IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Channel Detection in Coded Communication

Nir Weinberger and Neri Merhav

CCIT Report #887
September 2015

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #887 September 2015

Channel Detection in Coded Communication

Nir Weinberger and Neri Merhav
Dept. of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 3200004, Israel
{nirwein@campus, merhav@ee}.technion.ac.il

Abstract

We consider the problem of block-coded communication, where in each block, the channel law belongs to one of two disjoint sets. The decoder is aimed to decode only messages that have undergone a channel from one of the sets, and thus has to detect the set which contains the prevailing channel. We begin with the simplified case where each of the sets is a singleton. For any given code, 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, and also introduce sub-optimal detection/decoding rules which are simpler to implement. Then, various achievable bounds on the error exponents are derived, including the exact single-letter characterization of the random coding exponents for the optimal detector/decoder. We then extend the random coding analysis to general sets of channels, and show that there exists a universal detector/decoder which performs asymptotically as well as the optimal detector/decoder, when tuned to detect a channel from a specific pair of channels. The case of sets of binary symmetric channels is discussed in detail.

Index Terms

Joint detection/decoding, error exponent, false alarm, misdetection, random coding, expurgation, mismatch detection, detection complexity, universal detection.

I. Introduction

Consider communicating over a channel, for which the prevailing channel law $P_{Y|X}$ (X and Y being the channel input and output, respectively) is supposed to belong to a family of channels $\mathcal{W}$. For example, $\mathcal{W}$ could be a singleton $\mathcal{W}=\{W\}$, or some ball centered at W with respect to (w.r.t.) a given metric (say, total variation). This ball represents some uncertainty regarding the channel, which may result, e.g., from estimation errors. The receiver would also like to examine an alternative hypothesis, in which the channel $P_{Y|X}$ is not in $\mathcal{W}$, and belongs to a different set $\mathcal{V}$, disjoint from $\mathcal{W}$. Such a detection procedure will be useful, for example, in the following cases:

1) *Time-varying channels*: In many protocols, communication begins with a channel estimation phase, and later on, at the data transmission phase, the channel characteristics are tracked using adaptive algorithms [1, Chapters

8 and 9]. However, it is common, that apart from its slow variation, the channel may occasionally also change abruptly, for some reason. Then, the tracking mechanism totally fails, and it is necessary to initialize communication again with a channel estimation phase. The detection of this event is usually performed at high communication layers, e.g., by inspecting the output data bits of the decoder, and verifying their correctness in some way. This procedure could be aided, or even replaced, by identifying a distinct change in the channel as part of the decoding. Note that this problem is a block-wise version of the change-point detection problem from sequential analysis [2], [3] (see, also [4] and referenced therein for a recent related work).

2) *Arbitrarily varying channels in blocks*: In the same spirit, consider a bursty communication system, where within each burst, the underlying channel may belong to either one of two sets, resulting from two very distinctive physical conditions. For example, a wireless communication signal may occasionally be blocked by some large obstacle which results in low channel gain compared to the case of free-space propagation, or it may experience strong interference from other users [5]. The receiver should then decide if the current channel enables reliable decoding.

3) *Secure decoding*: In channels that are vulnerable to intrusions, the receiver would like to verify that an authorized transmitter has sent the message. In these cases, the channel behavior could serve as a proxy for the identity of the transmitter. For example, a channel with a significantly lower or larger signal-to-noise ratio (SNR) than predicted by the geographical distance between the transmitter and receiver, could indicate a possible attempt to intrude the system. The importance of identifying such cases is obvious, e.g., if the messages are used to control a sensitive equipment at the receiver side.

4) *Multiple access channels with no collisions*: Consider a slotted sparse multiple access channel, for which two transmitters are sending messages to a common receiver only in a very small portion of the available slots[1], via *different* channels. Thus, it may be assumed that at each slot, at most one transmitter is active. The receiver would like to identify the sender with high reliability. As might be dictated by practical considerations, the same codebook is used by both transmitters and the receiver identifies the transmitter via a short header, which is common to all codewords of the same transmitter.[2] The receiver usually identifies the transmitter based on the received header only. Of course, this header is an undesired overhead, and so it is important to maximize the detection performance for any given header. To this end, the receiver can also use the codeword sent, and identify the transmitter using the different channel.

Thus, beyond the ordinary task of decoding the message, the receiver would also like to detect the event $P_{Y|X} \in \mathcal{V}$, or, in other words, perform *hypothesis testing* between the null hypothesis $P_{Y|X} \in \mathcal{W}$ and the alternative hypothesis $P_{Y|X} \in \mathcal{V}$. For example, if the channel quality is gauged by a single parameter, say, the crossover probability of a binary symmetric channel (BSC), or the SNR of an additive white Gaussian noise channel (AWGN), then $\mathcal{W}$ and

[1]For simplicity, assume that each codeword occupies exactly a single slot.

[2]Also, if senders simply use different codebooks, then the detection performance would be related to the error probability of the codebook which is comprised from joining the two codebooks. The random coding exponents for the case that the codebook of each transmitter is chosen independently from the codebook of the other user can be obtained by slightly modifying the results of [6].

$\mathcal{V}$ could be two disjoint intervals of this parameter.

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 [7], [8], the problem of joint hypothesis testing and Bayesian estimation was considered, and in [9] 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 a more recent work [10], we have studied the related problem of joint detection and decoding for sparse communication [11], which is motivated by strongly asynchronous channels [12], [13]. In these channels the transmitter is either completely silent or transmits a codeword from a given codebook. The task of the detector/decoder is to decide whether transmission has taken place, and if so, to decode the message. 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 *inclusive error* (IE) - namely, not deciding on the correct message sent, namely, either misdetection of erroneous decoding. We have then found the optimum detector/decoder that minimizes the IE probability subject to given constraints on the FA and the MD probabilities for a given codebook, and also provided single-letter expressions for the exact random coding exponents. While this is a *joint* detector/decoder, we have also observed that an *asymptotic separation principle* holds, in the following sense: A detector/decoder which achieves the optimal exponents may be comprised of an optimal detector in the Neyman-Pearson sense for the FA and MD probabilities, followed by ordinary maximum likelihood (ML) decoding.

In this paper, we study the problem of joint channel detection between two disjoint sets of memoryless channels $\mathcal{W}, \mathcal{V}$, and decoding. We mainly consider discrete alphabets, but some of the results are easily adapted to continuous alphabets. We begin by considering the case of simple hypotheses, namely $\mathcal{W} = \{W\}$ and $\mathcal{V} = \{V\}$. As in [10], we measure the performance of the detector/decoder by its FA, MD and IE probabilities, derive the optimal detector/decoder, and show that here too, an asymptotic separation principle holds. Due to the numerical instability of the optimal detector, we also propose two simplified detectors, each of which suits better a different rate range. Then, we discuss a plethora of lower bounds on the achievable exponents: For the optimal detector/decoder, we derive single-letter expressions for the *exact* random coding exponents, as well as expurgated bounds which improve the bounds at low rates. The exact random coding exponents are also derived for the simplified detectors/decoders. In addition, we also derive Gallager/Forney-style random coding and expurgated bounds, which are simpler to compute, and can be directly adapted to continuous channels. However, as we show in a numerical example, the Gallager/Forney-style exponents may be strictly loose when compared to the exact exponents, even in simple cases. Thus, using the refined analysis technique which is based on type class enumeration (see, e.g., [14], [15] and references therein) and provides the *exact* random coding exponents is beneficial in this case. Afterwards, we discuss a generalization to composite hypotheses, i.e., $\mathcal{W}, \mathcal{V}$ that are not singletons. Finally, we discuss in detail

the archetype example for which W, V are a pair BSCs.

The detection problem addressed in [10] can be seen to be a special case of the problem studied here, for which the the output of the channel V is completely independent of its input, and plays the role of noise. It turns out that the optimal detector/decoder and its properties for the problem studied here are straightforward generalizations of [10], and thus we will discuss them rather briefly and only cite the relevant results from [10]. However, there is a substantial difference in the analysis of the random coding detection exponents in [10], compared to the analysis here. In [10], the discrimination is between the codebook and noise. The detector compares a likelihood which depends on the codebook with a likelihood function that depends on the noise. So, when analyzing the performance of random coding, the random choice of codebook only affects the distribution of the likelihood of the 'codebook hypothesis'. By contrast, here, since we would like to detect the channel, the random choice of codebook affects the likelihood of *both* hypotheses, and consequently, the two hypotheses may be highly dependent. One consequence of this situation, is that to derive the random coding exponents, it is required to analyze the joint distribution of type class enumerators (cf. Subsection V-A), and not just rely on their marginal distributions. The expurgated and Gallager/Forney-style exponents, as well as the simplified detectors/decoders are studied here for the first time.

The outline of the rest of the paper is as follows. In Section II, we establish notation conventions and provide some preliminaries, and in Section III, we formulate the problem of detecting between two channels. In Section IV, we derive the optimum detector/decoder and discuss some of its properties, and also introduce sub-optimal detectors/decoders. In Section V, we present our main results regarding various single-letter achievable exponents. In Section VI, we discuss the problem of detection of composite hypotheses. Finally, in Section VII, we exemplify the results for a pair of BSCs. We defer most of the proofs to the appendices.

II. 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.

A joint distribution of a pair of random variables (X, Y) on $\mathcal{X} \times \mathcal{Y}$, the Cartesian product alphabet of $\mathcal{X}$ and $\mathcal{Y}$, will be denoted by Q_{XY} and similar forms, e.g. $\tilde{Q}_{XY}$. Since usually Q_{XY} will represent a joint distribution of X and Y, we will abbreviate this notation by omitting the subscript XY, and denote, e.g, Q_{XY} by Q. The X-marginal (Y-marginal), induced by Q will be denoted by Q_X (respectively, Q_Y), and the conditional distributions will be denoted by $Q_{Y|X}$ and $Q_{X|Y}$. In accordance with this notation, the joint distribution induced by Q_X and $Q_{Y|X}$ will be denoted by $Q = Q_X \times Q_{Y|X}$.

For a given vector $\mathbf{x}$, let $\hat{Q}_{\mathbf{x}}$ denote the empirical distribution, that is, the vector $\{\hat{Q}_{\mathbf{x}}(x),\ x \in \mathcal{X}\}$, where $\hat{Q}_{\mathbf{x}}(x)$ is the relative frequency of the letter x in the vector $\mathbf{x}$. Let $\mathcal{T}(P_X)$ denote the type class[3] associated with P_X, that is, the set of all sequences $\{\mathbf{x}\}$ for which $\hat{Q}_{\mathbf{x}} = P_X$. Similarly, for a pair of vectors $(\mathbf{x}, \mathbf{y})$, the empirical joint distribution will be denoted by $\hat{Q}_{\mathbf{xy}}$.

The mutual information of a joint distribution Q will be denoted by $I(Q)$, where Q may also be an empirical joint distribution. The information divergence between Q_X and P_X will be denoted by $D(Q_X\|P_X)$, and the conditional information divergence between the empirical conditional distribution $Q_{Y|X}$ and $P_{Y|X}$, averaged over Q_X, will be denoted by $D(Q_{Y|X}\|P_{Y|X}|Q_X)$. Here too, the distributions may be empirical.

