \min_{Q_{X|Y} \in \mathcal{Q}_P:\, D(Q) \le \Delta} I(Q) \tag{33}$$

as well as

$$\mathbf{S}(\Delta; Q_Y) \triangleq \inf_{Q_{X|Y} \in \mathcal{Q}_P:\, D(Q) > \Delta} \mathcal{D}(Q_{Y|X} \| W | P). \tag{34}$$

Also, let $R_1(Q_Y)$ and $D_1(Q_Y)$ denote $I(Q^*)$ and $D(Q^*)$, where Q^* minimizes $I(Q) + D(Q)$ subject to the constraint $Q_X = P$, and

$$R_0(\alpha, \beta; Q_Y) \triangleq \begin{cases} R_1(Q_Y) + D_1(Q_Y) + \beta - \alpha & -\beta < D_1(Q_Y) \\ \mathbf{R}(-\beta; Q_Y) & \text{otherwise} \end{cases}. \tag{35}$$

Let $E_{\mathrm{r}}(R)$ be the ordinary random coding exponent function for fixed composition codes, i.e.,

$$E_{\mathrm{r}}(R) = \min_{Q} \left\{ \mathcal{D}(Q_{Y|X} \| W | P) + [I(Q) - R]_+ \right\} \tag{36}$$

where $Q = P \times Q_{Y|X}$ here, and in all the other expressions of exponents which follow.

For the FA exponent, define the following functions

$$E_A(R) \triangleq \min_{Q_{Y|X}:\, D(Q) \le -\beta + \alpha + [R - I(Q)]_+} \left\{ \mathcal{D}(Q_Y || Q_0) + [I(Q) - R]_+ \right\}, \tag{37}$$

$$E_B(R) \triangleq \min_{Q_{Y|X}:\, D(Q) \le -\beta} \left\{ \mathcal{D}(Q_Y || Q_0) + [I(Q) - R]_+ \right\}, \tag{38}$$

and

$$E_{\text{FA}}(R) \triangleq \min\left\{ E_A(R), E_B(R) \right\}. \tag{39}$$

For the MD exponent, define

$$E_{\text{MD}}(R) \triangleq \inf_{Q_Y:\, R_0(\alpha,\beta;Q_Y) > R} \mathbf{S}(-\beta; Q_Y). \tag{40}$$

For the decoding error exponents, define[8]

$$E_2''(R) \triangleq \min_{Q_Y} \left\{ \mathbf{S}(\alpha-\beta; Q_Y) + \left[\mathbf{R}(\alpha-\beta; Q_Y) - R \right]_+ \right\}, \tag{41}$$

$$E_3'(R) \triangleq \min_{Q_{Y|X}:\, D(Q) \le -\beta} \left\{ \mathcal{D}(Q_{Y|X} || W | P) + \left[I(Q) - R \right]_+ \right\}, \tag{42}$$

and

$$E_3''(R) \triangleq \min_{Q_Y} \left\{ \mathbf{S}(-\beta; Q_Y) + \left[\mathbf{R}(-\beta; Q_Y) - R \right]_+ \right\}. \tag{43}$$

as well as

$$E_{\text{EE}}(R) \triangleq \min \left\{ E_2''(R), E_3'(R), E_3''(R) \right\}, \tag{44}$$

and

$$E_{\text{IE}}(R) \triangleq \min \left\{ E_{\text{r}}(R), E_{\text{MD}}(R) \right\}. \tag{45}$$

Theorem 2. *Let W be a DMC and let $\mathcal{R}^*$ be defined as in Section II-B. Let the codewords of $\mathcal{C} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$, $M = e^{nR}$, be selected independently at random under the uniform distribution across a given type class $\mathcal{T}_P$. Then, the asymptotic exponents associated with $\bar{P}_{FA}$, $\bar{P}_{MD}$, $\bar{P}_{EE}$, and $\bar{P}_{IE}$ are given, respectively, by $E_{FA}(R)$, $E_{MD}(R)$, $E_{EE}(R)$, and $E_{IE}(R)$, as defined in eqs.* (39)*,* (40)*,* (44) *and* (45)*.*

Discussion. Before delving into the proof of the theorem in Section V, we make a few comments:

1) Eq. (45) follows from a simple application of the union bound, since the inclusive decoding error event is defined to include the misdetection event.

2) As discussed in Section III, for $\alpha \ge 0$, depend only on the difference $\alpha - \beta$, and thus can be computed by replacing $\alpha \to 0$ and $\beta \to \beta - \alpha$.

3) As mentioned previously, for any given rate R, one should tune the parameters α and β to meet prescribed constraints on E_{MD} and E_{FA}. However, observing eq. (45), and recalling that $\alpha = 0$ provides the maximal E_{MD} for any given E_{FA} (cf. comment 2 in Section III), it is evident that any strictly negative value of α may be replaced by $\alpha = 0$ and both E_{MD} and E_{IE} may only improve. Nonetheless, choosing $\alpha < 0$ may be still interesting if one is interested in increasing E_{EE} (perhaps at the price of decreasing E_{MD} and E_{IE}). Of course, for an actual finite block length decoder, even positive values of α may be used, in order to optimally fine-tune the error probabilities obtained (where as mentioned, asymptotically, the error exponents only depend on $\alpha - \beta$) . Notice that when $\alpha = 0$, a slight simplification in the exponents expressions is possible. In this case $E_{\text{FA}}(R) = E_A(R)$ and $E_{\text{IE}}(R) = \min\{E_3'(R), E_3''(R)\}$.

4) It is straightforward to observe that for a given $E_{\text{MD}} > 0$ (and $E_{\text{FA}} > 0$), there is no rate loss in terms

[8]The subscripts 2 and 3 of the following exponents are related to derivations in the proof (cf. Section V-C). There is also an $E_1(R)$ exponent, but it is not needed for the final minimization, and thus omitted here.

of the maximum achievable information rate for which the average probability of the inclusive decoding error still tends to zero, that is, the smallest rate R for which $E_{\text{IE}} = 0$, for given E_{MD} and E_{FA}. This is easily seen from eq. (45). Since $E_{\text{MD}} > 0$ is given, and since $E_{\text{r}}(R) \to 0$ as R approaches $I(P \times W)$, then it is clear that beyond a certain rate R, we have $E_{\text{IE}} = E_{\text{r}}(R)$, and hence E_{IE} also vanishes at $R = I(P \times W)$. Of course, if P is chosen to be the capacity-achieving input distribution, then the capacity is still achieved in this setting.

5) Notice that by definition $E_{\text{EE}}(R) \geq E_{\text{r}}(R)$. Now, for a given rate R, let $Q^*_{Y|X}$ be the minimizer of the objective function of $E_{\text{r}}(R)$ (in eq. (36)). If $D(Q^*_{Y|X}) \leq -\beta$ then eqs. (42) and (44) imply (see also Appendix A-C) that for this rate $E_{\text{EE}}(R) = E_{\text{r}}(R)$. In other cases $E_{\text{EE}}(R)$ may be strictly larger than $E_{\text{r}}(R)$ (but always $E_{\text{IE}}(R) \leq E_{\text{r}}(R)$).

6) The achievable exponents derived in [26, Theorem 4.2] for an ensemble of fixed composition codes of input type P with $E_{\text{IE}} = 0$, have a form similar to the optimal exponents for $\alpha = 0$. For example, for a given E_{MD}, the achievable FA exponent is given by

$$E_{\text{FA}}(R) = \min_{Q_{Y|X}:\mathcal{D}(Q_{Y|X}||W|P) \leq E_{\text{MD}}} \left\{ \mathcal{D}(Q_Y||Q_0) + [I(Q) - R]_+ \right\}. \tag{46}$$

The only difference from $E_{\text{FA}}(R) = E_A(R)$ is the domain of the minimization, which for $E_A(R)$, depends also on $Q_0(\mathbf{y})$ (via $D(Q)$) and not only on the channel W.

7) As mentioned in the Introduction, in many practical systems, the problem of transmission detection and decoding is performed separately, using a pattern that marks the beginning of transmission (also called *training approach*). The allowed block length n is divided into two parts of lengths γn and $(1-\gamma)n$, for some $\gamma \in (0, 1)$. In the first part, a training word of γn letters is transmitted, identical for all possible messages. Then, in the second part, an ordinary codebook of block length $(1-\gamma)n$ is used. At the receiver, the existence of a message is detected solely on the basis of the first γn output letters, and if transmission is detected, then an ordinary decoder is used for the remaining $(1-\gamma)n$ output letters. As is well known, in order to obtain an exponential decrease in the FA and MD probabilities, one must choose γ to be strictly positive. In turn, this means that effective block length of the decoder is $(1-\gamma)n$, which is strictly less than n, and so this separation approach decreases the error exponent (see also [7, Appendix I] for the related problem of training for channel estimation). In light of eq. (45), it is evident that at least for large rates, such that $E_{\text{MD}}(R) > E_{\text{r}}(R)$, the inclusive decoding error exponent for optimal detector/decoder is not reduced, compared to ordinary decoding. This implies that the training approach is in general sub-optimal (see also [25], [26] for observations in the same spirit).

8) In some applications of sparse communication, it is required to maximize the FA exponent, since noise is present almost the entire communication time. Thus, it is of interest to maximize $E_{\text{FA}}(R)$ under

the assumption that $E_{\text{MD}}(R) = 0$ (more precisely, a small strictly positive constant, so that $P_{\text{MD}} \to 0$ as $n \to 0$). As stated in comment 3, choosing $\alpha = 0$ provides the optimal trade-off between the FA and MD exponents, so the designer need only find the maximal β such that $E_{\text{MD}}(R) = 0$, and then calculate the resulting $E_{\text{FA}}(R)$. Some insight may be obtained by analyzing the extreme case of $R = 0$. For zero rate, notice that

$$E_{\text{MD}}(0) = \inf_{Q:\, D(Q) > -\beta} \mathcal{D}(Q \| P \times W) \tag{47}$$

where $Q = P \times Q_{Y|X}$ and

$$\begin{aligned} E_{\text{FA}}(0) &= E_A(0) \\ &= \min_{Q_{Y|X}:\, D(Q) \le -\beta} \{\mathcal{D}(Q_Y || Q_0) + I(Q)\} \\ &= \min_{Q:\, D(Q) \le -\beta} \mathcal{D}(Q || P \times Q_0). \end{aligned} \tag{48}$$

Now, let us ignore, for a moment, the constraints on β in $E_{\text{MD}}(0)$ and $E_{\text{FA}}(0)$. For any given $\delta \ge 0$, recall that the minimizer of

$$\min_{Q:\, \mathcal{D}(Q||P\times Q_0) \le \delta} \mathcal{D}(Q || P \times W) \tag{49}$$

is given by

$$Q_\lambda(x, y) \propto P(x) \cdot \left(W^\lambda(y|x) \cdot Q_0^{1-\lambda}(y)\right) \tag{50}$$

where $\lambda \in [0, 1]$ is chosen to satisfy the constraint on the divergence, and Q_λ sums to 1 over $\mathcal{X} \times \mathcal{Y}$ (see, e.g. [6, Chapter 11]. Now, using continuity arguments, as λ traces the interval $[0, 1]$, if

$$\beta = D(Q_\lambda) \tag{51}$$

is chosen, then the exponents achieved are

$$E_{\text{MD}}(0) = \mathcal{D}(Q_\lambda \| P \times W) \tag{52}$$

and

$$E_{\text{FA}}(0) = \mathcal{D}(Q_\lambda \| P \times Q_0). \tag{53}$$

Specifically, for $E_{\text{MD}}(0) = 0$ we get

$$E_{\text{FA,max}}(0) = \mathcal{D}(P \times W \| P \times Q_0). \tag{54}$$

Moreover, let $x^* \in \mathcal{X}$ be the input letter which maximizes $\mathcal{D}(Q(\cdot|x)||Q_0)$, and choose the input type to

assign $P(x^*) = 1$. Then,

$$E_{\text{FA,max}}(0) = \mathcal{D}(W(\cdot|x^*)\|Q_0). \tag{55}$$

This recovers the *synchronization threshold* derived in [5].