The probability of an event $\mathcal{A}$ will be denoted by $\mathbb{P}\{\mathcal{A}\}$, and the expectation operator will be denoted by $\mathbb{E}\{\cdot\}$. Whenever there is room for ambiguity, the underlying probability distribution Q will appear as a subscript, i.e., $\mathbb{P}_Q\{\cdot\}$ and $\mathbb{E}_Q\{\cdot\}$. The indicator function will be denoted by $\mathbb{I}\{\cdot\}$. Sets will normally be denoted by calligraphic letters. The complement of a set $\mathcal{A}$ will be denoted by $\overline{\mathcal{A}}$. Logarithms and exponents will be understood to be taken to the natural base. The notation $[t]_+$ will stand for $\max\{t, 0\}$. We adopt the standard convention that when a minimization (respectively, maximization) problem is performed on an empty set the result is ∞ (respectively, $-\infty$).

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, $\lim_{n\to\infty} \frac{1}{n}\log(\frac{a_n}{b_n}) = 0$, and similar standard notations $\dot{\leq}$ and $\dot{\geq}$ will also be used. When a_n is a sequence of conditional probabilities, i.e, $a_n = \mathbb{P}\left(\mathcal{A}_n|\mathcal{B}_n\right)$ for some pair of sequence of events $\{\mathcal{A}_n\}_{n=1}^{\infty}$ and $\{\mathcal{B}_n\}_{n=1}^{\infty}$, the notation $\mathbb{P}(\mathcal{A}_n|\mathcal{B}_n) \doteq b_n$ will mean

$$\lim_{l\to\infty} \frac{1}{n_l} \log\left(\frac{a_{n_l}}{b_{n_l}}\right) = 0, \tag{1}$$

where $\{n_l\}_{l=1}^{\infty}$ is the sequence of blocklengths such that $\mathbb{P}(\mathcal{B}_{n_l}) > 0$. We shall use the notation $a_n \doteq e^{-n\infty}$ when a_n decays super-exponentially to zero.

Throughout the sequel, we will make a 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,

$$\mathbb{P}\left\{\sum_{i=1}^{k_n} A_n(i) \geq b_n\right\} \doteq \mathbb{P}\left\{\max_{1\leq i\leq k_n} A_n(i) \geq b_n\right\} \tag{2}$$

$$= \mathbb{P}\bigcup_{i=1}^{k_n}\{A_n(i) \geq b_n\} \tag{3}$$

$$\doteq \sum_{i=1}^{k_n} \mathbb{P}\left\{A_n(i) \geq b_n\right\} \tag{4}$$

$$\doteq \max_{1\leq i\leq k_n} \mathbb{P}\left\{A_n(i) \geq b_n\right\}, \tag{5}$$

[3]The blocklength will not be displayed since it will be understood from the context.

provided that $b'_n \doteq b_n$ implies $\mathbb{P}\{A_n(i) \geq b'_n\} \doteq \mathbb{P}\{A_n(i) \geq b_n\}$.[4] In simple words, summations and maximizations are equivalent and can be both "pulled out outside" $\mathbb{P}\{\cdot\}$ without changing the exponential order, as long as $k_n \doteq 1$. The equalities in (5) will be termed henceforth 'the *union rule*' (UR). By the same token,

$$\mathbb{P}\left\{\sum_{i=1}^{k_n} A_n(i) \leq b_n\right\} \doteq \mathbb{P}\left\{\max_{1\leq i\leq k_n} A_n(i) \leq b_n\right\} \tag{6}$$

$$= \mathbb{P}\bigcap_{i=1}^{k_n}\{A_n(i) \leq b_n\}, \tag{7}$$

and these equalities will be termed henceforth 'the *intersection rule*' (IR).

The natural candidate for k_n is the number of joint types possible for a given block length n, and this fact, along with all other rules of the *method of types* [16] will be used extensively henceforth, without explicit reference.

III. Problem Formulation

Consider a discrete memoryless channel (DMC), characterized by a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, and a given matrix of single-letter transition probabilities $\{P_{Y|X}(y|x)\}_{x\in\mathcal{X},y\in\mathcal{Y}}$. Let $\mathcal{C}_n = \{\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_M\} \subset \mathcal{X}^n$, denote a codebook for blocklength n and rate R, for which the transmitted codeword is chosen with a uniform probability distribution over the $M = \lceil e^{nR} \rceil$ codewords. The conditional distribution $P_{Y|X}$ may either satisfy $P_{Y|X} = W$ (the *null hypothesis*), or $P_{Y|X} = V$ (the *alternative hypothesis*). It is required to design a detector/decoder which is oriented to decode messages only arriving via the channel W. Formally, such a detector/decoder ϕ is a partition of $\mathcal{Y}^n$ into $M+1$ regions, denoted by $\{\mathcal{R}_m\}_{m=0}^M$.[5] If $\mathbf{y} \in \mathcal{R}_m$ for some $1 \leq m \leq M$ then the m-th message is decoded. If $\mathbf{y} \in \mathcal{R}_0$ (the *rejection region*) then the channel V is identified, and no decoding takes place.

For a codebook $\mathcal{C}_n$ and a given detector/decoder ϕ, the probability of *false alarm* (FA) is given by

$$P_{\text{FA}}(\mathcal{C}_n, \phi) \triangleq \frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m), \tag{8}$$

the probability of *misdetection* (MD) is given by

$$P_{\text{MD}}(\mathcal{C}_n, \phi) \triangleq \frac{1}{M}\sum_{m=1}^{M} V(\overline{\mathcal{R}_0}|\mathbf{x}_m), \tag{9}$$

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

$$P_{\text{IE}}(\mathcal{C}_n, \phi) \triangleq \frac{1}{M}\sum_{m=1}^{M} W\left(\overline{\mathcal{R}_m}|\mathbf{x}_m\right). \tag{10}$$

Thus, the IE event is the total error event, namely, when the correct codeword $\mathbf{x}_m$ is not decoded either because

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

[5]The decoder ϕ naturally depends on the blocklength via the codebook $\mathcal{C}_n$, but this will be omitted.

of a FA or an ordinary erroneous decoding.[6] The probability of decoding to an erroneous codeword, excluding the rejection region, is termed the *exclusive error* (EE) probability and is defined as

$$P_{\text{EE}}(\mathcal{C}_n, \phi) \triangleq P_{\text{IE}}(\mathcal{C}_n, \phi) - P_{\text{FA}}(\mathcal{C}_n, \phi). \tag{11}$$

When obvious from context, we will omit the notation of the dependence of these probabilities on $\mathcal{C}_n$ and ϕ.

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

$$\begin{aligned} \text{minimize} \quad & P_{\text{IE}} \\ \text{subject to} \quad & P_{\text{FA}} \leq \epsilon_{\text{FA}} \\ & P_{\text{MD}} \leq \epsilon_{\text{MD}} \end{aligned} \tag{12}$$

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 [18, Theorem 11.7.1]. For a given ϵ_{FA}, the minimum achievable P_{MD} 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. While this is true for any *finite* blocklength, as we shall see (Proposition 3), an asymptotic separation principle holds, and the optimal detector in terms of exponents has full tension between the FA and MD exponents. The optimal detector/decoder for the problem (12) will be denoted by ϕ^*.

Remark 1. Naturally, one can use the detector/decoder ϕ^* for messages sent via V. The detection performance for this detector/decoder would simply be obtained by exchanging the meaning of FA with MD.

Our goal is to find the optimum detector/decoder for the problem (12), and then analyze the achievable exponents associated with the resulting error probabilities.

IV. Joint Detectors/Decoders

In this section, we discuss the optimum detector/decoder for the problem (12), and some of its properties. We will also derive an asymptotically optimal version, and discuss simplified decoders, whose performance is close to optimal in some regimes.

A. The Optimum Detector/Decoder

Let $a, b \in \mathbb{R}$, and define the detector/decoder $\phi^* = \{\mathcal{R}_m^*\}_{m=0}^M$, where:

$$\mathcal{R}_0^* \triangleq \left\{ \mathbf{y} : \ a \cdot \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) + \max_m W(\mathbf{y}|\mathbf{x}_m) \leq b \cdot \sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m) \right\}, \tag{13}$$

[6]This definition is conventional in related problems. For example, in Forney's error/erasure setting [17], 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.

and

$$\mathcal{R}_m^* \triangleq \overline{\mathcal{R}_0^*} \bigcap \left\{ \mathbf{y} : \max_m W(\mathbf{y}|\mathbf{x}_m) \geq \max_{k \neq m} W(\mathbf{y}|\mathbf{x}_k) \right\}, \tag{14}$$

where ties are broken arbitrarily.

Lemma 2. *Let a codebook $\mathcal{C}_n$ be given, let ϕ^* be as above, and let ϕ be any other partition of $\mathcal{Y}^n$ into $M+1$ regions. If $P_{\text{FA}}(\mathcal{C}_n, \phi) \leq P_{\text{FA}}(\mathcal{C}_n, \phi^*)$ and $P_{\text{MD}}(\mathcal{C}_n, \phi) \leq P_{\text{MD}}(\mathcal{C}_n, \phi^*)$ then $P_{\text{IE}}(\mathcal{C}_n, \phi) \geq P_{\text{IE}}(\mathcal{C}_n, \phi^*)$.*

Proof: The proof is almost identical to the proof of [10, Lemma 1] and thus omitted. ■

Note that this detector/decoder is optimal (in the Neyman-Pearson sense) for any *given* blocklength n and codebook $\mathcal{C}_n$. Thus, upon a suitable choice of the coefficients a and b, its solves the problem (12) *exactly*. As common, to assess the achievable performance, we resort to large blocklength analysis of error exponents. For a given sequence of codes $\mathcal{C} \triangleq \{\mathcal{C}_n\}_{n=1}^{\infty}$ and a detector/decoder ϕ, the FA exponent is defined as

$$E_{\text{FA}}(\mathcal{C}, \phi) \triangleq \liminf_{n \to \infty} -\frac{1}{n} \log P_{\text{FA}}(\mathcal{C}_n, \phi), \tag{15}$$

and the MD exponent $E_{\text{MD}}(\mathcal{C}, \phi)$ and the IE exponent $E_{\text{IE}}(\mathcal{C}, \phi)$ are defined similarly. The asymptotic version of (12) is then stated as finding the detector/decoder which achieves the largest E_{IE} under constraints on E_{FA} and E_{MD}. To affect these error exponents, the coefficients a, b in (13) need to exponentially increase/decrease as a functions of n. Denoting $a \triangleq e^{n\alpha}$ and $b \triangleq e^{n\beta}$, the rejection region of Lemma 2 becomes

$$\mathcal{R}_0^* = \left\{ \mathbf{y} : \ e^{n\alpha} \cdot \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) + \max_m W(\mathbf{y}|\mathbf{x}_m) \leq e^{n\beta} \cdot \sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m) \right\}. \tag{16}$$

For $\alpha \geq 0$, the ML term on the right-hand side (r.h.s.) of (16) is negligible w.r.t. the left-hand side (l.h.s.), and the obtained rejection region is asymptotically equivalent to

$$\mathcal{R}_0' \triangleq \left\{ \mathbf{y} : \ e^{n\alpha} \cdot \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) \leq e^{n\beta} \cdot \sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m) \right\} \tag{17}$$

which corresponds to an ordinary Neyman-Pearson test between the hypotheses that the channel is W or V. Thus, unlike the fixed blocklength case, asymptotically, we obtain a complete tension between the FA and MD probabilities. Also, comparing (17), and (16), we may observe that the term $\max_m W(\mathbf{y}|\mathbf{x}_m)$ in $\mathcal{R}_0^*$ is added in favor of the alternative hypothesis W. So, in case of a tie in the ordinary Neyman-Pearson test (17), the optimal detector/decoder will actually decide in favor of W.

As the next proposition shows, the above discussion implies that there is no loss in error exponents when using the detector/decoder ϕ', whose rejection region is as in (17), and if $\mathbf{y} \notin \mathcal{R}_0'$ then ordinary ML decoding for W is used, as in (14). This implies an *asymptotic separation principle* between detection and decoding: the optimal detector can be used without considering the subsequent decoding, and the optimal decoder can be used without considering the preceding detection. As a result, asymptotically, there is only a single degree of freedom to control the exponents. Thus, when analyzing error exponents in Section V, we will assume that ϕ' is used, and since (17)

depends on the difference $\alpha-\beta$ only, we will set henceforth $\beta=0$ for ϕ'. The parameter α will be used to control the trade-off between the FA and MD exponents, just as in ordinary hypothesis testing.

Proposition 3. *For any given sequence of codes $\mathcal{C}=\{\mathcal{C}_n\}_{n=1}^{\infty}$, and given constraints on the FA and MD exponents, the detector/decoder ϕ' achieves the same IE exponent as ϕ^*.*

Proof: Assume that the coefficients α,β of ϕ^* (in (16)) are tuned to satisfy constraints on the FA and MD exponents, say $\overline{E}_{\text{FA}}$ and $\overline{E}_{\text{MD}}$. Let us consider replacing ϕ^* by ϕ', with the same α,β. Now, given that the mth codeword was transmitted, the conditional IE probability (10) is the union of the FA event and the event

$$\left\{W(\mathbf{Y}|\mathbf{x}_m)<\max_{k\neq m}W(\mathbf{Y}|\mathbf{x}_k)\right\}, \tag{18}$$

namely, an ordinary ML decoding error. The union bound then implies

$$P_{\text{IE}}(\mathcal{C}_n,\phi)\leq P_{\text{O}}^*(\mathcal{C}_n)+P_{\text{FA}}(\mathcal{C}_n,\phi) \tag{19}$$

where $P_{\text{O}}^*(\mathcal{C}_n)$ is the ordinary decoding error probability, assuming the ML decoder tuned to W. As the union bound is asymptotically exponentially *tight* for a union of two events, then

$$P_{\text{IE}}\left(\mathcal{C}_n,\phi^*\right)\doteq P_{\text{O}}\left(\mathcal{C}_n,\phi^*\right)+P_{\text{FA}}\left(\mathcal{C}_n,\phi^*\right) \tag{20}$$

$$\doteq\max\left\{P_{\text{O}}\left(\mathcal{C}_n,\phi^*\right),P_{\text{FA}}\left(\mathcal{C}_n,\phi^*\right)\right\}, \tag{21}$$

or

$$E_{\text{IE}}\left(\mathcal{C},\phi^*\right)=\min\left\{E_{\text{O}}\left(\mathcal{C},\phi^*\right),E_{\text{FA}}\left(\mathcal{C},\phi^*\right)\right\}. \tag{22}$$

Now, the ordinary decoding error probability is the same for ϕ^* and ϕ' and so the first term in (21) is the same for both detectors/decoders. Also, given any constraint on the MD exponent, the detector defined by $\mathcal{R}_0'$ achieves the maximal FA exponent, and so

$$E_{\text{FA}}\left(\mathcal{C},\phi^*\right)\leq E_{\text{FA}}(\mathcal{C},\phi'). \tag{23}$$

In light of (22), this implies that ϕ' satisfies the MD and FA constraints, and at the same time, achieves an IE exponent at least as large as that of ϕ^*. ■

The achievable exponent bounds will be proved by random coding over some ensemble of codes. Letting over-bar denote an average w.r.t. some ensemble, we will define the random coding exponents, as

$$E_{\text{FA}}\left(\phi\right)\triangleq\lim_{l\to\infty}-\frac{1}{n_l}\log\overline{P_{\text{FA}}}\left(\mathcal{C}_{n_l},\phi\right), \tag{24}$$

where $\{n_l\}_{l=1}^{\infty}$ is a sub-sequence of blocklengths. When we assume a fixed composition ensemble with distribution P_X, this sub-sequence will simply be the blocklengths such that $\mathcal{T}(P_X)$ is not empty, and when we will assume the independent identically distributed (i.i.d.) ensemble, all blocklengths are valid. To comply with definition (15), one

can obtain codes which are good for *all* sufficiently large blocklength by slightly modifying the input distribution. The MD exponent $E_{\text{MD}}(\phi)$ and the IE exponent $E_{\text{IE}}(\phi)$ are defined similarly, where the three exponents share the *same* sequence of blocklengths.

Now, if we provide random coding exponents for the FA, MD and ordinary decoding exponents, then the existence of a good sequence of codes can be easily shown. Indeed, Markov inequality implies that

$$\mathbb{P}\left(\overline{P_{\text{FA}}}(\mathcal{C}_{n_l},\phi) \geq \exp\left[-n_l\left(E_{\text{FA}}(\phi)-\delta\right)\right]\right) \leq e^{-n_l\frac{\delta}{2}}, \tag{25}$$

for all l sufficiently large. Thus, with probability tending to 1, the chosen codebook will have FA probability not larger than $\exp\left[-n\left(E_{\text{FA}}(\phi)-\delta\right)\right]$. As the same can be said on the MD probability and the ordinary error probability, then one can find a sequence of codebooks with simultaneously good FA, MD and ordinary decoding error probabilities, and from (22), also good IE probability. For this reason, henceforth we will only focus on the detection performance, namely the FA and MD exponents. The IE exponent can be simply obtained by (22) and the known bounds of ordinary decoding, namely: (i) the standard Csiszár and Körner random coding bounds [16, Theorem 10.2] (and its tightness [16, Problem 10.34][7]) and the expurgated bound [16, Problem 10.18] for fixed composition ensembles, (ii) the random coding bound [21, Theorem 5.6.2], and the expurgated bound [21, Theorem 5.7.1] for the ensemble of i.i.d. codes.

Beyond the fact that ϕ' is slightly a simpler detector/decoder than ϕ^*, it also enables to prove a very simple relation between its FA and MD exponents. For the next proposition, we will use the notation ϕ'_α and $\mathcal{R}'_{0,\alpha}$ to explicitly denote their dependence on α.

Proposition 4. *For any ensemble of codes such that $E_{\text{FA}}(\mathcal{C},\phi'_\alpha)$ and $E_{\text{MD}}(\mathcal{C},\phi'_\alpha)$ are continuous in α, the FA and MD exponents of ϕ'_α satisfy*

$$E_{\text{FA}}(\mathcal{C},\phi'_\alpha) = E_{\text{MD}}(\mathcal{C},\phi'_\alpha) + \alpha. \tag{26}$$

Proof: For typographical convenience, let us assume that the sub-sequence of blocklengths is simply $\mathbb{N}$. The detector/decoder ϕ'_α is the one which minimizes the FA probability under an MD probability constraint. Considering $e^{-n\alpha} \geq 0$ as a positive Lagrange multiplier, it is readily seen that for any given code, ϕ'_α minimizes the following Lagrangian:

$$L(\mathcal{C}_n,\phi,\alpha) \triangleq P_{\text{FA}}(\mathcal{C}_n,\phi) + e^{-n\alpha}P_{\text{MD}}(\mathcal{C}_n,\phi) \tag{27}$$

$$= \sum_{\mathbf{y}}\left\{\frac{1}{M}\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)\mathbb{I}\{\mathbf{y}\in\mathcal{R}_0\} + e^{-n\alpha}\frac{1}{M}\sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m)\mathbb{I}\left\{\mathbf{y}\in\overline{\mathcal{R}_0}\right\}\right\} \tag{28}$$

Hence,

$$\overline{L(\mathcal{C}_n,\phi,\alpha)} \geq \overline{L(\mathcal{C}_n,\phi'_\alpha,\alpha)} = \overline{P_{\text{FA}}}(\mathcal{C}_n,\phi'_\alpha) + e^{-n\alpha}\overline{P_{\text{MD}}}(\mathcal{C}_n,\phi'_\alpha), \tag{29}$$

[7]See also the extended version [19, Appendix C], which provides a simple proof to the tightness of the random coding exponent of Slepian-Wolf coding [20]. A very similar method can show the tightness of the random coding exponent of channel codes.

or, after taking limits

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{L(\mathcal{C}_n,\phi,\alpha)} = \min\left\{E_{\text{FA}}(\phi), E_{\text{MD}}(\phi)+\alpha\right\}. \tag{30}$$

$$\leq \lim_{n\to\infty} -\frac{1}{n}\log\overline{L(\mathcal{C}_n,\phi'_\alpha,\alpha)} \tag{31}$$

$$= \min\left\{E_{\text{FA}}(\phi'_\alpha), E_{\text{MD}}(\phi'_\alpha)+\alpha\right\}. \tag{32}$$

Now, assume by contradiction that

$$E_{\text{FA}}(\phi'_\alpha) > E_{\text{MD}}(\phi'_\alpha)+\alpha. \tag{33}$$

Then, from continuity of the FA and MD exponents, one can expand $\mathcal{R}'_{0,\alpha}$ to some $\mathcal{R}'_{0,\overline{\alpha}}$ with $\overline{\alpha}<\alpha$ and obtain a decoder $\phi'_{\overline{\alpha}}$ for which

$$E_{\text{MD}}(\phi'_\alpha)+\alpha < E_{\text{MD}}(\mathcal{C},\phi'_{\overline{\alpha}})+\alpha = E_{\text{FA}}(\mathcal{C},\phi'_{\overline{\alpha}}) < E_{\text{FA}}(\mathcal{C},\phi'_\alpha). \tag{34}$$

Thus,

$$\overline{L(\mathcal{C}_n,\phi'_{\overline{\alpha}},\alpha)} \geq \overline{L(\mathcal{C}_n,\phi'_\alpha,\alpha)} \tag{35}$$

which contradicts (33), and so

$$E_{\text{FA}}(\mathcal{C},\phi'_\alpha) \leq E_{\text{MD}}(\mathcal{C},\phi'_\alpha)+\alpha. \tag{36}$$

Similarly, it can be shown that reversed strict inequality in (33) contradicts the optimality of ϕ'_α, and so (26) follows. ■

Remark 5. Consider the following related problem

$$\begin{aligned} \text{minimize} \quad & P_{\text{EE}} \\ \text{subject to} \quad & P_{\text{FA}} \leq \epsilon_{\text{FA}} \\ & P_{\text{MD}} \leq \epsilon_{\text{MD}} \end{aligned} \tag{37}$$

and let ϕ^{**} be the optimal detector/decoder for the problem (37). Now, as $P_{\text{IE}} = P_{\text{EE}} + P_{\text{FA}}$, it may be easily verified that when $P_{\text{FA}} = \epsilon_{\text{FA}}$ for the optimal detector/decoder ϕ^* (of the problem (12)), then ϕ^* is also the optimal detector/decoder for the problem (37). However, when $P_{\text{FA}} < \epsilon_{\text{FA}}$ for ϕ^*, then ϕ^{**} is different, since it easy to check that for the problem (37), the constraint $P_{\text{FA}} \leq \epsilon_{\text{FA}}$ for ϕ^{**} must be achieved with equality. To gain some intuition why (37) is more complicated than (12), see the discussion in [10, Section III].