9) While the exponent expressions in Theorem 2 are relatively compact, one may wonder if and how they can be computed efficiently. In Appendix A we discuss aspects of the computation of these exponents, by revealing their similarity to the ordinary random coding exponent. It is shown there that all the optimization problems involved, except for $E_{\text{MD}}(R)$, are in fact convex optimization problems and thus can be computed efficiently [3]. The computation of $E_{\text{MD}}(R)$, which is more complex, is also briefly discussed in Appendix A.

V. Proof of Theorem 2

First, observe that, as mentioned in the above discussion, since the inclusive decoding error event includes the misdetection event, then P_{IE} is lower bounded by $\max\{P_{\text{ODE}}, P_{\text{IE}}\}$ and upper bounded by $P_{\text{ODE}} + P_{\text{IE}}$, where P_{ODE} is the probability of error associated with ordinary ML decoding, without the decision region $\mathcal{R}_0$. This readily yields eq. (45). It remains then to establish the single-letter formulas of the FA, MD and EE exponents, i.e., eqs. (39), (40) and (44). Accordingly, this section is divided into three subsections, each one devoted to the analysis of one of these exponents.

A. The False Alarm Error Exponent

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

$$\begin{aligned}
\bar{P}_{\text{FA}}(\mathbf{y}) &\triangleq \Pr\left\{e^{n\alpha}\cdot\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{X}_m) + \max_m W(\mathbf{y}|\mathbf{X}_m) > e^{n\beta}Q_0(\mathbf{y})\right\} && (56)\\
&\doteq \Pr\left\{e^{n\alpha}\cdot\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{X}_m) > e^{n\beta}Q_0(\mathbf{y})\right\} + \\
&\quad \Pr\left\{\max_m W(\mathbf{y}|\mathbf{X}_m) > e^{n\beta}Q_0(\mathbf{y})\right\} && (57)\\
&= \Pr\left\{\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{X}_m) > e^{n(\beta-\alpha)}Q_0(\mathbf{y})\right\} + \\
&\quad \Pr\left\{\max_m W(\mathbf{y}|\mathbf{X}_m) > e^{n\beta}Q_0(\mathbf{y})\right\} && (58)\\
&\triangleq A(\mathbf{y}) + B(\mathbf{y}), && (59)
\end{aligned}$$

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

with $\mathbf{y}$ is $\hat{Q} = \{\hat{Q}(x,y),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$ (with $\hat{Q}_X = P$). Next, define

$$f(\hat{Q}) \triangleq \sum_{x,y} \hat{Q}(x,y) \ln W(y|x) \tag{60}$$

and

$$g(\hat{Q}_Y) \triangleq \sum \hat{Q}_Y(y) \ln Q_0(y) + \beta - \alpha \tag{61}$$

as well as

$$u(\hat{Q}) \triangleq g(\hat{Q}_Y) - f(\hat{Q}) = \sum_{x,y \in \mathcal{X}\times\mathcal{Y}} \hat{Q}(x,y) \ln \frac{Q_0(y)}{W(y|x)} + \beta - \alpha = D(\hat{Q}) + \beta - \alpha. \tag{62}$$

We begin with the analysis of $A(\mathbf{y})$,

$$\begin{aligned}
A(\mathbf{y}) &\triangleq \Pr\left\{\sum_{\hat{Q}_{X|Y}} N(\hat{Q}|\mathbf{y}) e^{nf(\hat{Q})} > e^{ng(\hat{Q}_Y)}\right\} && (63)\\
&\dot{=} \Pr\left\{\max_{\hat{Q}_{X|Y}} N(\hat{Q}|\mathbf{y}) e^{nf(\hat{Q})} > e^{ng(\hat{Q}_Y)}\right\} && (64)\\
&= \Pr \bigcup_{\hat{Q}_{X|Y}} \left\{N(\hat{Q}|\mathbf{y}) e^{nf(\hat{Q})} > e^{ng(\hat{Q}_Y)}\right\} && (65)\\
&\dot{=} \sum_{\hat{Q}_{X|Y}} \Pr\left\{N(\hat{Q}|\mathbf{y}) > e^{n[g(\hat{Q}_Y) - f(\hat{Q})}\right\} && (66)\\
&\dot{=} \max_{\hat{Q}_{X|Y}} \Pr\left\{N(\hat{Q}|\mathbf{y}) > e^{nu(\hat{Q})}\right\}, && (67)
\end{aligned}$$

where we have used again eq. (3). Now, since $N(\hat{Q}|\mathbf{y})$ is a binomial 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 [14, Subsection 6.3]

$$\Pr\{N(\hat{Q}|\mathbf{y}) \geq e^{nu(\hat{Q})}\} \dot{=} \exp\left\{-e^{n[u(\hat{Q})]_+}(n[I(\hat{Q}) - R + [u(\hat{Q})]_+] - 1)\right\}, \tag{68}$$

provided that for $u(\hat{Q}) > 0$, $I(\hat{Q}) - R + u(\hat{Q}) > 0$ (otherwise, $\Pr\{N(\hat{Q}|\mathbf{y}) \geq e^{nu(\hat{Q})}\} \to 1$).[9] Therefore, the exponential rate $E(\hat{Q})$ of $\Pr\{N(\hat{Q}|\mathbf{y}) \geq e^{nu(\hat{Q})}\}$ is as follows:

$$E(\hat{Q}) = \begin{cases} [I(\hat{Q}) - R]_+ & u(\hat{Q}) \leq 0 \\ \infty & u(\hat{Q}) > 0,\ u(\hat{Q}) > R - I(\hat{Q}) \\ 0 & u(\hat{Q}) > 0,\ u(\hat{Q}) < R - I(\hat{Q}) \end{cases} \tag{69}$$

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

$$= \begin{cases} I(\hat{Q}) - R & u(\hat{Q}) \leq 0, R \leq I(\hat{Q}) \\ 0 & u(\hat{Q}) \leq 0, R > I(\hat{Q}) \\ \infty & u(\hat{Q}) > 0, u(\hat{Q}) > R - I(\hat{Q}) \\ 0 & u(\hat{Q}) > 0, u(\hat{Q}) \leq R - I(\hat{Q}) \end{cases} \tag{70}$$

$$= \begin{cases} \Big[I(\hat{Q}) - R\Big]_+ & u(\hat{Q}) \leq \Big[R - I(\hat{Q})\Big]_+ \\ \infty & u(\hat{Q}) > \Big[R - I(\hat{Q})\Big]_+ \end{cases}. \tag{71}$$

Using the convention that the minimum over an empty set is infinity, we may succinctly write

$$\min_{\hat{Q}_{X|Y} \in \mathcal{Q}_P} E(\hat{Q}) = \min_{\hat{Q}_{X|Y} \in \mathcal{Q}_P} \left\{ \Big[I(\hat{Q}) - R\Big]_+ \text{ s.t. } u(\hat{Q}) \leq \Big[R - I(\hat{Q})\Big]_+ \right\} \tag{72}$$

$$= \min_{\hat{Q}_{X|Y} \in \mathcal{Q}_P} \left\{ \Big[I(\hat{Q}) - R\Big]_+ \text{ s.t. } D(\hat{Q}) \leq -\beta + \alpha + \Big[R - I(\hat{Q})\Big]_+ \right\}. \tag{73}$$

For the overall exponent associated with A, we need to average over $\mathbf{Y}$, which gives $A \doteq e^{-nE_A}$ with

$$E_A(R) = \min_{\hat{Q}_{Y|X}} \left\{ \mathcal{D}(\hat{Q}_Y || Q_0) + \Big[I(\hat{Q}) - R\Big]_+ \text{ s.t. } D(\hat{Q}) \leq -\beta + \alpha + \Big[R - I(\hat{Q})\Big]_+ \right\}. \tag{74}$$

Moving on to the analysis of $B(\mathbf{y})$, we have

$$B(\mathbf{y}) \triangleq \Pr\left\{ \max_m W(\mathbf{y}|\mathbf{X}_m) > e^{n\beta} Q_0(\mathbf{y}) \right\} \tag{75}$$

$$= \Pr \bigcup_{m=1}^{M} \left\{ W(\mathbf{y}|\mathbf{X}_m) > e^{n\beta} Q_0(\mathbf{y}) \right\} \tag{76}$$

$$\doteq \min\left\{1, M \cdot \Pr\{W(\mathbf{y}|\mathbf{X}_1) > e^{n\beta} Q_0(\mathbf{y})\}\right\}, \tag{77}$$

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

$$\Pr\{W(\mathbf{y}|\mathbf{X}_1) > e^{n\beta} Q_0(\mathbf{y})\} \doteq e^{-nI_0(\hat{Q}_Y)}, \tag{78}$$

where

$$I_0(\hat{Q}_Y) \triangleq \min_{\hat{Q}_{X|Y}} \Big\{ I(\hat{Q}) : \ D(\hat{Q}) \leq -\beta, \ \ \hat{Q}_{X|Y} \in \mathcal{Q}_P \Big\} \tag{79}$$

$$= \mathbf{R}(-\beta; \hat{Q}_Y). \tag{80}$$

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

$$E_B(R) \triangleq \min_{Q_Y} \left\{ \mathcal{D}(Q_Y || Q_0) + [\mathbf{R}(-\beta; Q_Y) - R]_+ \right\} \tag{81}$$

$$= \min_{Q_Y}\left\{\mathcal{D}(Q_Y||Q_0)+\left[\min_{Q_{Y|X}:D(Q)\le-\beta,(P\times Q_{Y|X})_Y=Q_Y} I(Q)-R\right]_+\right\} \quad (82)$$

$$= \min_{Q_Y}\left\{\mathcal{D}(Q_Y||Q_0)+\min_{Q_{Y|X}:D(Q)\le-\beta,(P\times Q_{Y|X})_Y=Q_Y}[I(Q)-R]_+\right\} \quad (83)$$

$$= \min_{Q_Y}\min_{Q_{Y|X}:\,D(Q)\le-\beta,(P\times Q_{Y|X})_Y=Q_Y}\left\{\mathcal{D}(Q_Y||Q_0)+[I(Q)-R]_+\right\} \quad (84)$$

$$= \min_{Q_{Y|X}:\,D(P\times Q_{Y|X})\le-\beta}\left\{\mathcal{D}(Q_Y||Q_0)+[I(Q)-R]_+\right\} \quad (85)$$

where $Q = P \times Q_{Y|X}$.