B. Simplified Detectors/Decoders

Unfortunately, the asymptotically optimal detector/decoder (17) is very difficult to implement in its current form. The reason is that the computation of $\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)$ is usually intractable, as it is the sum of exponentially many likelihood terms, where each likelihood term is exponentially small. This is in sharp contrast to ordinary

decoders, based on comparison of single likelihood terms which can be carried out in the logarithmic scale, rendering them numerically feasible. In a recent related work [22] dealing with the optimal erasure/list decoder [17], it was observed that a much simplified decoder is asymptotically optimal. For the detector/decoder discussed in this paper, this simplification of (17) implies that the rejection region

$$\mathcal{R}_0'' \triangleq \left\{ \mathbf{y} : \ e^{n\alpha} \cdot \max_Q \tilde{N}(Q|\mathbf{y}) e^{nf_W(Q)} \leq e^{n\beta} \cdot \max_Q \tilde{N}(Q|\mathbf{y}) e^{nf_V(Q)} \right\}, \tag{38}$$

is asymptotically optimal, where the *type class enumerators* are defined as

$$\tilde{N}(Q|\mathbf{y}) \triangleq \left| \left\{ \mathbf{x} \in \mathcal{C}_n : \ \hat{Q}_{\mathbf{xy}} = Q_{XY} \right\} \right|. \tag{39}$$

While the above mentioned numerical problem does not arise in $\mathcal{R}_0''$, there is still room for additional simplification which significantly facilitates implementation, at the cost of degrading the performance, perhaps only slightly. For zero rate, the type class enumerators cannot increase exponentially, and so either $\tilde{N}(Q|\mathbf{y}) = 0$ or $\tilde{N}(Q|\mathbf{y}) \doteq 1$. Thus, for low rates, we propose the use of a sub-optimal detector/decoder, which has the following rejection region

$$\mathcal{R}_{0,\text{L}} \triangleq \left\{ \mathbf{y} : \ e^{n\alpha} \cdot \max_{1 \leq m \leq M} W(\mathbf{y}|\mathbf{x}_m) < \max_{1 \leq m \leq M} V(\mathbf{y}|\mathbf{x}_m) \right\}. \tag{40}$$

We will denote the resulting detector/decoder by ϕ_{L}. In this context, this is a *generalized likelihood ratio test* [23], in which the codeword is the 'nuisance parameter' for the detection problem. For high rates (close to the capacity of the channel), the output distribution $\frac{1}{M}\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)$ of a 'good' code [24] tends to be close to a memoryless distribution $\tilde{W} \triangleq (P_X \times W)_Y$ for some distribution P_X. Thus, for high rates, a possible approximation is a sub-optimal detector/decoder, which has the following rejection region

$$\mathcal{R}_{0,\text{H}} \triangleq \left\{ \mathbf{y} : \ e^{n\alpha} \cdot \tilde{W}(\mathbf{y}) < \tilde{V}(\mathbf{y}) \right\}, \tag{41}$$

where $\tilde{V} \triangleq (P_X \times W)_Y$. We will denote the resulting detector/decoder by ϕ_{H}.

As was recently demonstrated in [22], while ϕ_{L} and ϕ_{H} are much simpler to implement than ϕ', they have the potential to cause only slight loss in exponents compared to ϕ'. Since the random coding performance of ϕ_{H} is simply obtained by the standard analysis of hypothesis testing between two memoryless hypotheses (cf. Subsection V-C), we will mainly focus on ϕ_{L}.

V. Achievable Error Exponents

In this section, we derive various achievable exponents for the joint detection/decoding problem (12), for a given pair of DMCs (W, V), at rate R. In Subsection V-A, we derive the *exact* random coding performance of the asymptotically optimal detector/decoder ϕ'. In Subsection V-B, we derive an improved bound for low rates using the expurgation technique. In Subsection V-C, we discuss the exponents achieved by the sub-optimal detectors/decoders ϕ_{L} and ϕ_{H}. In Subsection V-D, we provide Gallager/Forney-style lower bounds on the exponents. While these bounds can be loose and only lead to inferior exponents when compared to Subsections V-A and V-B, it is indeed

useful to derive them since: (i) they are simpler to compute, since they require solving at most two-dimensional optimization problems[8], irrespective of the input/output alphabet sizes, (ii) the bounds are translated almost verbatim to memoryless channels with continuous input/output alphabets, like the AWGN channel. For brevity, in most cases the notation of the dependence on the problem parameters (i.e. R, P_X, α, W, V) will be omitted, and will be reintroduced only when necessary.

A. Exact Random Coding Exponents

We begin with a sequence of definitions. Throughout, $\tilde{Q}$ will represent the joint type of the true transmitted codeword and the output, and $\overline{Q}$ is some type of competing codewords. We denote the *normalized log-likelihood ratio* of a channel W by

$$f_W(Q) \triangleq \sum_{x\in\mathcal{X},y\in\mathcal{Y}} Q(x,y)\log W(y|x), \tag{42}$$

with the convention $f_W(\hat{Q}_{\mathbf{xy}}) = -\infty$ if $W(\mathbf{y}|\mathbf{x}) = 0$. We define the set

$$\mathcal{Q}_W \triangleq \{Q:\ f_W(Q) > -\infty\} \tag{43}$$

and for $\gamma \in \mathbb{R}$,

$$\mathbf{s}(\tilde{Q}_Y, \gamma) \triangleq \min_{Q\in\mathcal{Q}_W:\, Q_Y=\tilde{Q}_Y} I(Q) + \left[-\alpha - f_W(Q) + \gamma\right]_+ . \tag{44}$$

Now, define the sets

$$\mathcal{J}_1 \triangleq \left\{\tilde{Q}:\ f_W(\tilde{Q}) \le -\alpha + f_V(\tilde{Q})\right\}, \tag{45}$$

$$\mathcal{J}_2 \triangleq \left\{\tilde{Q}:\ \mathbf{s}\left(\tilde{Q}_Y, f_V(\tilde{Q})\right) \ge R\right\}, \tag{46}$$

the exponent

$$E_A \triangleq \min_{\tilde{Q}\in\cap_{i=1}^2\mathcal{J}_i} D(\tilde{Q}_{Y|X}\|W|P_X), \tag{47}$$

the sets

$$\mathcal{K}_1 \triangleq \left\{(\tilde{Q},\overline{Q}):\ \overline{Q}_Y = \tilde{Q}_Y\right\}, \tag{48}$$

$$\mathcal{K}_2 \triangleq \left\{(\tilde{Q},\overline{Q}):\ f_W(\overline{Q}) \le -\alpha + f_V(\overline{Q})\right\}, \tag{49}$$

$$\mathcal{K}_3 \triangleq \left\{(\tilde{Q},\overline{Q}):\ f_V(\overline{Q}) \ge \alpha + f_W(\tilde{Q}) - \left[R - I(\overline{Q})\right]_+\right\}, \tag{50}$$

$$\mathcal{K}_4 \triangleq \left\{(\tilde{Q},\overline{Q}):\ \mathbf{s}\left(\tilde{Q}_Y, f_V(\overline{Q}) + \left[R - I(\overline{Q})\right]_+\right) \ge R\right\}, \tag{51}$$

and the exponent

$$E_B \triangleq \min_{(\tilde{Q},\overline{Q})\in\cap_{i=1}^4\mathcal{K}_i} \left\{D(\tilde{Q}_{Y|X}\|W|P_X) + \left[I(\overline{Q}) - R\right]_+\right\}. \tag{52}$$

[8]When there are no input constraints. When input constraints are given, as e.g. in the power limited AWGN channel, it is required to solve four-dimensional optimization problem (cf. (159)).

In addition, let us define the *type-enumeration detection random coding exponent* as

$$E_{\text{TE}}^{\text{RC}}\left(R,\alpha,P_X,W,V\right)\triangleq\min\left\{E_A,E_B\right\}. \tag{53}$$

Theorem 6. *Let a distribution P_X and a parameter $\alpha\in\mathbb{R}$ be given. Then, there exists a sequence of codes $\mathcal{C}=\{\mathcal{C}_n\}_{n=1}^{\infty}$ of rate R such that for any $\delta>0$*

$$E_{\text{FA}}\left(\mathcal{C},\phi^*\right)\geq E_{\text{TE}}^{\text{RC}}\left(R,\alpha,P_X,W,V\right)-\delta, \tag{54}$$

$$E_{\text{MD}}\left(\mathcal{C},\phi^*\right)\geq E_{\text{TE}}^{\text{RC}}\left(R,\alpha,P_X,W,V\right)-\alpha-\delta. \tag{55}$$

The main challenge in analyzing the random coding FA exponent, is that the *likelihoods* of both hypotheses, namely $\sum_{m=1}^{M}W(\mathbf{Y}|\mathbf{X}_m)$ and $\sum_{m=1}^{M}V(\mathbf{Y}|\mathbf{X}_m)$ are very correlated due to the fact the once the codewords are drawn, they are common for both likelihoods. This is significantly different from the situation in [10], in which the likelihood $\sum_{m=1}^{M}W(\mathbf{Y}|\mathbf{X}_m)$ was compared to a likelihood $Q_0(\mathbf{Y})$, of a completely different distribution[9].

We first make the following observation.

Fact 7. *For the detector/decoder ϕ'*

$$P_{\text{FA}}(\mathcal{C}_n,\phi')=\mathbb{P}_W\left(\mathbf{Y}\in\mathcal{R}_0'\right) \tag{56}$$

$$=\mathbb{P}_W\left(\frac{\sum_{m=1}^{M}W(\mathbf{Y}|\mathbf{x}_m)}{\sum_{m=1}^{M}V(\mathbf{Y}|\mathbf{x}_m)}\leq e^{-n\alpha}\right) \tag{57}$$

where $\mathbb{P}_W\left(\mathcal{A}\right)$ is the probability of the event $\mathcal{A}$ under the hypothesis that the channel is W. Similarly,

$$P_{\text{MD}}(\mathcal{C}_n,\phi')=\mathbb{P}_V\left(\mathbf{Y}\notin\mathcal{R}_0'\right) \tag{58}$$

$$=\mathbb{P}_V\left(\frac{\sum_{m=1}^{M}W(\mathbf{Y}|\mathbf{x}_m)}{\sum_{m=1}^{M}V(\mathbf{Y}|\mathbf{x}_m)}\geq e^{-n\alpha}\right) \tag{59}$$

$$=\mathbb{P}_V\left(\frac{\sum_{m=1}^{M}V(\mathbf{Y}|\mathbf{x}_m)}{\sum_{m=1}^{M}W(\mathbf{Y}|\mathbf{x}_m)}\leq e^{n\alpha}\right). \tag{60}$$

Thus, the random coding MD exponent can be obtained by replacing α with $-\alpha$, and W with V in the FA exponent, i.e.

$$\lim_{l\to\infty}-\frac{1}{n_l}\log\overline{P_{\text{MD}}}(\mathcal{C}_{n_l},\phi^*)=E_{\text{TE}}^{\text{RC}}\left(R,-\alpha,P_X,V,W\right) \tag{61}$$

where $\{n_l\}$ is the sub-sequence of blocklengths such that $\mathcal{T}(P_X)$ is not empty.

Before rigorously proving Theorem 6, we make a short detour to present the *type class enumerators* concept [14], and also derive two useful lemmas. Recall that when analyzing the performance of a randomly chosen code, a common method is to first evaluate the error probability conditioned on the transmitted codeword (assumed, without loss of generality, to be $\mathbf{x}_1$) and the output vector $\mathbf{y}$, and average only over $\{\mathbf{X}_m\}_{m=2}^{M}$. Afterwards, the ensemble

[9] In [10], $Q_0(\mathbf{Y})$ represented the hypothesis that no codeword was transmitted and only noise was received.

average error probability is obtained by averaging w.r.t. the random choice of $(\mathbf{X}_1, \mathbf{Y})$. We will assume that the codewords are drawn randomly and uniformly from $\mathcal{T}(P_X)$, and so all joint types Q mentioned henceforth will satisfy $Q_X = P_X$, even if this is not explicitly displayed.

To analyze the conditional error probability, it is useful [14] to define the *type class enumerators*

$$N(Q|\mathbf{y}) \triangleq \left|\left\{\mathbf{x} \in \mathcal{C}_n \backslash \mathbf{x}_1 : \ \hat{Q}_{\mathbf{xy}} = Q\right\}\right|, \tag{62}$$

which, for a given $\mathbf{y}$, count the number of codewords, excluding $\mathbf{x}_1$, which have joint type Q with $\mathbf{y}$. As the codewords in the ensemble are drawn independently, $N(Q|\mathbf{y})$ is a binomial random variable pertaining to $M = \left\lceil e^{nR} \right\rceil$ trials and probability of success of the exponential order of $e^{-nI(Q)}$, and consequently, $\mathbb{E}\left[N(Q|\mathbf{y})\right] \doteq \exp\left[n(R - I(Q))\right]$. A more refined analysis, similar to the one carried in [14, Subsection 6.3], shows that for any given $u \in \mathbb{R}$

$$\mathbb{P}\left\{N(Q|\mathbf{y}) \geq e^{nu}\right\} \doteq \exp\left\{-e^{n[u]_+}\left(n\left[I(Q) - R + [u]_+\right] - 1\right)\right\}. \tag{63}$$

Consequently, if $I(Q) < R$, $N(Q|\mathbf{y})$ concentrates double-exponentially rapidly around its average $\doteq e^{n[R-I(Q)]}$, and if $I(Q) > R$, then with probability tending to 1 we have $N(Q|\mathbf{y}) = 0$, and $\mathbb{P}\left\{N(Q|\mathbf{y}) \geq 1\right\} \doteq e^{-n[I(Q)-R]}$, as well as $\mathbb{P}\left\{N(Q|\mathbf{y}) \geq e^{nu}\right\} \doteq e^{-n\infty}$ for any $u > 0$.

We now derive two useful lemmas. In the first lemma, we show that if a single joint type $\overline{Q}$ is excluded from the possible joint types for a randomly chosen codeword $\mathbf{X}_l$ and $\mathbf{y}$, then the probability of drawing some other joint type is not significantly different from its unconditional counterpart. In the second lemma we characterize the behavior of the probability of the intersection of events in which the type class enumerators are upper bounded.

Lemma 8. *For any $Q \neq \overline{Q}$*

$$\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = Q | \hat{Q}_{\mathbf{X}_l\mathbf{y}} \neq \overline{Q}\right) \doteq \mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = Q\right) \doteq e^{-nI(Q)}. \tag{64}$$

Proof: For any given $\overline{Q}$

$$\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = \overline{Q}\right) \doteq e^{-nI(\overline{Q})}, \tag{65}$$

and if $I(\overline{Q}) = 0$ then

$$\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = \overline{Q}\right) \to 0, \tag{66}$$

as $n \to \infty$, although sub-exponentially [16, Problem 2.2]. Thus, for any $Q \neq \overline{Q}$,

$$\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = Q | \hat{Q}_{\mathbf{X}_l\mathbf{y}} \neq \overline{Q}\right) = \frac{\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = Q, \hat{Q}_{\mathbf{X}_l\mathbf{y}} \neq \overline{Q}\right)}{\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} \neq \overline{Q}\right)} \tag{67}$$

$$= \frac{\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = Q\right)}{1 - \mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = \overline{Q}\right)} \tag{68}$$

$$\doteq e^{-nI(Q)}. \tag{69}$$

■

Lemma 9. *Let a set $\mathcal{Q}$ of joint types, a continuous function $J(Q)$ in $\mathcal{Q}$, and a type $\tilde{Q}_Y$ be given. Let $\{\hat{N}(Q|\mathbf{y})\}_{Q\in\mathcal{Q}}$ be a sequence of sets of binomial random variables pertaining to K_n trials and probability of success p_n. Then, if $K_n \doteq e^{nR}$ and $p_n \doteq e^{-nI(Q)}$*

$$\mathbb{P}\left(\bigcap_{Q\in\mathcal{Q}:\, Q_Y=\tilde{Q}_Y}\left\{\hat{N}(Q|\mathbf{y}) < e^{nJ(Q)}\right\}\right) \begin{cases} = 1-o(n), & \mathbf{S}(\tilde{Q}_Y; J, \mathcal{Q}) > R \\ \doteq e^{-n\infty}, & \textit{otherwise} \end{cases}, \tag{70}$$

where $\mathbf{y} \in \mathcal{T}(\tilde{Q}_Y)$, and

$$\mathbf{S}(\tilde{Q}_Y; J, \mathcal{Q}) \triangleq \min_{Q\in\mathcal{Q}:\, Q_Y=\tilde{Q}_Y} I(Q) + [J(Q)]_+ . \tag{71}$$

Proof: A similar statement was proved in [10, pp. 5086-5087], but for the sake of completeness, we include its short proof. If there exists at least one $Q \in \mathcal{Q}$ with $Q_Y = \tilde{Q}_Y$ for which $I(Q) < R$ and $R - I(Q) > J(Q)$, then this Q alone is responsible for a double exponential decay of the intersection probability, because then the event in question would be a large deviations event whose probability decays exponentially with $M = \lceil e^{nR} \rceil$, thus double-exponentially with n, let alone the intersection over all $Q \in \mathcal{Q}$. The condition for this to happen is $R > \mathbf{S}(\tilde{Q}_Y; J, \mathcal{Q})$. Conversely, if for every $Q \in \mathcal{Q}$ with $Q_Y = \tilde{Q}_Y$, we have $I(Q) > R$ or $R - I(Q) < J(Q)$, i.e., $R < \mathbf{S}(\tilde{Q}_Y; J, \mathcal{Q})$, then the intersection probability is close to 1, since the intersection is over a sub-exponential number of events with very high probability. Thus (70) follows. ■

Remark 10. A natural choice for $\hat{N}(Q|\mathbf{y})$ is simply $N(Q|\mathbf{y})$. However, in what follows, we will need to analyze a conditional version of the type enumerators, namely, events of the form $\{N(Q|\mathbf{y}) = N_1 | N(\overline{Q}|\mathbf{y}) = N_2\}$ for some $0 \le N_1, N_2 \le M$. As Lemma 8 above hints, in some cases the conditional distribution of $N(Q|\mathbf{y})$ is asymptotically the same as the unconditional distribution. In this respect, it should be noted that the result of Lemma 9 is proved using the marginal distribution of each $\hat{N}(Q|\mathbf{y})$ alone, and not their *joint* distribution. It should also be noted that the second argument of $\mathbf{S}(\tilde{Q}_Y; \cdot, \cdot)$ in (71) is a *function* of the joint type Q, and the third argument is a set of joint types. Finally, since the types are dense in the subspace of the simplex of all the type satisfying $Q_Y = \tilde{Q}_Y$, then the exclusion of a *single* type form the intersection in (70) does not change the result of the lemma.

Remark 11. As $Q_X = P_X$ the minimization in (71) is in fact over the variables $\{Q_{Y|X}(y|x)\}_{x\in\mathcal{X}, y\in\mathcal{Y}}$. Thus, whenever $J(Q)$ is convex in $Q_{Y|X}$, then

$$\mathbf{S}(\tilde{Q}_Y; J, \mathcal{Q}) = \min_{Q\in\mathcal{Q}:\, Q_Y=\tilde{Q}_Y} \max_{0\le\lambda\le 1} [I(Q) + \lambda J(Q)] \tag{72}$$

$$\stackrel{(a)}{=} \max_{0\le\lambda\le 1} \min_{Q\in\mathcal{Q}:\, Q_Y=\tilde{Q}_Y} [I(Q) + \lambda J(Q)] \tag{73}$$

where (a) is by the minimax theorem [25], as both $I(Q)$ and $J(Q)$ are convex in $Q_{Y|X}$ and the minimization

set involves only linear constraints and thus convex. This dual form is simpler to compute than (71), since the inner minimization in (73) is a convex optimization problem [26], and the outer maximization problem requires only a simple line-search. Note that the function $\mathbf{s}(\tilde{Q}_Y;\gamma)$ is a specific instance of $\mathbf{S}(\tilde{Q}_Y;\cdot,\cdot)$ defined in (71) with $\mathcal{Q}=\mathcal{Q}_W$ and $J(Q)=-\alpha-f_W(Q)+\gamma$ which is convex in $Q_{Y|X}$ (in fact, linear).

We are now ready to prove Theorem 6.

Proof of Theorem 6: We begin by analyzing the FA exponent. Assume, without loss of generality, that the first message is transmitted. Let us condition on the event $\mathbf{X}_1=\mathbf{x}_1$ and $\mathbf{Y}=\mathbf{y}$, and analyze the average over the ensemble of fixed composition codes of type P_X. For brevity, we will denote $\tilde{Q}=\hat{Q}_{\mathbf{x}_1\mathbf{y}}$. The average conditional FA probability for the decoder ϕ' with parameter α is given by

$$\overline{P_{\text{FA}}}(\mathbf{x}_1,\mathbf{y}) \triangleq \mathbb{P}\left(\mathbf{y}\in\mathcal{R}_0'|\mathbf{X}_1=\mathbf{x}_1,\mathbf{Y}=\mathbf{y}\right) \tag{74}$$

$$\overset{(a)}{=} \mathbb{P}\left(W(\mathbf{y}|\mathbf{x}_1)+\sum_{m=2}^{M}W(\mathbf{y}|\mathbf{X}_m)\le e^{-n\alpha}\cdot V(\mathbf{y}|\mathbf{x}_1)+e^{-n\alpha}\cdot\sum_{m=2}^{M}V(\mathbf{y}|\mathbf{X}_m)\right) \tag{75}$$

$$\begin{aligned}\overset{(UR)}{\doteq}\ & \mathbb{P}\left(W(\mathbf{y}|\mathbf{x}_1)+\sum_{m=2}^{M}W(\mathbf{y}|\mathbf{X}_m)\le e^{-n\alpha}\cdot V(\mathbf{y}|\mathbf{x}_1)\right)\\ &+\mathbb{P}\left(W(\mathbf{y}|\mathbf{x}_1)+\sum_{m=2}^{M}W(\mathbf{y}|\mathbf{X}_m)\le e^{-n\alpha}\cdot\sum_{m=2}^{M}V(\mathbf{y}|\mathbf{X}_m)\right)\end{aligned} \tag{76}$$

$$\begin{aligned}\overset{(IR)}{\doteq}\ & \mathbb{P}\left(\sum_{m=2}^{M}W(\mathbf{y}|\mathbf{X}_m)\le e^{-n\alpha}\cdot V(\mathbf{y}|\mathbf{x}_1)\right)\cdot\mathbb{I}\left\{W(\mathbf{y}|\mathbf{x}_1)\le e^{-n\alpha}\cdot V(\mathbf{y}|\mathbf{x}_1)\right\}\\ &+\mathbb{P}\left(W(\mathbf{y}|\mathbf{x}_1)+\sum_{m=2}^{M}W(\mathbf{y}|\mathbf{X}_m)\le e^{-n\alpha}\cdot\sum_{m=2}^{M}V(\mathbf{y}|\mathbf{X}_m)\right)\end{aligned} \tag{77}$$