B. The Misdetection Error Exponent

Without loss of generality, we will assume that $\mathbf{X}_1 = \mathbf{x}_1$ was transmitted. We first condition on $\mathbf{x}_1$ and $\mathbf{y}$, and use the fact that $(\mathbf{X}_1, \mathbf{Y})$ are independent of $\{\mathbf{X}_m\}_{m=2}^M$:

$$\bar{P}_{\text{MD}}(\mathbf{x}_1,\mathbf{y}) \triangleq \Pr\left\{e^{n\alpha}\sum_{m=1}^M W(\mathbf{y}|\mathbf{X}_m)+\max_m W(\mathbf{y}|\mathbf{X}_m)\le e^{n\beta}Q_0(\mathbf{y})\middle|\mathbf{X}_1=\mathbf{x}_1,\mathbf{Y}=\mathbf{y}\right\} \quad (86)$$

$$= \Pr\left\{e^{n\alpha}\sum_{m=1}^M W(\mathbf{y}|\mathbf{X}_m)+ \max\{W(\mathbf{y}|\mathbf{x}_1),\max_{m>1}W(\mathbf{y}|\mathbf{X}_m)\}\le e^{n\beta}Q_0(\mathbf{y})\middle|\mathbf{X}_1=\mathbf{x}_1,\mathbf{Y}=\mathbf{y}\right\} \quad (87)$$

$$\doteq \Pr\left\{e^{n\alpha}\left[W(\mathbf{y}|\mathbf{x}_1)+\sum_{m>1}W(\mathbf{y}|\mathbf{X}_m)\right] +W(\mathbf{y}|\mathbf{x}_1)+\max_{m>1}W(\mathbf{y}|\mathbf{X}_m)\le e^{n\beta}Q_0(\mathbf{y})\right\} \quad (88)$$

$$\doteq \Pr\left\{e^{n[\alpha]_+}W(\mathbf{y}|\mathbf{x}_1)+e^{n\alpha}\sum_{m>1}W(\mathbf{y}|\mathbf{X}_m)+\max_{m>1}W(\mathbf{y}|\mathbf{X}_m)\le e^{n\beta}Q_0(\mathbf{y})\right\} \quad (89)$$

$$\doteq \Pr\left\{e^{n[\alpha]_+}W(\mathbf{y}|\mathbf{x}_1)<e^{n\beta}Q_0(\mathbf{y}), \; e^{n\alpha}\sum_{m>1}W(\mathbf{y}|\mathbf{X}_m)+\max_{m>1}W(\mathbf{y}|\mathbf{X}_m)\le e^{n\beta}Q_0(\mathbf{y})\right\} \quad (90)$$

$$= \mathcal{I}\left\{e^{n[\alpha]_+}W(\mathbf{y}|\mathbf{x}_1)<e^{n\beta}Q_0(\mathbf{y})\right\}\times \Pr\left\{e^{n\alpha}\sum_{m>1}W(\mathbf{y}|\mathbf{X}_m)+\max_{m>1}W(\mathbf{y}|\mathbf{X}_m)\le e^{n\beta}Q_0(\mathbf{y})\right\} \quad (91)$$

$$\triangleq C\cdot D. \quad (92)$$

Using the identity

$$\max_{m>1}W(\mathbf{y}|\mathbf{x}_m)\equiv\max_{\hat{Q}_{X|Y}}\mathcal{I}\{N(\hat{Q}|\mathbf{y})\ge1\}\cdot e^{nf(\hat{Q})} \quad (93)$$

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

$$\begin{aligned}
D &= \Pr\left\{e^{n\alpha}\sum_{m>1}W(\mathbf{y}|\mathbf{X}_m)+\max_{m>1}W(\mathbf{y}|\mathbf{X}_m)\le e^{n\beta}Q_0(\mathbf{y})\Big|\mathbf{x}_1,\mathbf{y}\right\} && (94)\\
&= \Pr\left\{e^{n\alpha}\sum_{\hat{Q}_{X|Y}}N(\hat{Q}|\mathbf{y})e^{nf(\hat{Q})}+\max_{\hat{Q}_{X|Y}}\mathcal{I}\{N(\hat{Q}|\mathbf{y})\ge 1\}\cdot e^{nf(\hat{Q})}\le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\mathbf{x}_1,\mathbf{y}\right\} && (95)\\
&\doteq \Pr\left\{e^{n\alpha}\sum_{\hat{Q}_{X|Y}}N(\hat{Q}|\mathbf{y})e^{nf(\hat{Q})}+\sum_{\hat{Q}_{X|Y}}\mathcal{I}\{N(\hat{Q}|\mathbf{y})\ge 1\}e^{nf(\hat{Q})}\le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\mathbf{x}_1,\mathbf{y}\right\} && (96)\\
&= \Pr\left\{\sum_{\hat{Q}_{X|Y}}[e^{n\alpha}N(\hat{Q}|\mathbf{y})+\mathcal{I}\{N(\hat{Q}|\mathbf{y})\ge 1\}]e^{nf(\hat{Q})}\le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\mathbf{x}_1,\mathbf{y}\right\} && (97)\\
&\doteq \Pr\left\{\max_{\hat{Q}_{X|Y}}[e^{n\alpha}N(\hat{Q}|\mathbf{y})+\mathcal{I}\{N(\hat{Q}|\mathbf{y})\ge 1\}]e^{nf(\hat{Q})}\le e^{n[g(\hat{Q}_Y)+\alpha]}\Big|\mathbf{x}_1,\mathbf{y}\right\} && (98)\\
&= \Pr\bigcap_{\hat{Q}_{X|Y}}\left\{e^{n\alpha}N(\hat{Q}|\mathbf{y})+\mathcal{I}\{N(\hat{Q}|\mathbf{y})\ge 1\}\le e^{n[u(\hat{Q})+\alpha]}\Big|\mathbf{x}_1,\mathbf{y}\right\} && (99)\\
&= \Pr\bigcap_{\hat{Q}_{X|Y}}\left\{N(\hat{Q}|\mathbf{y})\le e^{nv(\hat{Q})}\Big|\mathbf{x}_1,\mathbf{y}\right\}, && (100)
\end{aligned}$$

where

$$v(\hat{Q})=\begin{cases}u(\hat{Q}) & u(\hat{Q})+\alpha>0\\ -\infty & u(\hat{Q})+\alpha\le 0\end{cases} \tag{101}$$

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 $R>R_0(\alpha,\beta;Q_Y)\triangleq\min_{Q_{X|Y}\in\mathcal{Q}_P}\max\{I(\hat{Q}),I(\hat{Q})+v(\hat{Q})\}$. 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, $R<R_0(\alpha,\beta;Q_Y)$, 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 $\mathcal{I}\{R_0(\alpha,\beta;Q_Y)>R\}$, Thus,

$$\begin{aligned}
P_{\text{MD}} &\doteq \mathbf{E}\mathcal{I}\left\{R_0(\alpha,\beta;Q_Y)>R,\ W(\mathbf{Y}|\mathbf{X}_1)\le e^{n(\beta-[\alpha]_+)}Q_0(\mathbf{Y})\right\} && (102)\\
&= \exp\left[-n\inf_{Q_{Y|X}}\left\{\mathcal{D}(Q_{Y|X}\|W|P):\ R_0(\alpha,\beta;Q_Y)>R,\ D(Q)>[\alpha]_+-\beta\right\}\right]. && (103)
\end{aligned}$$

Thus,

$$E_{\text{MD}} = \inf_{Q_{Y|X}} \left\{ \mathcal{D}(Q_{Y|X}\|W|P) : \ R_0(\alpha,\beta;Q_Y) > R, \ D(Q) > [\alpha]_+ - \beta \right\} \tag{104}$$

$$= \inf_{Q_Y:\, R_0(\alpha,\beta;Q_Y)>R} \left\{ \inf_{Q_{Y|X}:\, (P\times Q_{Y|X})_Y = Q_Y, D(Q)>[\alpha]_+-\beta} \mathcal{D}(Q_{Y|X}\|W|P) \right\} \tag{105}$$

$$= \inf_{Q_Y:\, R_0(\alpha,\beta;Q_Y)>R} \mathbf{S}([\alpha]_+ - \beta; Q_Y) \tag{106}$$

Now, let us take a closer look at $R_0(\alpha,\beta;Q_Y)$:

$$\max\{I(Q), I(Q)+v(Q)\} = \begin{cases} \max\{I(Q), I(Q)+u(Q)\} & u(Q) > -\alpha \\ I(Q) & u(Q) \le -\alpha \end{cases} \tag{107}$$

$$= I(Q) + u(Q)\cdot \mathcal{I}\{u(Q) > [-\alpha]_+\}. \tag{108}$$

Thus,

$$R_0(\alpha,\beta;Q_Y) = \min_{Q_{X|Y}\in\mathcal{Q}_P} [I(Q) + u(Q)\cdot\mathcal{I}\{u(Q) > [-\alpha]_+\}] \tag{109}$$

$$= \min\left\{ \min_{Q_{X|Y}\in\mathcal{Q}_P:\, u(Q)\le[-\alpha]_+} I(Q), \min_{Q_{X|Y}\in\mathcal{Q}_P:\, u(Q)>[-\alpha]_+} [I(Q)+u(Q)] \right\}. \tag{110}$$

Now,

$$\min_{Q_{X|Y}\in\mathcal{Q}_P:\, u(Q)\le[-\alpha]_+} I(Q) = \mathbf{R}(\alpha + [-\alpha]_+ - \beta; Q_Y) \tag{111}$$

$$= \mathbf{R}([\alpha]_+ - \beta; Q_Y) \tag{112}$$

and

$$\min_{Q_{X|Y}\in\mathcal{Q}_P:\, u(Q)>[-\alpha]_+} [I(Q)+u(Q)] \tag{113}$$

$$= \beta - \alpha + \min_{Q_{X|Y}\in\mathcal{Q}_P:\, D(Q)>[\alpha]_+-\beta} [I(Q)+D(Q)] \tag{114}$$

$$= \beta - \alpha + \begin{cases} R_1(Q_Y) + D_1(Q_Y) & [\alpha]_+ - \beta < D_1(Q_Y) \\ \mathbf{R}([\alpha]_+ - \beta; Q_Y) + [\alpha]_+ - \beta & \text{otherwise} \end{cases} \tag{115}$$

$$= \begin{cases} R_1(Q_Y) + D_1(Q_Y) + \beta - \alpha & [\alpha]_+ - \beta < D_1(Q_Y) \\ \mathbf{R}([\alpha]_+ - \beta; Q_Y) + [\alpha]_+ - \alpha & \text{otherwise} \end{cases} \tag{116}$$

$$= \begin{cases} R_1(Q_Y) + D_1(Q_Y) + \beta - \alpha & [\alpha]_+ - \beta < D_1(Q_Y) \\ \mathbf{R}([\alpha]_+ - \beta; Q_Y) + [-\alpha]_+ & \text{otherwise} \end{cases} \tag{117}$$

The term in the first line corresponds to an unconstrained minimization problem. Thus, whenever the term in the second line is active, it is larger than the term in the first line. Thus, using the fact that $\alpha \le 0$ we

get

$$R_0(\alpha,\beta;Q_Y)=\begin{cases} R_1(Q_Y)+D_1(Q_Y)+\beta-\alpha & -\beta<D_1(Q_Y) \\ \mathbf{R}(-\beta;Q_Y) & \text{otherwise} \end{cases}. \tag{118}$$

C. The Exclusive Decoding Error Exponent

Let us denote

$$\Omega_m \triangleq \left\{\mathbf{y}:\ W(\mathbf{y}|\mathbf{x}_m)>\max_{k\neq m} W(\mathbf{y}|\mathbf{x}_k)\right\}. \tag{119}$$

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

$$P_{\text{EE}} \ = \ \frac{1}{M}\sum_{m=1}^{M} W(\overline{\mathcal{R}_0^*}\cap\overline{\Omega_m}|\mathbf{x}_m). \tag{120}$$

Let $\mathbf{X}_1=\mathbf{x}_1$ be transmitted and let $\mathbf{Y}=\mathbf{y}$ be received, and let $\tilde{Q}$ denote their empirical joint distribution. As before, we first condition on $(\mathbf{x}_1,\mathbf{y})$.