$$\begin{aligned}&=\mathbb{P}\left(\sum_{Q}N(Q|\mathbf{y})e^{nf_W(Q)}\le e^{-n\alpha}\cdot e^{nf_V(\tilde{Q})}\right)\cdot\mathbb{I}\left\{f_W(\tilde{Q})\le-\alpha+f_V(\tilde{Q})\right\}\\ &\quad+\mathbb{P}\left(e^{nf_W(\tilde{Q})}+\sum_{Q}N(Q|\mathbf{y})e^{nf_W(Q)}\le e^{-n\alpha}\cdot\sum_{Q}N(Q|\mathbf{y})e^{nf_V(Q)}\right)\end{aligned} \tag{78}$$

$$\triangleq A(\tilde{Q})+B(\tilde{Q}) \tag{79}$$

$$\doteq \max\left\{A(\tilde{Q}),B(\tilde{Q})\right\}, \tag{80}$$

where $A(\tilde{Q})$ and $B(\tilde{Q})$ were implicitly defined, and (a) is because $\{\mathbf{X}_m\}_{m=2}^M$ are chosen independently of $(\mathbf{X}_1,\mathbf{Y})$. For the first term,

$$A(\tilde{Q}) \overset{(IR)}{\doteq} \mathbb{P}\left(\bigcap_{Q:\,f_W(Q)>-\infty}\left\{N(Q|\mathbf{y})<e^{n\left[-\alpha+f_V(\tilde{Q})-f_W(Q)\right]}\right\}\right)\cdot\mathbb{I}\left\{f_W(\tilde{Q})\le-\alpha+f_V(\tilde{Q})\right\} \tag{81}$$

$$\overset{(a)}{\doteq} \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y;-\alpha+f_V(\tilde{Q})-f_W(Q),\mathcal{Q}_W)>R\right\}\cdot\mathbb{I}\left\{f_W(\tilde{Q})\le-\alpha+f_V(\tilde{Q})\right\}, \tag{82}$$

where (a) is by Lemma 9 . Upon averaging over $(\mathbf{X}_1,\mathbf{Y})$, we obtain the exponent E_A of (47), when utilizing the

definition in (44). Moving on to the second term, we first assume that $e^{nf_W(\tilde{Q})} > 0$. Then,

$$B(\tilde{Q}) \overset{(UR)}{\doteq} \sum_{\overline{Q}} \mathbb{P}\left(e^{nf_W(\tilde{Q})} + \sum_{Q} N(Q|\mathbf{y}) e^{nf_W(Q)} \leq e^{-n\alpha} \cdot N(\overline{Q}|\mathbf{y}) e^{nf_V(\overline{Q})} \right) \tag{83}$$

$$\overset{(IR)}{\doteq} \sum_{\overline{Q}} \mathbb{P}\left(\bigcap_{Q \neq \overline{Q}} \left\{ N(Q|\mathbf{y}) e^{nf_W(Q)} \leq e^{-n\alpha} \cdot N(\overline{Q}|\mathbf{y}) e^{nf_V(\overline{Q})} \right\} \cap \right.$$

$$\left. \left\{ N(\overline{Q}|\mathbf{y}) e^{nf_W(\overline{Q})} \leq e^{-n\alpha} \cdot N(\overline{Q}|\mathbf{y}) e^{nf_V(\overline{Q})} \right\} \cap \left\{ e^{nf_W(\tilde{Q})} \leq e^{-n\alpha} \cdot N(\overline{Q}|\mathbf{y}) e^{nf_V(\overline{Q})} \right\} \right) \tag{84}$$

$$\overset{(a)}{=} \sum_{\overline{Q}\colon f_W(\overline{Q}) \leq -\alpha + f_V(\overline{Q})} \mathbb{P}\left(\bigcap_{Q \neq \overline{Q}} \left\{ N(Q|\mathbf{y}) e^{nf_W(Q)} \leq e^{-n\alpha} \cdot N(\overline{Q}|\mathbf{y}) e^{nf_V(\overline{Q})} \right\} \right.$$

$$\left. \left\{ e^{nf_W(\tilde{Q})} \leq e^{-n\alpha} \cdot N(\overline{Q}|\mathbf{y}) e^{nf_V(\overline{Q})} \right\} \right) \tag{85}$$

$$\overset{(b)}{=} \sum_{\overline{Q}\colon f_W(\overline{Q}) \leq -\alpha + f_V(\overline{Q})} \mathbb{P}\left(\bigcap_{Q \neq \overline{Q}\colon f_W(Q) > -\infty} \left\{ N(Q|\mathbf{y}) \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(Q)\right]} \cdot N(\overline{Q}|\mathbf{y}) \right\} \right.$$

$$\left. \left\{ 1 \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q})\right]} \cdot N(\overline{Q}|\mathbf{y}) \right\} \right) \tag{86}$$

$$\triangleq \sum_{\overline{Q}\colon f_W(\overline{Q}) \leq -\alpha + f_V(\overline{Q})} \zeta(\overline{Q}), \tag{87}$$

where (a) is since when $f_W(\overline{Q}) > -\alpha + f_V(\overline{Q})$ the second event in the intersection implies $N(\overline{Q}|\mathbf{y}) = 0$, but this implies that the third event does not occur, and in (b) we have rearranged the terms. To continue the analysis of the exponential behavior of $B(\tilde{Q})$, we split the analysis into three cases:

Case 1: $0 < I(\overline{Q}) \leq R$. For any $0 < \epsilon < R - I(\overline{Q})$ let

$$\mathcal{G}_n \triangleq \left\{ e^{n\left[R - I(\overline{Q}) - \epsilon\right]} \leq N(\overline{Q}|\mathbf{y}) \leq e^{n\left[R - I(\overline{Q}) + \epsilon\right]} \right\}, \tag{88}$$

which satisfies $\mathbb{P}\left[\mathcal{G}_n\right] \doteq 1$. Thus,

$$\zeta(\overline{Q}) = \mathbb{P}\left(\bigcap_{Q \neq \overline{Q}\colon f_W(Q) > -\infty} \left\{ N(Q|\mathbf{y}) \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(Q)\right]} \cdot N(\overline{Q}|\mathbf{y}) \right\} \cap \right.$$

$$\left. \left\{ 1 \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q})\right]} \cdot N(\overline{Q}|\mathbf{y}) \right\} \right) \tag{89}$$

$$\leq \mathbb{P}\left(\bigcap_{Q \neq \overline{Q}\colon f_W(Q) > -\infty} \left\{ N(Q|\mathbf{y}) \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(Q)\right]} \cdot N(\overline{Q}|\mathbf{y}) \right\} \cap \right.$$

$$\left. \left\{ 1 \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q})\right]} \cdot N(\overline{Q}|\mathbf{y}) \right\} \right) \mathbb{P}(\mathcal{G}_n) + \mathbb{P}(\overline{\mathcal{G}_n}) \tag{90}$$

$$\doteq \mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\, f_W(Q)>-\infty}\left\{N(Q|\mathbf{y})\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}\cap\right.$$
$$\left.\left\{1\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}|\mathcal{G}_n\right) \tag{91}$$

$$\dot{\leq}\mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\, f_W(Q)>-\infty}\left\{N(Q|\mathbf{y})\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)+R-I(\overline{Q})+\epsilon\right]}\right\}\cap\right.$$
$$\left.\left\{1\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})+R-I(\overline{Q})+\epsilon\right]}\right\}|\mathcal{G}_n\right) \tag{92}$$

$$\stackrel{(a)}{\doteq} \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y;-\alpha+f_V(\overline{Q})-f_W(Q)+R-I(\overline{Q})+\epsilon,\mathcal{Q}_W)>R\right\}\times$$
$$\mathbb{I}\left\{-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})+R-I(\overline{Q})+\epsilon\geq 0\right\}, \tag{93}$$

where (a) is since conditioned on $\mathcal{G}_n$, $N(Q|\mathbf{y})$ is a binomial random variable with probability of success $\doteq e^{-nI(Q)}$ (see Lemma 8), and more than $e^{nR}-e^{n\left[R-I(\overline{Q})-\epsilon\right]}\doteq e^{nR}$ trials (whenever $Q_Y=\overline{Q}_Y$, and $N(Q|\mathbf{y})=0$ otherwise), and by using Lemma 9 and Remark 10.[10] Similarly,

$$\zeta(\overline{Q})=\mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\, f_W(Q)>-\infty}\left\{N(Q|\mathbf{y})\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}\right.$$
$$\left.\left\{1\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}\right) \tag{94}$$

$$\geq \mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\, f_W(Q)>-\infty}\left\{N(Q|\mathbf{y})\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}\cap\right.$$
$$\left.\left\{1\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}|\mathcal{G}_n\right)\mathbb{P}(\mathcal{G}_n) \tag{95}$$

$$\doteq \mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\, f_W(Q)>-\infty}\left\{N(Q|\mathbf{y})\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}\cap\right.$$
$$\left.\left\{1\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})\right]}\cdot N(\overline{Q}|\mathbf{y})\right\}|\mathcal{G}_n\right) \tag{96}$$

$$\dot{\geq}\mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\, f_W(Q)>-\infty}\left\{N(Q|\mathbf{y})\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)+R-I(\overline{Q})-\epsilon\right]}\right\}\cap\right.$$
$$\left.\left\{1\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})+R-I(\overline{Q})-\epsilon\right]}\right\}|\mathcal{G}_n\right) \tag{97}$$

[10]We have also implicitly used the following obvious monotonicity property: If N_1 and N_2 are two binomial random variables pertaining to the same probability of success but the number of trials of N_1 is larger than the number of trials of N_2 then $\mathbb{P}\left(N_1\leq L\right)\leq\mathbb{P}\left(N_2\leq L\right)$.

$$\stackrel{(a)}{\doteq} \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q) + R - I(\overline{Q}) - \epsilon, \mathcal{Q}_W) > R\right\} \times \mathbb{I}\left\{-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q}) + R - I(\overline{Q}) - \epsilon \geq 0\right\}, \tag{98}$$

where (a) is now since conditioned on $\mathcal{G}_n$, $N(Q|\mathbf{y})$ is a binomial random variable, with probability of success $\doteq e^{-nI(Q)}$ (see Lemma 8), and less than e^{nR} trials (whenever $Q_Y = \overline{Q}_Y$, and $N(Q|\mathbf{y}) = 0$ otherwise), and by utilizing again Lemma 9 and Remark 10. As $\epsilon > 0$ is arbitrary,

$$\zeta(\overline{Q}) \doteq \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q) + R - I(\overline{Q}), \mathcal{Q}_W) > R\right\} \times \mathbb{I}\left\{-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q}) + R - I(\overline{Q}) > 0\right\} \tag{99}$$

Case 2: Assume that $I(\overline{Q}) = 0$. This case is not significantly different from Case 1. Indeed, for any $0 < \epsilon < R$, let

$$\mathcal{G}_n \triangleq \left\{e^{n(R-\epsilon)} \leq N(\overline{Q}|\mathbf{y}) \leq \frac{1}{2}e^{nR}\right\}, \tag{100}$$

then $\mathbb{P}[\mathcal{G}_n] \doteq 1$. To see this, we note that for $\mathbf{X}_l$ drawn uniformly within $\mathcal{T}(P_X)$.

$$\mathbb{E}\left[N(\overline{Q}|\mathbf{y})\right] = e^{nR} \cdot \mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = \overline{Q}\right) \tag{101}$$

$$\stackrel{(a)}{\leq} \frac{1}{4}e^{nR} \tag{102}$$

for all n sufficiently large, where (a) is since $\mathbb{P}\left(\hat{Q}_{\mathbf{X}_l\mathbf{y}} = Q\right) \to 0$ as $n \to \infty$. So, by Markov inequality

$$\mathbb{P}\left\{N(\overline{Q}|\mathbf{y}) \leq \frac{1}{2}e^{nR}\right\} \geq \mathbb{P}\left\{N(\overline{Q}|\mathbf{y}) \leq 2\mathbb{E}\left[N(\overline{Q}|\mathbf{y})\right]\right\} \geq \frac{1}{2}. \tag{103}$$

Since, as before $\mathbb{P}\left\{e^{n(R-\epsilon)} \leq N(\overline{Q}|\mathbf{y})\right\} \doteq 1$, and the intersection of two high probability sets also has high probability, we obtain $\mathbb{P}[\mathcal{G}_n] \doteq 1$. The rest of the analysis follows as in Case 1, and the result is the same, when setting $I(\overline{Q}) = 0$.

Case 3: Assume that $I(\overline{Q}) > R$. Then, for any $\epsilon > 0$

$$\zeta(\overline{Q}) = \mathbb{P}\left(\bigcap_{Q \neq \overline{Q}:\, f_W(Q) > -\infty} \left\{N(Q|\mathbf{y}) \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(Q)\right]} \cdot N(\overline{Q}|\mathbf{y})\right\} \cap \left\{1 \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q})\right]} \cdot N(\overline{Q}|\mathbf{y})\right\}\right) \tag{104}$$

$$\stackrel{(a)}{\doteq} \mathbb{P}\left(\bigcap_{Q \neq \overline{Q}:\, f_W(Q) > -\infty} \left\{N(Q|\mathbf{y}) \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(Q)\right]} \cdot N(\overline{Q}|\mathbf{y})\right\} \cap \left\{1 \leq e^{n\left[-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q})\right]} \cdot N(\overline{Q}|\mathbf{y})\right\} | 1 \leq N(\overline{Q}|\mathbf{y}) \leq e^{n\epsilon}\right) \mathbb{P}\left(1 \leq N(\overline{Q}|\mathbf{y}) \leq e^{n\epsilon}\right) \tag{105}$$

$$\overset{(b)}{\geq} \mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\, f_W(Q)>-\infty} \left\{N(Q|\mathbf{y}) \leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)\right]}\right\} \cap \right.$$

$$\left. \left\{1 \leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})\right]}\right\} | 1 \leq N(\overline{Q}|\mathbf{y}) \leq e^{n\epsilon}\right) e^{-n\left(I(\overline{Q})-R\right)} \tag{106}$$

$$\overset{(c)}{\doteq} \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q), \mathcal{Q}_W) > R\right\} \mathbb{I}\left\{-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q}) \geq 0\right\} e^{-n\left(I(\overline{Q})-R\right)}, \tag{107}$$

where (a) is since conditioned on $N(\overline{Q}|\mathbf{y}) = 0$ the probability of the event is 0, and

$$\mathbb{P}\left[N(\overline{Q}|\mathbf{y}) \geq e^{n\epsilon}\right] \doteq 0, \tag{108}$$

(b) is since

$$\mathbb{P}\left(1 \leq N(\overline{Q}|\mathbf{y}) \leq e^{n\epsilon}\right) \geq \mathbb{P}\left(N(\overline{Q}|\mathbf{y}) = 1\right) \tag{109}$$

$$\doteq e^{-n\left(I(\overline{Q})-R\right)}, \tag{110}$$

and (c) is since conditioned on $1 \leq N(\overline{Q}|\mathbf{y}) \leq e^{n\epsilon}$, $N(Q|\mathbf{y})$ is a binomial random variable, with probability of success $\doteq e^{-nI(Q)}$ (see Lemma 8), and $\doteq e^{nR}$ trials (whenever $Q_Y = \overline{Q}_Y$, and $N(Q|\mathbf{y}) = 0$ otherwise), and by utilizing once again Lemma 9 and Remark 10. Similarly, using

$$\mathbb{P}\left(1 \leq N(\overline{Q}|\mathbf{y}) \leq e^{n\epsilon}\right) \leq e^{n\epsilon}\mathbb{P}\left(N(\overline{Q}|\mathbf{y}) = 1\right) \doteq e^{-n\left(I(\overline{Q})-R-\epsilon\right)}, \tag{111}$$

the same analysis as in the previous case, shows a reversed inequality. As $\epsilon > 0$ is arbitrary, then

$$\zeta(\overline{Q}) \doteq \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q), \mathcal{Q}_W) > R\right\} \mathbb{I}\left\{-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q}) > 0\right\} e^{-n\left(I(\overline{Q})-R\right)}. \tag{112}$$

Returning to (87), we obtain that $B(\tilde{Q})$ is exponentially equal to the maximum between

$$\max_{\overline{Q}:\, f_W(\overline{Q})<-\alpha+f_V(\overline{Q}),\, I(\overline{Q})\leq R,\, f_V(\overline{Q})>\alpha+f_W(\tilde{Q})-R+I(\overline{Q})} \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q) + R - I(\overline{Q}), \mathcal{Q}_W) > R\right\}, \tag{113}$$

and

$$\max_{\overline{Q}:\, f_W(\overline{Q})<-\alpha+f_V(\overline{Q}),\, I(\overline{Q})>R,\, f_V(\overline{Q})>\alpha+f_W(\tilde{Q})} \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q), \mathcal{Q}_W) > R\right\} e^{-n\left(I(\overline{Q})-R\right)}, \tag{114}$$

or, more succinctly,

$$B(\tilde{Q}) = \max_{\overline{Q}} \mathbb{I}\left\{\mathbf{S}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q) + \left[R - I(\overline{Q})\right]_+, \mathcal{Q}_W) > R\right\} e^{-n\left[I(\overline{Q})-R\right]_+} \tag{115}$$

where the maximization is over

$$\left\{\overline{Q}:\ f_W(\overline{Q}) < -\alpha + f_V(\overline{Q}),\ f_V(\overline{Q}) > \alpha + f_W(\tilde{Q}) - \left[R - I(\overline{Q})\right]_+\right\}. \tag{116}$$

Now, in the evaluation of $B(\tilde{Q})$ we have assumed that $e^{nf_W(\tilde{Q})} > 0$. However, there is no need to analyze the case $e^{nf_W(\tilde{Q})} = 0$ since as

$$f_W(\tilde{Q}) = -D(\tilde{Q}_{Y|X}||W|P_X) - H_{\tilde{Q}}(Y|X) \tag{117}$$

and $H_{\tilde{Q}}(Y|X) \leq \log|\mathcal{Y}| < \infty$, then $e^{nf_W(\tilde{Q})} = 0$ implies $\mathbb{P}(\hat{Q}_{\mathbf{x}_1\mathbf{y}} = \tilde{Q}) \doteq \exp\left[-nD(\tilde{Q}_{Y|X}||W|P_X)\right] \doteq e^{-n\infty}$. Thus, upon averaging over $(\mathbf{X}_1, \mathbf{Y})$ we obtain the exponent E_B of (52), utilizing (44). Then, we obtain the required result from (80).

Next, for the MD exponent, we observe that as $E_{\text{TE}}^{\text{RC}}(R, \alpha, P_X, W, V)$ is continuous in α, Fact 7 above implies that the MD exponent will be also continuous in α. So, Proposition 4 implies that when the codewords are drawn from a fixed composition ensemble with distribution P_X,

$$\lim_{l\to\infty} -\frac{1}{n_l} \log \overline{P_{\text{MD}}}(\mathcal{C}_{n_l}, \phi^*) = E_{\text{TE}}^{\text{RC}}(R, \alpha, P_X, W, V) - \alpha. \tag{118}$$

Finally, the continuity of $E_{\text{TE}}^{\text{RC}}(R, \alpha, P_X, W, V)$ in P_X implies that for all sufficiently large n, one can find a distribution P'_X close enough to P_X such that (54) and (55) hold, which completes the proof of the theorem. ■

To keep the flow of the proof, we have omitted a technical point which we now address.

Remark 12. The ensemble average FA probability should be obtained by averaging $\overline{P_{\text{FA}}}(\mathbf{X}_1, \mathbf{Y})$ w.r.t. $(\mathbf{X}_1, \mathbf{Y})$. However, we have averaged its asymptotic equivalence in the exponential scale, resulting from analyzing the terms $A(\tilde{Q})$ and $B(\tilde{Q})$. Thus, in a sense, we have interchanged the expectation and limit order. This is possible due to the fact that all the asymptotic equivalence relations become tight for n sufficiently large, which *does not depend on* $\tilde{Q}$ (i.e. on $(\mathbf{X}_1, \mathbf{Y})$). Indeed, the union and intersection rules add a negligible term to the exponent. This term depends only on the number of types, which is polynomial in n, independent of the specific type $\tilde{Q}$. The asymptotic equivalence relations that stem from Lemma 9 do not depend on $\tilde{Q}$, as functions of $\tilde{Q}$ only play the role of bounds on the sums of weighted type enumerators. Indeed, it is evident from the proof of Lemma 9 that the required blocklength n to approach convergence of the probability does not depend on $J(Q)$.

B. Expurgated Exponents

We begin again with several definitions. Throughout, $P_{X\tilde{X}}$ will represent a joint type of a pair of codewords. Let us define the Chernoff distance[11]

$$d_s(x, \tilde{x}) \triangleq -\log\left(\sum_{y\in\mathcal{Y}} W^{1-s}(y|x)V^s(y|\tilde{x})\right) \tag{119}$$

and the set

$$\mathcal{L} \triangleq \left\{P_{X\tilde{X}} :\ P_{\tilde{X}} = P_X,\ I(P_{X\tilde{X}}) \leq R\right\}. \tag{120}$$

[11]When s is maximized, then the result is the Chernoff information [18, Section 11.9]. For $s = \frac{1}{2}$ this is the Bhattacharyya distance.

In addition, let us define the *type-enumeration detection expurgated exponent* as

$$E_{\text{TE}}^{\text{EX}}(R,\alpha,P_X,W,V) \triangleq \max_{0\le s\le 1}\min_{P_{X\tilde{X}}\in\mathcal{L}}\left\{\alpha s+\mathbb{E}\left[d_s(X,\tilde{X})\right]+I(P_{X\tilde{X}})-R\right\}. \tag{121}$$

Theorem 13. *Let a distribution P_X and a parameter $\alpha\in\mathbb{R}$ be given. Then, there exists a sequence of codes $\mathcal{C}=\{\mathcal{C}_n\}_{n=1}^{\infty}$ of rate R such that for any $\delta>0$*

$$E_{\text{FA}}(\mathcal{C},\phi^*)\ge E_{\text{TE}}^{\text{EX}}(R,\alpha,P_X,W,V)-\delta, \tag{122}$$

$$E_{\text{MD}}(\mathcal{C},\phi^*)\ge E_{\text{TE}}^{\text{EX}}(R,\alpha,P_X,W,V)-\alpha-\delta. \tag{123}$$

The proof can be found in Appendix A.

Remark 14. Hölder inequality shows that $d_s(x,\tilde{x})\ge 0$. In (121), there is freedom to maximize over $0\le s\le 1$, and naturally, $s=\frac{1}{2}$ is a valid choice. Due to the symmetry of $d_s(x,\tilde{x})$ in s around $s=\frac{1}{2}$ when $W=V$, for the ordinary decoding exponent, the optimal choice is $s=\frac{1}{2}$ (as also manifested at $R=0$ by the Shannon-Gallager-Berlekamp upper bound [27, Theorem 4]), but here, no such symmetry exists.