$$\begin{aligned}
\Pr\{\overline{\mathcal{R}_0^*}\cap\overline{\Omega_1}|\mathbf{x}_1,\mathbf{y}\} \ &= \ \Pr\left\{e^{n\alpha}\sum_m W(\mathbf{y}|\mathbf{X}_m)+\max_m W(\mathbf{y}|\mathbf{X}_m)>e^{n\beta}Q_0(\mathbf{y}),\right. \\
&\qquad \left.\max_{m>1} W(\mathbf{y}|\mathbf{X}_m)\geq W(\mathbf{y}|\mathbf{x}_1)\middle|\mathbf{x}_1,\mathbf{y}\right\} &(121)\\
&\doteq \ \Pr\left\{e^{n[\alpha]_+}W(\mathbf{y}|\mathbf{x}_1)+e^{n\alpha}\sum_{m>1} W(\mathbf{y}|\mathbf{X}_m)+\max_{m>1} W(\mathbf{y}|\mathbf{X}_m)>e^{n\beta}Q_0(\mathbf{y}),\right. \\
&\qquad \left.\max_{m>1} W(\mathbf{y}|\mathbf{X}_m)\geq W(\mathbf{y}|\mathbf{x}_1)\middle|\mathbf{x}_1,\mathbf{y}\right\} &(122)\\
&\doteq \ F_1(\mathbf{x}_1,\mathbf{y})+F_2(\mathbf{x}_1,\mathbf{y})+F_3(\mathbf{x}_1,\mathbf{y}) &(123)
\end{aligned}$$

where

$$F_1(\mathbf{x}_1,\mathbf{y})=\mathcal{I}\left\{W(\mathbf{y}|\mathbf{x}_1)\geq e^{n(\beta-[\alpha]_+)}Q_0(\mathbf{y})\right\}\cdot\Pr\left\{\max_{m>1} W(\mathbf{y}|\mathbf{X}_m)\geq W(\mathbf{y}|\mathbf{x}_1)\middle|\mathbf{x}_1,\mathbf{y}\right\}, \tag{124}$$

$$F_2(\mathbf{x}_1,\mathbf{y})=\Pr\left\{\sum_{m>1} W(\mathbf{y}|\mathbf{X}_m)\geq e^{n(\beta-\alpha)}Q_0(\mathbf{y}),\ \max_{m>1} W(\mathbf{y}|\mathbf{X}_m)\geq W(\mathbf{y}|\mathbf{x}_1)\middle|\mathbf{x}_1,\mathbf{y}\right\}, \tag{125}$$

and

$$F_3(\mathbf{x}_1,\mathbf{y})=\Pr\left\{\max_{m>1} W(\mathbf{y}|\mathbf{X}_m)\geq\max\{e^{n\beta}Q_0(\mathbf{y}),\ W(\mathbf{y}|\mathbf{x}_1)\}\middle|\mathbf{x}_1,\mathbf{y}\right\}. \tag{126}$$

We next analyze separately each one of these three terms. For $F_1(\mathbf{x}_1,\mathbf{y})$, we may use an expression similar to the regular random coding exponent, but with an additional constraint on $\tilde{Q}$ (similar to the derivation

of $E_B(R)$ in the FA exponent)

$$F_1(\mathbf{x}_1,\mathbf{y}) \;=\; \mathcal{I}\left\{W(\mathbf{y}|\mathbf{x}_1) \geq e^{n(\beta-[\alpha]_+)}Q_0(\mathbf{y})\right\}\cdot \Pr\left\{\max_{m>1} W(\mathbf{y}|\mathbf{X}_m) > W(\mathbf{y}|\mathbf{x}_1)\right\} \tag{127}$$

$$\doteq\; \exp\left[-n\cdot \min_{\tilde{Q}_{Y|X}\in\mathcal{Q}_P:\, D(\tilde{Q})\leq[\alpha]_+-\beta}\left\{\mathcal{D}(\tilde{Q}_{Y|X}||W|P)+\Big[I(\tilde{Q})-R\Big]_+\right\}\right] \tag{128}$$

with $\tilde{Q}=P\times\tilde{Q}_{Y|X}$. So we obtain $F_1\triangleq \mathbf{E}\{F_1(\mathbf{X}_1,\mathbf{Y})\}\doteq e^{-nE_1}$ with

$$E_1=\min_{Q_{Y|X}:\, D(Q)\leq[\alpha]_+-\beta}\left\{\mathcal{D}(Q_{Y|X}||W|P)+[I(Q)-R]_+\right\} \tag{129}$$

with $Q=P\times Q_{Y|X}$.

Next, we continue with $F_3(\mathbf{x}_1,\mathbf{y})$. We consider the set

$$\mathcal{S}\triangleq\left\{\tilde{Q}:\;\; D(\tilde{Q})>-\beta\right\} \tag{130}$$

We split the evaluation of $F_3(\mathbf{x}_1,\mathbf{y})$ into two cases, depending whether $\tilde{Q}\in\mathcal{S}$ or not.

Case 1: $\tilde{Q}\notin\mathcal{S}$. In this case $\tilde{Q}$ is such that $e^{n\beta}Q_0(\mathbf{y})\leq W(\mathbf{y}|\mathbf{x}_1)$ and so

$$F_3(\mathbf{x}_1,\mathbf{y}) \;=\; \Pr\left\{\max_{m>1} W(\mathbf{y}|\mathbf{X}_m) > W(\mathbf{y}|\mathbf{x}_1)\right\} \tag{131}$$

$$\doteq\; \exp\left[-n\cdot\Big[I(\tilde{Q})-R\Big]_+\right] \tag{132}$$

as in the evaluation of $F_1(\mathbf{x}_1,\mathbf{y})$.

Case 2: $\tilde{Q}\in\mathcal{S}$. In this case $\tilde{Q}$ is such that $e^{n\beta}Q_0(\mathbf{y})> W(\mathbf{y}|\mathbf{x}_1)$ and so

$$F_3(\mathbf{x}_1,\mathbf{y}) \;=\; \Pr\left\{\max_{m>1} W(\mathbf{y}|\mathbf{X}_m) > e^{n\beta}Q_0(\mathbf{y})\right\} \tag{133}$$

$$\doteq\; \exp\left[-n\cdot\left[\min_{Q_{X|Y}\in\mathcal{Q}_P:\, D(\tilde{Q}_Y\times Q_{X|Y})\leq-\beta} I(Q)-R\right]_+\right] \tag{134}$$

$$=\; \exp\left[-n\cdot\Big[\mathbf{R}(-\beta,\tilde{Q}_Y)-R\Big]_+\right]. \tag{135}$$

Then

$$F_3 \;\triangleq\; \mathbf{E}\left[F_3(\mathbf{X}_1,\mathbf{Y})\right] \tag{136}$$

$$\doteq\; \mathbf{E}\Big[F_3(\mathbf{X}_1,\mathbf{Y})\cdot\mathcal{I}(\tilde{Q}\notin\mathcal{S})\Big]+\mathbf{E}\Big[F_3(\mathbf{X}_1,\mathbf{Y})\cdot\mathcal{I}(\tilde{Q}\in\mathcal{S})\Big] \tag{137}$$

$$\doteq\; \max\left\{e^{-nE_3'},e^{-nE_3''}\right\} \tag{138}$$

$$\doteq\; e^{-n\cdot\min\left\{E_3',E_3''\right\}} \tag{139}$$

where

$$E_3' = \min_{\tilde{Q}:\, D(\tilde{Q})\le -\beta} \left\{ \mathcal{D}(\tilde{Q}_{Y|X}||W|P) + \Big[I(\tilde{Q}) - R \Big]_+ \right\} \tag{140}$$

and

$$E_3'' = \inf_{\tilde{Q}:\, D(\tilde{Q})> -\beta} \left\{ \mathcal{D}(\tilde{Q}_{Y|X}||W|P) + \Big[\mathbf{R}(-\beta, \tilde{Q}_Y) - R \Big]_+ \right\}. \tag{141}$$

Using the fact that $\alpha \le 0$, we get $E_3' = E_1$. Thus if we denote $F_3 \triangleq \mathbf{E}\{F_3(\mathbf{X}_1, \mathbf{Y})\} \doteq e^{-nE_3}$, we obtain

$$E_3 = \min\left\{E_3', E_3''\right\} = \min\left\{E_1, E_3''\right\} \le E_1 \tag{142}$$

and so, E_1 is always dominated by E_3.

Finally, let us consider $F_2(\mathbf{x}_1, \mathbf{y})$. Again, let us consider a set

$$\mathcal{S} \triangleq \left\{ \tilde{Q} :\ D(\tilde{Q}) \ge \alpha - \beta \right\} \tag{143}$$

and split into two cases, such that

$$F_2 \triangleq \mathbf{E}\left[F_2(\mathbf{X}_1, \mathbf{Y})\right] = \max\left\{ e^{-nE_2'}, e^{-nE_2''} \right\} = e^{-n\cdot\min\left\{E_2', E_2''\right\}} \tag{144}$$

and each of the two exponents fits one of the following cases.

Case 1: $\tilde{Q} \notin \mathcal{S}$. In this case, $e^{n(\beta-\alpha)} Q_0(\mathbf{y}) \le W(\mathbf{y}|\mathbf{x}_1)$ so using Bayes rule

$$\begin{aligned} F_2(\mathbf{x}_1, \mathbf{y}) &= \Pr\left\{ \sum_{m>1} W(\mathbf{y}|\mathbf{X}_m) > e^{n(\beta-\alpha)} Q_0(\mathbf{y}), \max_{m>1} W(\mathbf{y}|\mathbf{X}_m) \ge W(\mathbf{y}|\mathbf{x}_1) \right\} && (145)\\ &= \Pr\left\{ \max_{m>1} W(\mathbf{y}|\mathbf{X}_m) \ge W(\mathbf{y}|\mathbf{x}_1) \right\} \times \\ &\quad \Pr\left\{ \sum_{m>1} W(\mathbf{y}|\mathbf{X}_m) > e^{n(\beta-\alpha)} Q_0(\mathbf{y}) | \max_{m>1} W(\mathbf{y}|\mathbf{X}_m) \ge W(\mathbf{y}|\mathbf{x}_1) \right\} && (146)\\ &= \Pr\left\{ \max_{m>1} W(\mathbf{y}|\mathbf{X}_m) \ge W(\mathbf{y}|\mathbf{x}_1) \right\} \times 1. && (147) \end{aligned}$$

The resulting exponent is

$$E_2' = \min_{Q:\, D(Q)\le \alpha-\beta} \left\{ \mathcal{D}(Q_{Y|X}||W|P) + \left[I(Q) - R \right]_+ \right\}, \tag{148}$$

and evidently, as $\alpha \le 0$ then $E_2' \ge E_1$.

Case 2: $\tilde{Q} \in \mathcal{S}$. In this case, $e^{n(\beta-\alpha)} Q_0(\mathbf{y}) > W(\mathbf{y}|\mathbf{x}_1)$ so we use the following derivation.

$$F_2(\mathbf{x}_1, \mathbf{y}) \ \doteq\ \Pr\left\{ \sum_{\hat{Q}_{X|Y}} N(\hat{Q}|\mathbf{y}) e^{nf(\hat{Q})} \ge e^{ng(\hat{Q})},\ \sum_{\hat{Q}_{X|Y}} \mathcal{I}\{N(\hat{Q}|\mathbf{y}) \ge 1\} \cdot e^{nf(\hat{Q})} \ge e^{nf(\tilde{Q})} \right\} \tag{149}$$

$$\dot{=} \quad \Pr\left\{\max_{\hat{Q}_{X|Y}} N(\hat{Q}|\mathbf{y})e^{nf(\hat{Q})} \geq e^{ng(\hat{Q})},\ \max_{\hat{Q}_{X|Y}} \mathcal{I}\{N(\hat{Q}|\mathbf{y}) \geq 1\} \cdot e^{nf(\hat{Q})} \geq e^{nf(\tilde{Q})}\right\} \tag{150}$$

$$\dot{=} \quad \Pr\left[\bigcup_{\hat{Q}_{X|Y}} \{N(\hat{Q}|\mathbf{y}) \geq e^{nu(\hat{Q})}\}\right] \bigcap \left[\bigcup_{\hat{Q}_{X|Y}} \{\mathcal{I}\{N(\hat{Q}|\mathbf{y}) \geq 1\} \geq e^{n[f(\tilde{Q})-f(\hat{Q})]}\}\right] \tag{151}$$

$$= \quad \Pr\left[\bigcup_{\hat{Q}_{X|Y}} \{N(\hat{Q}|\mathbf{y}) \geq e^{nu(\hat{Q})}\}\right] \bigcap$$

$$\left[\bigcup_{\hat{Q}_{X|Y}:\, f(\tilde{Q}) \leq f(\hat{Q})} \{\mathcal{I}\{N(\hat{Q}|\mathbf{y}) \geq 1\} \geq e^{n[f(\tilde{Q})-f(\hat{Q})]}\}\right] \tag{152}$$

$$= \quad \Pr \bigcup_{\{\hat{Q}_{X|Y}, Q'_{X|Y}:\, f(\tilde{Q}) \leq f(Q')\}} \left\{N(\hat{Q}|\mathbf{y}) \geq e^{nu(\hat{Q})},\ N(Q'|\mathbf{y}) \geq 1\right\} \tag{153}$$

$$\dot{=} \quad \Pr \bigcup_{\hat{Q}_{X|Y}:\, f(\tilde{Q}) \leq f(\hat{Q})} \left\{N(\hat{Q}|\mathbf{y}) \geq e^{n[u(\hat{Q})]_+}\right\} +$$

$$\sum_{\hat{Q}_{X|Y} \neq Q'_{X|Y}:\, f(\tilde{Q}) \leq f(Q')} \Pr\{N(\hat{Q}|\mathbf{y}) \geq e^{n[u(\hat{Q})]_+},\ N(Q'|\mathbf{y}) \geq 1\} \tag{154}$$

$$\dot{=} \quad \max_{\hat{Q}_{X|Y}:\, f(\tilde{Q}) \leq f(\hat{Q})} \Pr\left\{N(\hat{Q}|\mathbf{y}) \geq e^{n[u(\hat{Q})]_+}\right\} +$$

$$\max_{\hat{Q}_{X|Y} \neq Q'_{X|Y}:\, f(\tilde{Q}) \leq f(Q')} \Pr\{N(\hat{Q}|\mathbf{y}) \geq e^{n[u(\hat{Q})]_+},\ N(Q'|\mathbf{y}) \geq 1\} \tag{155}$$

$$\dot{=} \quad \max_{\hat{Q}_{X|Y}:\, f(\tilde{Q}) \leq f(\hat{Q})} \Pr\{N(\hat{Q}|\mathbf{y}) \geq e^{n[u(\hat{Q})]_+}\} \tag{156}$$

$$= \quad \max_{\hat{Q}_{X|Y}:\, D(\tilde{Q}) \geq D(\hat{Q})} \Pr\{N(\hat{Q}|\mathbf{y}) \geq e^{n[u(\hat{Q})]_+}\} \tag{157}$$

$$\dot{=} \quad \exp\left[-n \cdot \left[\min_{\hat{Q}_{X|Y}:\, D(\hat{Q}) \leq D(\tilde{Q}), D(\hat{Q}) \leq \alpha-\beta} I(Q) - R\right]_+\right], \tag{158}$$

where all the conditional types satisfy $\tilde{Q}_{X|Y} \in \mathcal{Q}_P, Q'_{X|Y} \in \mathcal{Q}_P, \hat{Q}_{X|Y} \in \mathcal{Q}_P$, and the last passage follows from an analysis almost identical to that of E_A in Subsection V-A.

After averaging over $(\mathbf{X}_1, \mathbf{Y})$, the resulting exponent is

$$E''_2 \quad = \quad \inf_{\tilde{Q}:\, D(\tilde{Q}) > \alpha-\beta} \left\{\mathcal{D}(\tilde{Q}_{Y|X}||W|P) + \left[\min_{\hat{Q}_{X|Y} \in \mathcal{Q}_P:\, D(\hat{Q}) \leq D(\tilde{Q}), D(\hat{Q}) \leq \alpha-\beta} I(Q) - R\right]_+\right\} \tag{159}$$

$$= \quad \inf_{\tilde{Q}:\, D(\tilde{Q}) > \alpha-\beta} \left\{\mathcal{D}(\tilde{Q}_{Y|X}||W|P) + \left[\min_{\hat{Q}_{X|Y} \in \mathcal{Q}_P:\, D(\hat{Q}) \leq \alpha-\beta} I(Q) - R\right]_+\right\} \tag{160}$$

$$= \quad \inf_{\tilde{Q}:\, D(\tilde{Q}) > \alpha-\beta} \left\{\mathcal{D}(\tilde{Q}_{Y|X}||W|P) + \left[\mathbf{R}(\alpha-\beta, \tilde{Q}_Y) - R\right]_+\right\}. \tag{161}$$

To conclude, we have

$$E_{\text{EE}}(R) \triangleq \min\left\{E''_2, E'_3, E''_3\right\}. \tag{162}$$

VI. Numerical Examples

In this section, we demonstrate the results obtained via numerical examples. First, we consider an example of practical interest, and second, we compare our results to the non-optimal exponents derived in [26].

Consider transmitting over an additive white Gaussian noise (AWGN) channel in discrete time, in which the output at time $k \in \{1, \ldots, n\}$ is given by

$$\tilde{y}_k = x_k + w_k. \tag{163}$$

The channel input x_k is restricted to a ternary input alphabet $\mathcal{X} = \{-1, 0, 1\} \subset \mathbb{R}$, and a fixed composition codebook is drawn according to the input type $P = [1/2, 0, 1/2]$. When the transmitter is silent, $x_k = 0$. The noise is Gaussian i.i.d. with $\mathcal{N}(0, 1/\text{SNR})$ where SNR is the signal to noise ratio (SNR). The output $\tilde{y}_k$ is quantized into a ternary alphabet $\mathcal{Y} = \{-1, 0, 1\}$ as follows:

$$y_k = \begin{cases} -1 & \tilde{y}_k \leq -1/2 \\ 0 & -1/2 < \tilde{y}_k < 1/2 \\ 1 & \tilde{y}_k \geq 1/2 \end{cases}. \tag{164}$$

After output quantization, a DMC, parametrized by the SNR value, is obtained. For example, when SNR $= 4$dB then

$$\begin{bmatrix} W(-1|-1) & W(0|-1) & W(1|-1) \\ W(-1|0) & W(0|0) & W(1|0) \\ W(-1|1) & W(0|1) & W(1|1) \end{bmatrix} = \begin{bmatrix} 0.786, & 0.2053, & 0.0087 \\ 0.21405, & 0.5719, & 0.21405 \\ 0.0087, & 0.2053, & 0.786 \end{bmatrix}, \tag{165}$$

where $Q_0 = W(\cdot|0)$. The capacity of the original AWGN channel (with optimal input and no output quantization) is $C \approx 0.628$ (nats) and the mutual information of the resulting DMC with the assumed input type P is $I(P \times W) \approx 0.5$ (nats). The obtained exponents for the parameters $\alpha = 0, \beta = 0.2$ are shown in Figure 1 and 2 for SNR $= 0$dB and SNR $= 4$dB, respectively. For visual clarity, the exponent $E_{\text{IE}}(R)$ is not shown, as it is easily given by the minimum of $E_{\text{MD}}(R)$ and $E_{\text{r}}(R)$.

It can be observed that the FA exponent has the same form as the random coding exponent. For low rates it is an affine function of the rate with slope -1, and for larger rates it has decreases monotonically to 0. Moreover, E_{FA} is convex for all positive rates. In general, since $E_{\text{FA}}(R)$ is the pointwise minimum of two decreasing convex functions (see Appendix A), it is only convex in intervals where one exponent dominates the other. It is also worth to observe that the MD exponent is constant up to some critical rate, and then increases (this may also be deduced from eq. (40)). Indeed, the fact that many codewords exist

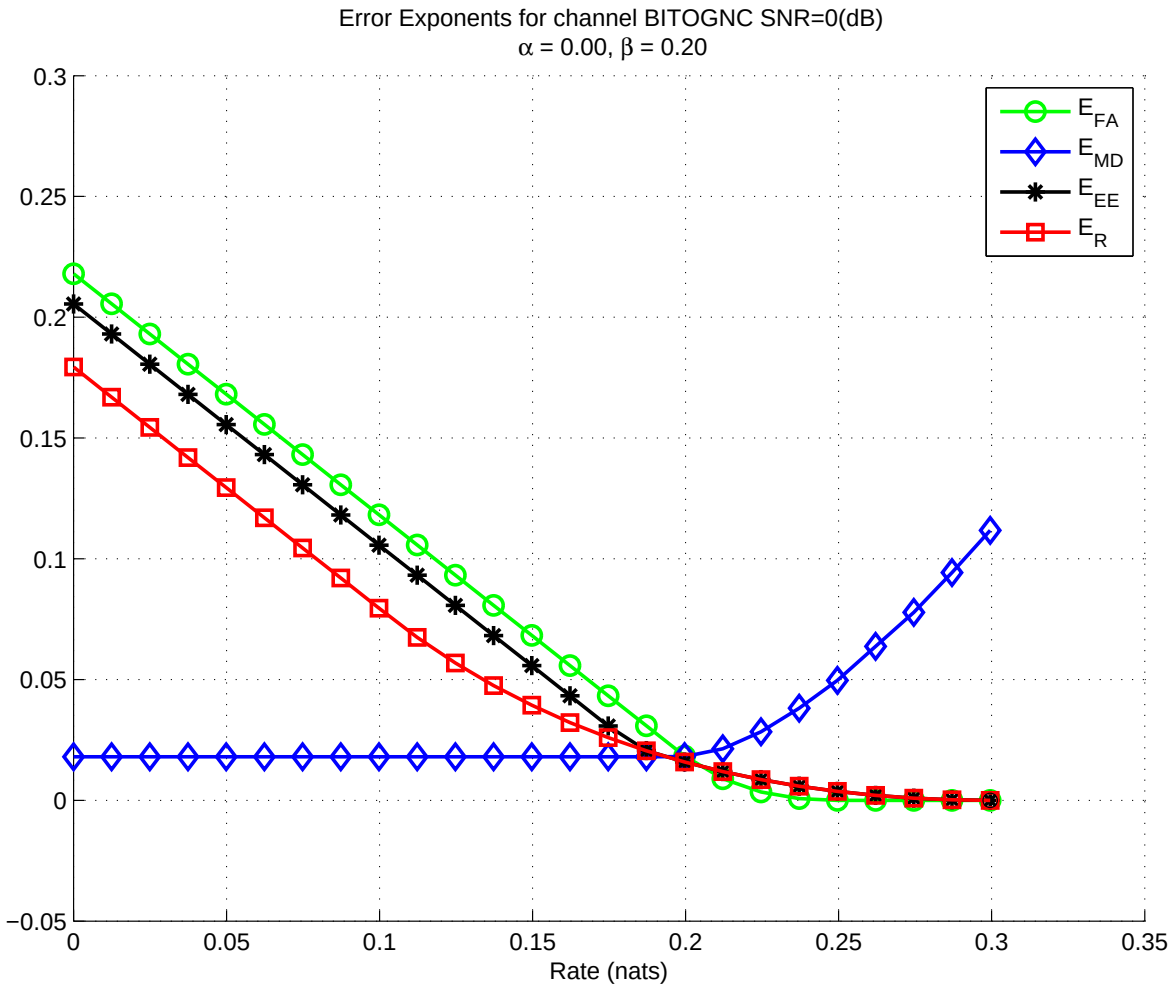

Figure 1. Error exponents for binary-input-ternary-output Gaussian channel and SNR $=$ 0dB, $\alpha = 0$, $\beta = 0.2$

in the codebook decreases the MD probability, even if the decoding will eventually not be to the true codeword. The IE exponent, given by the minimum of $E_{\text{MD}}(R)$ and $E_{\text{r}}(R)$ is increasing up to the rate where these two exponents equate (assuming $E_{\text{MD}}(0) \leq E_{\text{r}}(0)$), and then decreases to zero at $R = I(P \times W)$. The qualitative description of the EE exponent as a function of the rate is similar to the FA exponent. Finally, notice that for the larger SNR, $E_{\text{EE}}(R)$ reaches its lower bound $E_{\text{r}}(R)$ for all rates.

Next, we compare the optimal random coding exponents derived in this paper, with the results of [26]. In [26, Chapter 4], Wang studies the achievable tradeoff between $E_{\text{FA}}(R)$ and $E_{\text{MD}}(R)$ at the capacity $R = I(P \times W)$, which obviously implies that $E_{\text{IE}}(R) = 0$. Among the ensembles considered in [26], the best achievable trade-off is obtained for fixed composition codebooks, as assumed in this paper, but a heuristic decoding rule is used (see [26, Theorem 4.2] and comment 3 in Section (III)). Another difference is that the MD probability in [26] is not averaged over all codewords in the codebook (as in eq. 7), and is defined as the maximal probability over all codewords in the codebook. Nonetheless, this difference is immaterial since for any codebook with a given average misdetection probability, a different codebook

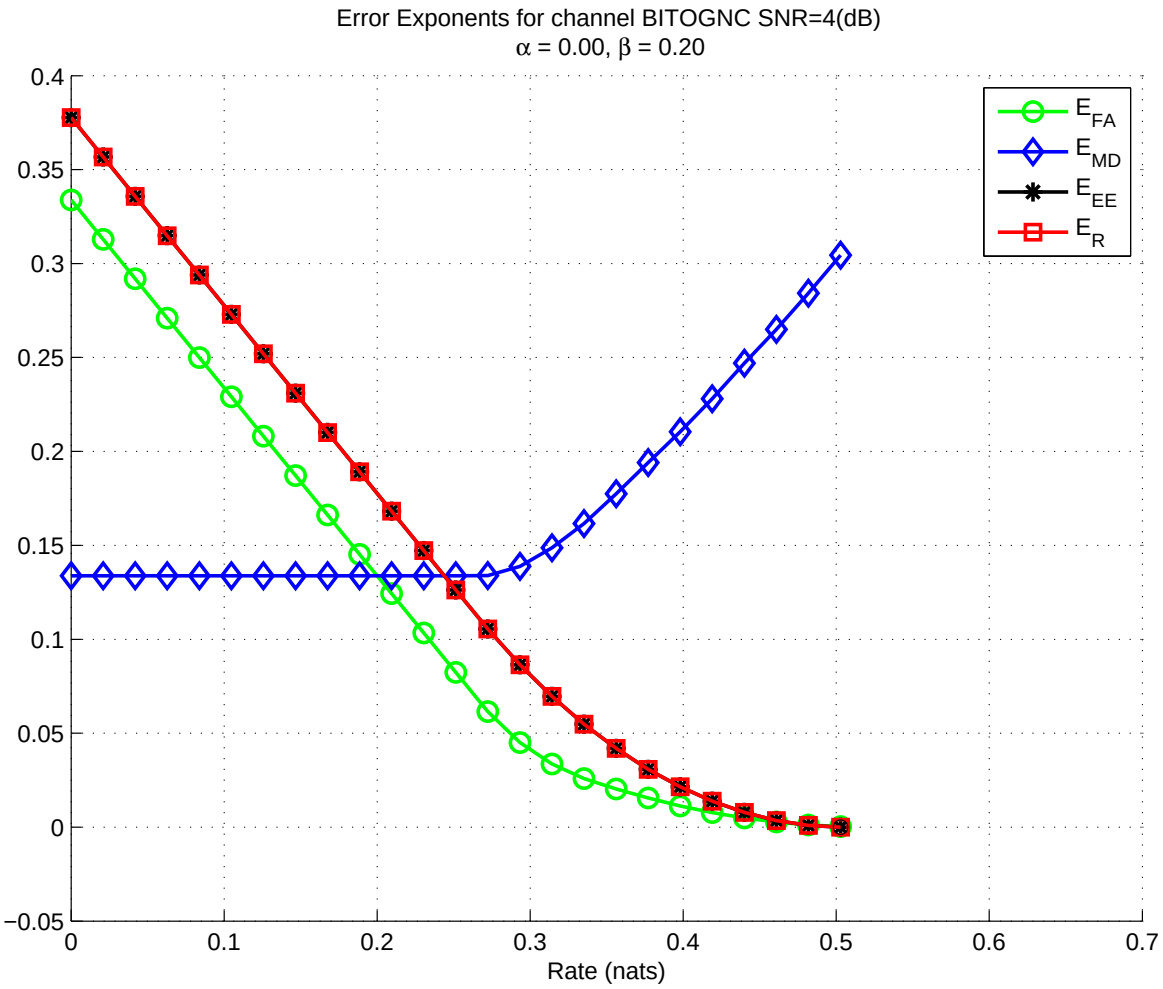

Figure 2. Error exponents for binary-input-ternary-output Gaussian channel and SNR $= 0$dB, $\alpha = -0.1$, $\beta = 0.2$

can be found with the same maximal misdetection probability, and a negligible loss in rate.

One of the numerical examples in [26, Section 4.4.4] assumes an input alphabet $\mathcal{X} = \{-1, 0, 1\}$ and output alphabet $\mathcal{Y} = \{-1, 1\}$. The relation between the inputs $\mathcal{X}\backslash\{0\} = \{-1, 1\}$ to $\mathcal{Y}$ is an ordinary binary symmetric channel (BSC) with crossover probability $\epsilon = 0.05$, and for the special symbol $Q_0 = [1/2, 1/2]$. Moreover, the symbol $'0'$ is restricted not to be used for coding. In this event, since the channel is symmetric, a uniform input type $P = [1/2, 0, 1/2]$ induces Q_0 on the output letters, just like the special symbol. Thus, good results are obtained when the input type is not uniform, and $P = [0.4, 0, 0.6]$ is chosen as an example, which results $I(P \times W) = 0.48$ (nats). A numerical calculation we have performed showed that in this case the exponents in [26, Section 4.4.4] and the optimal exponents derived in this paper coincide.[10] However, if one considers a binary *non-symmetric* channel between the inputs $\{-1, 1\}$ and $\mathcal{Y}$ with $\epsilon_1 = W(1|-1) = 0.01$ and $\epsilon_2 = W(1|0) = 0.3$ (and all other parameters are the same), then

[10]See [26, Figure 4.5 in Section 4.4.4] for the actual graphs. Notice that the figures in [26] are presented in base 2 logarithm, even though expressions of the form e^{-nE} are used for exponential decrease.

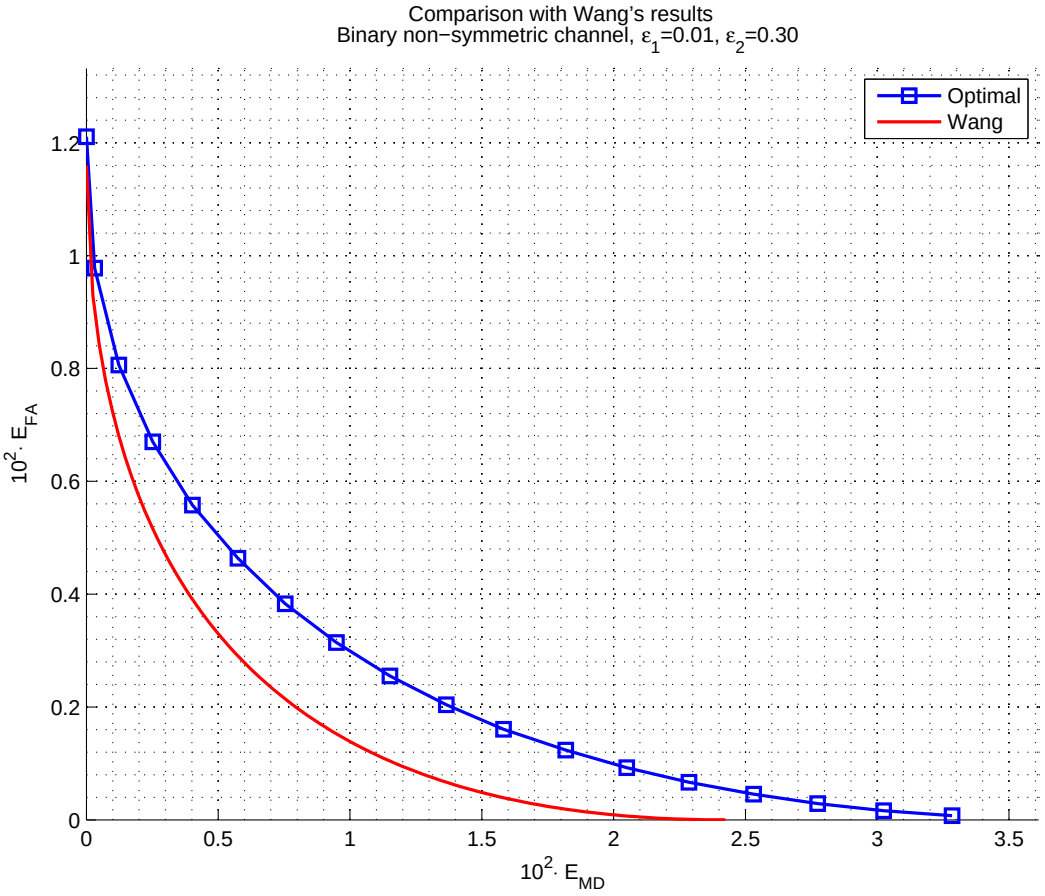

Figure 3. Comparison with [26, Theorem 4.2]

Wang's results are strictly sub-optimal, as can be seen in Figure 3. We remark that for optimal results, we have chosen $\alpha = 0$ (see discussion in Section III) and β was varied in order to trade-off between the false alarm and misdetection exponents.

VII. Directions for Future Research

The main contribution of this paper is the derivation of the optimal detector for the transmission of a codeword, as well as its exact random coding error exponents for the ensemble of fixed composition codes of a given type P. From this point, there is a variety of possible future research directions:

1) We have focused on discrete memoryless channels, but the same problem may be formulated for additive Gaussian noise channels. For these channels, the optimal detector/decoder is exactly the same as in Lemma 1, while the error exponents analysis may be performed by methods similar, e.g., to [13], where an analogue of the method of types is developed for the continuous alphabet case.

2) Throughout, we have assumed a given input type P. Clearly, P may be optimized to obtain the largest exponents possible, as long as $I(P \times W) \geq R$ is satisfied. Notice, however, that three exponents are involved in this optimization, and by varying P one may improve one exponent, but decrease either

of the two others. This can be circumvented by defining a cost function that weighs all three exponents as a Lagrangian, and then optimize the cost function with respect to P. As a subset of this problem, it may be only decided whether to use the special symbol for coding or not. If the P-probability of $'0'$ is strictly positive, then the mutual information $I(P \times W)$ and the IE and EE exponents may be increased, but clearly the FA and MD exponents will decrease.

3) A related direction is improving the ensemble of codebooks, e.g. using expurgated ensembles or an ensemble of linear codes. As for expurgated bounds, here too, one should define which codewords to expurgate, since removing some of the codewords from a given code may improve one exponent, but perhaps harm the other exponents. Again, a cost function may be defined, and then methods that similar to those of [16] and [19] may be applied. As for linear ensembles of codes, currently, the type-class enumeration method relies on the fact that type-class enumerator[11] is a binomial random variable, namely, it counts 'successes' of mutually independent experiments. For the linear ensemble, these experiments are merely pairwise independent, so it is not trivial to characterize the large deviations behavior of the enumerators. Though one can always resort to bounds, our aim for further research are the exact exponents.

4) The problem discussed in this paper is intimately related to the message-wise UEP problem [2] where a specific message out of the codebook has a special stature, and similarly to our problem, FA and MD probabilities may be defined with respect to the special message. It is apparent, that no matter what is the codeword of the special message, the optimal detector/decoder in Lemma 1 can be used verbatim, by letting $Q_0(\mathbf{y})$ to be the output distribution, given that the special codeword has been sent. Also, in case that the special message is $\mathbf{0}$[12], both problems are equivalent, and our results apply, and thus generalize the results of [2, Theorem 2, Theorem 10], which only considered rates close to capacity and either zero FA exponent or zero MD exponent. Exploring further relations between these problems is left for further study.

5) In many cases, in addition to the codeword detection problem, the classical problem of channel knowledge is also present [11]. To perform optimal detection/decoding, the receiver should have an up-to-date knowledge of the channel law W. The problem of designing joint detection/decoding rules remains relevant, and even becomes more involved, when the receiver is ignorant of the exact underlying law governing the channel. Optimal detectors/decoders should be developed for this scenario, and their performance should be evaluated.

6) The problem is easily extended to network scenarios, e.g., detecting, in a multiple access channel, which of the users is currently transmitting, and then decode the messages of the active users.

[11]The quantity defined as $N(\hat{Q}|\mathbf{y})$ in the proof of Theorem 2.

[12]As proposed in [2, Section III.B] in cases that maximal MD exponent is required, or maximal (lower bound on) FA exponent is required, an optimal strategy is to repeatedly send the same symbol.

ACKNOWLEDGMENTS

Useful comments made by the Associate Editor, Aaron Wagner, and by the anonymous referees are acknowledged with thanks.

APPENDIX A
ASPECTS OF EXPONENTS COMPUTATION

In this appendix we discuss several aspects of the computation of the various exponent functions appearing in Section IV. Let us recall the regular random coding expression for fixed composition codes is given by

$$E_{\mathrm{r}}(R) = \min_{Q_{Y|X}} \mathcal{D}(Q_{Y|X}||W|P) + \left[I(P \times Q_{Y|X}) - R\right]_+ . \tag{A.1}$$

The objective function of this minimization problem includes clipping, which complicates the computation of $E_{\mathrm{r}}(R)$, as it results a minimax problem. The standard method to solve this optimization problem is to first solve it for $R = 0$. In this case

$$E_{\mathrm{r}}(0) = \min_{Q_{Y|X}} \mathcal{D}(Q_{Y|X}||W|P) + I(P \times Q_{Y|X}). \tag{A.2}$$

which does not contain the problematic clipping. Now, letting the minimizer be $Q^0_{Y|X}$, and defining the critical rate

$$R_{cr} \triangleq I(P \times Q^0_{Y|X}) \tag{A.3}$$

it is easily noticed that for $R \leq R_{cr}$ we have

$$E_{\mathrm{r}}(R) = \mathcal{D}(Q^*_{Y|X}(0)||W|P) + R_{cr} - R \tag{A.4}$$

namely, $E_{\mathrm{r}}(R)$ decreases linear with slope -1. Then, for $R > R_{cr}$ we have

$$E_{\mathrm{r}}(R) = \min_{Q_{Y|X}: I(P \times Q_{Y|X}) \leq R} \mathcal{D}(Q_{Y|X}||W|P) \tag{A.5}$$

which is a constrained optimization problem. However, since $I(Q)$ is a convex functional of $Q_{Y|X}$ then this is a convex optimization problem which can efficiently be solved numerically. Moreover, using standard Lagrange formulation this problem may be written as the following an unconstrained problem

$$\begin{aligned} E_{\mathrm{r}}(R) &= \min_{Q_{Y|X}} \max_{\lambda \geq 0} \left\{\mathcal{D}(Q_{Y|X}||W|P) + \lambda \cdot \left(I(P \times Q_{Y|X}) - R\right)\right\} && \text{(A.6)} \\ &= \max_{\lambda \geq 0} \min_{Q_{Y|X}} \left\{\mathcal{D}(Q_{Y|X}||W|P) + \lambda \cdot \left(I(P \times Q_{Y|X}) - R\right)\right\}. && \text{(A.7)} \end{aligned}$$

For a given λ, let $Q_{Y|X}^{\lambda}$ be the minimizer of the inner problem. It is well known that as λ is varied from 1 to 0 then the random coding exponent is given parametrically by

$$\begin{cases} E_{\mathrm{r}}(R) = \mathcal{D}(Q_{Y|X}^{\lambda}||W|P) \\ R = I(P \times Q_{Y|X}^{\lambda}) \end{cases}, \tag{A.8}$$

where for $\lambda = 0$ the error exponent vanishes, and the rate is the mutual information $I(P \times W)$.

In what follows, we shall attempt to represent the various exponents in forms similar to the ordinary random coding exponent, in order to prove the feasibility of their computation. This is possible for E_{FA} and E_{EE}, however computing E_{MD} is more evolved as discussed in Subsection A-B of this appendix.

A. Computation of False Alarm Exponents

We begin with the expression $E_A(R)$ from eq. (37). While this is a very compact expression, it is difficult to compute its value because of the clipping operations. Thus, we split the minimization in eq. (37) into two cases $I(Q) \leq R$ and $I(Q) > R$. Then, we may write

$$E_A(R) = \min\left\{E_A'(R), E_A''(R)\right\} \tag{A.9}$$

where

$$E_A'(R) \triangleq \min_{Q_{Y|X}:\, I(Q)\leq R,\, D(Q)+I(Q)\leq -\beta+\alpha+R} \left\{\mathcal{D}(Q_Y||Q_0)\right\}, \tag{A.10}$$

$$E_A''(R) \triangleq \min_{Q_{Y|X}:\, I(Q)> R,\, D(Q)\leq -\beta+\alpha} \left\{\mathcal{D}(Q_Y||Q_0) + I(Q) - R\right\}. \tag{A.11}$$

Now, computing $E_A'(R)$ is a convex optimization problem since $D(Q)$ is a linear function of Q (cf. eq. (32) and the definition of $D(\cdot)$ immediately after). Next, let us analyze

$$E_A''(R) = \inf_{Q_{Y|X}:\, D(Q)\leq -\beta+\alpha, I(Q)>R} \mathcal{D}(Q_Y||Q_0) + I(Q) - R. \tag{A.12}$$

as a function of the rate. For $R = 0$ we have

$$E_A''(R=0) = \min_{Q_{Y|X}:\, D(Q)\leq -\beta+\alpha} \mathcal{D}(Q_Y||Q_0) + I(Q) \tag{A.13}$$

which is a convex optimization problem. Suppose that its unique solution is $(Q_{A,R=0})_{Y|X}$ and let

$$R_{cr,A} \triangleq I(P \times (Q_{A,R=0})_{Y|X}). \tag{A.14}$$

Then, for $R \leq R_{cr,A}$ the exponent is an affine function

$$E_A''(R) = E_A''(R=0) - R \tag{A.15}$$

and for $R > R_{cr,A}$

$$E''_A(R) = \inf_{Q_{Y|X}:\, D(Q)\le -\beta+\alpha, I(Q)>R} \mathcal{D}(Q_Y||Q_0) + I(Q) - R \tag{A.16}$$

$$= \min_{Q_{Y|X}:\, D(Q)\le -\beta+\alpha, I(Q)=R} \mathcal{D}(Q_Y||Q_0), \tag{A.17}$$

due to the convexity of the objective. But, since

$$E'_A(R) = \min_{Q_{Y|X}:\, I(Q)\le R, D(Q)+I(Q)\le -\beta+\alpha+R} \mathcal{D}(Q_Y||Q_0) \tag{A.18}$$

then for $R > R_{cr,A}$ we get

$$E'_A(R) \le E''_A(R). \tag{A.19}$$

Thus, for the purpose of computing $E_A(R)$ we may disregard $E''_A(R)$ for $R > R_{cr,A}$, or equivalently

$$E_A(R) = \min\{E'_A(R), \tilde{E}''_A(R)\} \tag{A.20}$$

where

$$\tilde{E}''_A(R) = \begin{cases} \min_{Q_{Y|X}:\, D(Q)\le -\beta+\alpha} \mathcal{D}(Q_Y||Q_0) + I(Q) - R, & R \le R_{cr,A} \\ \infty & R > R_{cr,A} \end{cases}. \tag{A.21}$$

Then, computing both $E'_A(R)$ and $\tilde{E}''_A(R)$ only requires solving convex optimization problems.

Next, let us discuss $E_B(R)$. Again, for $R = 0$

$$E_B(R=0) = \min_{Q_{Y|X}:\, D(Q)\le -\beta} \{\mathcal{D}(Q_Y||Q_0) + I(Q)\} \tag{A.22}$$

where as usual $P \times Q_{Y|X}$, and let $(Q_{B,R=0})_{Y|X}$ be the minimizer for this problem. Then we may similarly define

$$R_{cr,B} \triangleq I(P \times (Q_{B,R=0})_{Y|X}) \tag{A.23}$$

and for $R \le R_{cr,B}$

$$E_B(R) = \mathcal{D}((P \times Q^0_{Y|X})_Y||Q_0) + R_{cr} - R. \tag{A.24}$$

Then, for $R > R_{cr,B}$ we have

$$E_B(R) = \min_{Q_{Y|X}:\, I(P\times Q_{Y|X})\le R, D(P\times Q_{Y|X})\le -\beta} \mathcal{D}((P \times Q_{Y|X})_Y||Q_0). \tag{A.25}$$

All optimization problems involved in computing $E_B(R)$ are convex since the constraint $D(Q) \le -\beta$ is linear, and the divergence involved is convex. Moreover, a parametric form of $E_B(R)$ may be obtained, just as for the ordinary random coding exponent.

Further simplifications are possible, recalling that without loss of generality it may be assumed that $\alpha \leq 0$. When $\alpha = 0$ then $R_{cr,A} = R_{cr,B}$ and $E_B(R) = \tilde{E}''_A(R)$ for $R \leq R_{cr,A} = R_{cr,B}$. For $R > R_{cr,A}$ we have

$$\tilde{E}''_A(R) \geq E_B(R) \tag{A.26}$$

and thus

$$E_{FA}(R) = \min\left\{E'_A(R), E_B(R)\right\}. \tag{A.27}$$

Now if α is reduced to be strictly less than 0 then $E_B(R)$ does not change, and $E'_A(R)$ and $\tilde{E}''_A(R)$ increase, so

$$\tilde{E}''_A(R) \geq E_B(R) \tag{A.28}$$

and

$$E_{FA}(R) = \min\left\{E'_A(R), E_B(R)\right\}. \tag{A.29}$$

So to conclude, assuming without loss of generality $\alpha \leq 0$, we may disregard $\tilde{E}''_A(R)$. The resulting optimization problems for $E'_A(R)$ and $E_B(R)$ are all convex, and thus solvable.

On a final note, we remark that $E'_A(R)$ and $E_B(R)$ are both convex functions of R (but $E_{FA}(R)$ may not be, as it is a minimum of two convex functions).

B. Computation of Misdetection Exponent

It is observed from eqs. (35) and (40) that $\mathbf{R}(\Delta; Q_Y)$, $\mathbf{S}(\Delta; Q_Y)$ and $R_1(Q_Y), D_1(Q_Y)$ need to be computed for any given Q_Y. It can be easily proved that given Q_Y, computing each of these functions is a convex optimization problem, and thus can be efficiently computed. However, the optimization over Q_Y, needed in order to compute $E_{\text{MD}}(R)$ in eq. (40), is a non-convex problem, and thus global optimization methods are required (e.g. a simple algorithm is an exhaustive search over a fine grid of the $|\mathcal{Y}|$-dimensional probability simplex).

C. Computation of Exclusive Decoding Error Exponent

We begin by analyzing $E'_3(R)$. It can be easily seen that computing $E'_3(R)$ is a convex optimization problem, and thus finding the optimal solution $Q^{'}$ is a feasible task. Then, two cases are possible depending whether $Q^{'}$ has a slack constraint, i.e. $D(Q^{'}) < -\beta$, or not.

Case 1: $D(Q^{'}) < -\beta$. In this case

$$E'_3(R) \;=\; \min_{Q:\, D(Q)\leq -\beta} \left\{\mathcal{D}(Q_{Y|X}||W|P) + \left[I(Q) - R\right]_+\right\} \tag{A.30}$$

$$= \min_{Q} \left\{ \mathcal{D}(Q_{Y|X}||W|P) + [I(Q) - R]_+ \right\} \tag{A.31}$$

$$= E_{\rm r}(R) \tag{A.32}$$

i.e. the regular random coding exponent. Since by definition $E_{\rm EE}(R) \geq E_{\rm r}(R)$ this implies that $E_{\rm EE}(R) = E_{\rm r}(R)$.

Case 2: $D(Q^{'}) = -\beta$. In this case, continuity and convexity arguments imply

$$E_3'(R) = \min_{Q:\, D(Q) \leq -\beta} \left\{ \mathcal{D}(Q_{Y|X}||W|P) + [I(Q) - R]_+ \right\} \tag{A.33}$$

$$= \min_{Q:\, D(Q) = -\beta} \left\{ \mathcal{D}(Q_{Y|X}||W|P) + [I(Q) - R]_+ \right\} \tag{A.34}$$

$$\geq \min_{Q:\, D(Q) = -\beta} \left\{ \mathcal{D}(Q_{Y|X}||W|P) + [\mathbf{R}(-\beta, Q_Y) - R]_+ \right\} \tag{A.35}$$

$$\geq \min_{Q:\, D(Q) \geq -\beta} \left\{ \mathcal{D}(Q_{Y|X}||W|P) + [\mathbf{R}(-\beta, Q_Y) - R]_+ \right\} \tag{A.36}$$

$$= \inf_{Q:\, D(Q) > -\beta} \left\{ \mathcal{D}(Q_{Y|X}||W|P) + [\mathbf{R}(-\beta, Q_Y) - R]_+ \right\} \tag{A.37}$$

$$= E_3''(R) \tag{A.38}$$

and then $E_{\rm EE}(R) = \min\left\{E_2''(R), E_3''(R)\right\}$.

The above analysis may aid to choose β. Since a-priori $E_{\rm EE}(R) \geq E_{\rm r}(R)$, if a strict inequality $E_{\rm EE}(R) > E_{\rm r}(R)$ is desired, then the value of β should be chosen such that Q^*, the minimizer of $E_{\rm r}(R)$, is not feasible for $E_3'(R)$, namely $D(Q^*) > -\beta$.

We continue with $E_2''(R)$. For $R = 0$ we have

$$E_2''(R=0) = \min_{Q_Y} \left\{ \mathbf{S}(\alpha - \beta; Q_Y) + \mathbf{R}(\alpha - \beta; Q_Y) \right\}. \tag{A.39}$$

To find the minimal value we explicitly write

$$E_2''(R=0) = \min_{Q_Y} \left\{ \inf_{Q_{X|Y} \in \mathcal{Q}_P:\, D(Q) > \alpha - \beta} \mathcal{D}(Q_{Y|X}\|W|P) + \min_{\tilde{Q}_{X|Y} \in \mathcal{Q}_P:\, D(\tilde{Q}) \leq \alpha - \beta} I(\tilde{Q}) \right\} \tag{A.40}$$

or equivalently

$$E_2''(R=0) = \inf_{Q_{X|Y} \in \mathcal{Q}_P:\, D(Q) > \alpha - \beta} \; \min_{\tilde{Q}_{X|Y} \in \mathcal{Q}_P:\, D(\tilde{Q}) \leq \alpha - \beta} \mathcal{D}(Q_{Y|X}\|W|P) + I(\tilde{Q}) \quad \text{s.t.} \quad Q_Y = \tilde{Q}_Y. \tag{A.41}$$

Thus, in essence, computing $E_2''(R=0)$ requires minimizing over two conditional distributions $Q_{Y|X}$ and $\tilde{Q}_{Y|X}$ with the additional linear constraint that their Y-marginal are equal. As the objective function is jointly convex in the optimization variables $Q_{Y|X}$ and $\tilde{Q}_{Y|X}$, and the constraints are linear, this is a convex optimization problem.

Then, if the optimal solution is $((Q_{2,R=0})_{Y|X}, (\tilde{Q}_{2,R=0})_{Y|X})$ denoting

$$R_{cr,2} \triangleq I\left(P \times (\tilde{Q}_{2,R=0})_{Y|X}\right) \tag{A.42}$$

we get for $R \leq R_{cr,2}$ an affine section

$$E_2''(R) = \mathcal{D}((Q_{2,R=0})_{Y|X} \| W|P) + R_{cr,2} - R. \tag{A.43}$$

For $R > R_{cr,2}$ we may use the Lagrange formulation

$$\inf_{Q_{X|Y} \in \mathcal{Q}_P:\, D(Q) > \alpha - \beta} \; \min_{\tilde{Q}_{X|Y} \in \mathcal{Q}_P:\, D(\tilde{Q}) \leq \alpha - \beta} \mathcal{D}(Q_{Y|X} \| W|P) + \lambda \cdot I(\tilde{Q}) \quad \text{s.t.} \quad Q_Y = \tilde{Q}_Y. \tag{A.44}$$

Letting the solution of this problem be $((Q_{2,\lambda})_{Y|X}, (\tilde{Q}_{2,\lambda})_{Y|X})$, and solving this problem for $\lambda \in [0,1]$ provides the following parametric representation of the exponent:

$$\begin{cases} E_2''(R) = \mathcal{D}\left((Q_{2,\lambda})_{Y|X} \| W|P\right) \\ R = I\left(P \times (\tilde{Q}_{2,\lambda})_{Y|X}\right) \end{cases}. \tag{A.45}$$

Finally, $E_3''(R)$ has identical structure to $E_2''(R)$ (the only difference is that $-\beta$ is replaced by $\alpha - \beta$), and thus may be computed exactly in the same manner.

REFERENCES

[1] R. H. Barker. Group synchronization of binary digital systems. *Communication Theory*, pages 273–287, 1953.

[2] S. Borade, B. Nakiboglu, and Zheng L. Unequal error protection: An information-theoretic perspective. *Information Theory, IEEE Transactions on*, 55(12):5511–5539, Dec 2009.

[3] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge university press, 2004.

[4] V. Chandar, A. Tchamkerten, and D. Tse. Asynchronous capacity per unit cost. *Information Theory, IEEE Transactions on*, 59(3):1213–1226, March 2013.

[5] V. Chandar, A. Tchamkerten, and G. Wornell. Optimal sequential frame synchronization. *Information Theory, IEEE Transactions on*, 54(8):3725–3728, Aug 2008.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[7] M. Feder and A. Lapidoth. Universal decoding for channels with memory. *IEEE Trans. Inform. Theory*, 44(5):1726–1745, September 1998.

[8] G. D. Forney Jr. Exponential error bounds for erasure, list, and decision feedback schemes. *Information Theory, IEEE Transactions on*, 14(2):206–220, 1968.

[9] L Franks. Carrier and bit synchronization in data communication–a tutorial review. *Communications, IEEE Transactions on*, 28(8):1107–1121, 1980.

[10] S. Golomb, J. Davey, I. Reed, H. Van Trees, and J. Stiffler. Synchronization. *Communications Systems, IEEE Transactions on*, 11(4):481–491, December 1963.

[11] A. Lapidoth and P. Narayan. Reliable communications under channel uncertainty. *Information Theory, IEEE Transactions on*, 44(6):2148–2177, October 1998.

[12] J. L. Massey. Optimum frame synchronization. *Communications, IEEE Transactions on*, 20(2):115–119, 1972.

[13] N. Merhav. Universal decoding for memoryless Gaussian channels with a deterministic interference. *Information Theory, IEEE Transactions on*, 39(4):1261–1269, Jul 1993.

[14] N. Merhav. Statistical physics and information theory. *Foundations and Trends in Communications and Information Theory*, 6(1-2):1–212, 2009.

[15] N. Merhav. Asymptotically optimal decision rules for joint detection and source coding. *Submitted to IEEE Transactions on Information Theory*, October 2013. Available online: http://arxiv.org/abs/1310.4939.pdf.

[16] N. Merhav. List decoding - random coding exponents and expurgated exponents. *Submitted to IEEE Transactions on Information Theory*, November 2013. Available online: http://arxiv.org/pdf/1311.7298.pdf.

[17] G.V. Moustakides. Optimum joint detection and estimation. In *Proc. 2011 IEEE International Symposium on Information Theory*, pages 2984–2988, July 2011.

[18] G.V. Moustakides, G.H. Jajamovich, A. Tajer, and Xiaodong Wang. Joint detection and estimation: Optimum tests and applications. *Information Theory, IEEE Transactions on*, 58(7):4215–4229, July 2012.

[19] J. Scarlett, L. Peng, N. Merhav, A. Martinéz, and A. Guillén i Fàbregas. Expurgated random-coding ensembles: Exponents, refinements and connections. *To appear in IEEE Transactions on Information Theory*. Available online: http://arxiv.org/pdf/1307.6679v3.pdf.

[20] R. A. Scholtz. Frame synchronization techniques. *Communications, IEEE Transactions on*, 28:1204–1213, August 1980.

[21] N. Shulman. *Communication Over an Unknown Channel via Common Broadcasting*. PhD thesis, Tel Aviv University, 2003. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.407.7542&rep=rep1&type=pdf.

[22] A. Somekh-Baruch and N. Merhav. Achievable error exponents for the private fingerprinting game. *Information Theory, IEEE Transactions on*, 53(5):1827–1838, May 2007.

[23] A. Somekh-Baruch and N. Merhav. Exact random coding exponents for erasure decoding. *Information Theory, IEEE Transactions on*, 57(10):6444–6454, 2011.

[24] A. Tchamkerten, V. Chandar, and Gregory W. Wornell. Communication under strong asynchronism. *Information Theory, IEEE Transactions on*, 55(10):4508–4528, Oct 2009.

[25] A. Tchamkerten, V. Chandar, and G.W. Wornell. Asynchronous communication: Capacity bounds and suboptimality of training. *Information Theory, IEEE Transactions on*, 59(3):1227–1255, March 2013.

[26] D. Wang. Distinguishing codes from noise : fundamental limits and applications to sparse communication. Ms.c thesis, Massachusetts Institute of Technology, June 2010. http://dspace.mit.edu/bitstream/handle/1721.1/60710/696796175.pdf?sequence=1.

[27] D. Wang, V. Chandar, S.Y. Chung, and G. W. Wornell. Error exponents in asynchronous communication. In *Proc. 2011 IEEE International Symposium on Information Theory*, pages 1071–1075, 2011.