Remark 15. In Theorem 13 we have assumed a fixed composition code of type P_X. As discussed in [16, Problem 10.23 (b)], for ordinary decoding, the exponent (121) is at least as large as the corresponding exponent using Gallager's approach to expurgation [21, Section 5.7], and for the maximizing P_X, the two bounds coincide. Thus, for ordinary decoding, the exponent bound (121) offers an improvement over Gallager's approach when the input type P_X is constrained. For joint detection/decoding, there is an additional source of possible improvement - the input type P_X which best suits channel coding is not necessarily the best input type for the detection problem. We also mention that for $R=0$, an improvement at any given P_X can be obtained by taking the *upper concave envelope* of (121) (see [16, Problem 10.22] and the discussion in [28, Section II]).

Remark 16. This expurgation technique can be used also for continuous alphabet channels, and specifically, for AWGN channels, see [29, Section 4].

C. Exact Random Coding Exponents of Simplified Detectors/Decoders

We now discuss the random coding exponents achieved by the simplified detectors/decoders ϕ_{L} and ϕ_{H} introduced in Subsection IV-B. We begin with ϕ_{L}. For $\gamma\in\mathbb{R}$, let us define

$$\mathbf{t}(\tilde{Q}_Y,\gamma)\triangleq\min_{Q\in\mathcal{Q}_W:\,Q=\tilde{Q}_Y,\,-\alpha-f_W(Q)+\gamma\le 0} I(Q), \tag{124}$$

the sets $\mathcal{J}_{1,\text{L}}\triangleq\mathcal{J}_1$ and

$$\mathcal{J}_{2,\text{L}}\triangleq\left\{\tilde{Q}:\ \mathbf{t}\left(\tilde{Q}_Y,f_V(\tilde{Q})\right)\ge R\right\}, \tag{125}$$

the exponent

$$E_{A,\text{L}}\triangleq\min_{\tilde{Q}\cap_{i=1}^{2}\mathcal{J}_{i,\text{L}}} D(\tilde{Q}_{Y|X}\|W|P_X), \tag{126}$$

the sets $\mathcal{K}_{1,\mathrm{L}} \triangleq \mathcal{K}_1$, $\mathcal{K}_{2,\mathrm{L}} \triangleq \mathcal{K}_2$[12]

$$\mathcal{K}_{3,\mathrm{L}} \triangleq \left\{(\tilde{Q}, \overline{Q}) : \ f_V(\overline{Q}) \geq \alpha + f_W(\tilde{Q})\right\}, \tag{127}$$

$$\mathcal{K}_{4,\mathrm{L}} \triangleq \left\{(\tilde{Q}, \overline{Q}) : \ \mathbf{t}\left(\tilde{Q}_Y, f_V(\overline{Q})\right) \geq R\right\}, \tag{128}$$

and the exponent

$$E_{B,\mathrm{L}} \triangleq \min_{(\tilde{Q},\overline{Q}) \in \cap_{i=1}^{4} \mathcal{K}_{i,\mathrm{L}}} D(\tilde{Q}_{Y|X} \| W | P_X) + \left[I(\overline{Q}) - R\right]_+ . \tag{129}$$

In addition, let us define the *low-rate detection random coding exponent* as

$$E_{\mathrm{L}}^{\mathrm{RC}}(R, \alpha, P_X, W, V) \triangleq \min\left\{E_{A,\mathrm{L}}, E_{B,\mathrm{L}}\right\}. \tag{130}$$

Theorem 17. *Let a distribution P_X and a parameter $\alpha \geq 0$ be given. Then, there exists a sequence of codes $\mathcal{C} = \{\mathcal{C}_n\}_{n=1}^{\infty}$ of rate R such that for any $\delta > 0$*

$$E_{\mathrm{FA}}(\mathcal{C}, \phi^*) \geq E_{\mathrm{L}}^{\mathrm{RC}}(R, \alpha, P_X, W, V) - \delta, \tag{131}$$

$$E_{\mathrm{MD}}(\mathcal{C}, \phi^*) \geq E_{\mathrm{L}}^{\mathrm{RC}}(R, -\alpha, P_X, V, W) - \delta. \tag{132}$$

The proof can be found in Appendix B.

Next, we discuss the random coding exponents of ϕ_{H}. As this is a simple hypothesis testing between two memoryless sources $\tilde{W}$ and $\tilde{V}$, the standard analysis [30] and [18, Section 11.7] is applicable verbatim. For given $0 \leq \mu \leq 1$, let

$$Q_\mu(y) \triangleq \frac{\tilde{W}^\mu(y)\tilde{V}^{1-\mu}(y)}{\sum_{y' \in \mathcal{Y}} \tilde{W}^\mu(y')\tilde{V}^{1-\mu}(y')} \tag{133}$$

for all $x \in \mathcal{X}$, and let us define the high-rate detection random coding exponent as

$$E_{\mathrm{H}}^{\mathrm{RC}}(R, \alpha, P_X, W, V) \triangleq D(Q_{\mu(\alpha)} || \tilde{W}), \tag{134}$$

where $\mu(\alpha)$ is chosen so that

$$D(Q_{\mu(\alpha)} || \tilde{W}) - D(Q_{\mu(\alpha)} || \tilde{V}) = -\alpha. \tag{135}$$

Theorem 18. *Let a distribution P_X and a parameter $\alpha \geq 0$ be given. Then, there exists a sequence of codes $\mathcal{C} = \{\mathcal{C}_n\}_{n=1}^{\infty}$ of rate R such that for any $\delta > 0$*

$$E_{\mathrm{FA}}(\mathcal{C}, \phi^*) \geq E_{\mathrm{H}}^{\mathrm{RC}}(R, \alpha, P_X, W, V) - \delta, \tag{136}$$

$$E_{\mathrm{MD}}(\mathcal{C}, \phi^*) \geq E_{\mathrm{H}}^{\mathrm{RC}}(R, \alpha, P_X, W, V) - \alpha - \delta. \tag{137}$$

[12]It can be noticed that the only difference between $\mathcal{K}_{3,\mathrm{L}}, \mathcal{K}_{4,\mathrm{L}}$ and $\mathcal{K}_3, \mathcal{K}_4$ are the exclusion of $I(Q) - R$ terms and replacing $\mathbf{s}(\tilde{Q}_Y, \gamma)$ with $\mathbf{t}(\tilde{Q}_Y, \gamma)$.

Proof: The proof follows the standard analysis in [18, Section 11.7]. ■

Remark 19. The decoder ϕ_{H} and its random coding exponents do not depend on the rate R.

D. Gallager/Forney-Style Exponents

Next, we derive achievable exponents using the classical Gallager/Forney technique.

1) Random Coding Exponents: For a given distribution $\{P_X(x)\}_{x\in\mathcal{X}}$, and parameters s, ρ, define

$$E_0'(s,\rho) \triangleq -\log\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)W^{(1-s)/\rho}(y|x)V^{s/\rho}(y|x)\right)^{\rho}\right], \tag{138}$$

and

$$E_0''(s,\rho) \triangleq -\log\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)W^{(1-s)/\rho}(y|x)\right)^{\rho}\left(\sum_{x\in\mathcal{X}} P_X(x)V^{s/\rho}(y|x)\right)^{\rho}\right], \tag{139}$$

and let the *Gallager/Forney detection random coding exponent* be defined as

$$E_{\text{GF}}^{\text{RC}}(R,\alpha,P_X,W,V) \triangleq \max_{0\le s\le 1,\max\{s,1-s\}\le\rho\le 1} \min\big\{\alpha s + E_0'(s,\rho) - (\rho-1)R, \alpha s + E_0''(s,\rho) - (2\rho-1)R\big\}. \tag{140}$$

Theorem 20. *Let a distribution P_X and a parameter $\alpha\in\mathbb{R}$ be given. Then, there exists a sequence of codes $\mathcal{C}=\{\mathcal{C}_n\}_{n=1}^{\infty}$ of rate R such that for any $\delta>0$*

$$E_{\text{FA}}(\mathcal{C},\phi^*) \ge E_{\text{GF}}^{\text{RC}}(R,\alpha,P_X,W,V) - \delta, \tag{141}$$

$$E_{\text{MD}}(\mathcal{C},\phi^*) \ge E_{\text{GF}}^{\text{RC}}(R,\alpha,P_X,W,V) - \alpha - \delta. \tag{142}$$

The proof can be found in Appendix C.

2) Expurgated Exponents: For a given distribution $\{P_X(x)\}_{x\in\mathcal{X}}$ and parameters s, ρ, define

$$E_x'(s) \triangleq -\log\left[\sum_{x\in\mathcal{X}} P_X(x)\sum_{y\in\mathcal{Y}} W^{1-s}(y|x)V^{s}(y|x)\right], \tag{143}$$

and

$$E_x''(s) \triangleq -\log\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)W^{1-s}(y|x)\right)\left(\sum_{x\in\mathcal{X}} P_X(x)V^{s}(y|x)\right)\right], \tag{144}$$

and let the *Gallager/Forney detection expurgated exponent* be defined as

$$E_{\text{GF}}^{\text{EX}}(R,\alpha,P_X,W,V) \triangleq \sup_{0\le s\le 1,\rho\ge 1} \min\big\{s\alpha + E_x'(s), s\alpha + E_x''(s) - \rho R\big\}. \tag{145}$$

Theorem 21. *Let a distribution P_X and a parameter $\alpha\in\mathbb{R}$ be given. Then, there exists a sequence of codes*

$\mathcal{C} = \{\mathcal{C}_n\}_{n=1}^{\infty}$ *of rate* R *such that for any* $\delta > 0$

$$E_{\text{FA}}\left(\mathcal{C}, \phi^*\right) \geq E_{\text{GF}}^{\text{EX}}\left(R, \alpha, P_X, W, V\right) - \delta, \tag{146}$$

$$E_{\text{MD}}\left(\mathcal{C}, \phi^*\right) \geq E_{\text{GF}}^{\text{EX}}\left(R, \alpha, P_X, W, V\right) - \alpha - \delta. \tag{147}$$

The proof can be found in Appendix D.

E. Discussion

We summarize this section with the following discussion.

1) *Monotonicity in the rate*: The ordinary random coding exponents are decreasing with the rate R, and vanish at $I(P_X \times W)$. By contrast, the detection exponents are not necessarily so. Indeed, the exponent E_A of (47) is increasing with the rate. For the exponent E_B of (52), as R increases, the objective function decreases and $\mathcal{K}_3$ expands, but the set $\mathcal{K}_4$ diminishes[13], and so no monotonicity is assured for E_B, and as a results, also for $E_{\text{TE}}^{\text{RC}}\left(R, \alpha, P_X, W, V\right)$. The same holds for ϕ_{L}, whereas ϕ_{H} does not depend on R at all. The expurgated exponent $E_{\text{TE}}^{\text{EX}}\left(R, \alpha, P_X, W, V\right)$ of (121) decreases in R. To gain intuition, recall from (63), that when $I(Q) < R$ the type enumerator $N(Q|\mathbf{y})$ concentrates double-exponentially rapidly around its average $\doteq \exp\left[n(R - I(Q))\right]$. Thus, for any given $\mathbf{y}$, an increase of the rate will introduce codewords having a joint type that was not typically seen at lower rates, and this new joint type might dominate one of the likelihoods. However, it is not clear to which direction this new type will tip the scale in the likelihoods comparison, and so the rate increase does not necessarily imply an increase or a decrease of one of the exponents. In addition, the above discussion and (21) imply that the largest achievable rate such that $P_{\text{IE}} \to 0$ as $n \to \infty$, may still be the mutual information $I(P_X \times W)$, or, in other words, the detection does not cause a rate loss.

2) *Computation of the exponents*: Unfortunately, the optimization problems involved in computing the exact exponents of Subsections V-A and V-C are usually not convex, and might be complex to solve when the alphabets are large. For example, for the exact exponents, computing E_A of (47) is not a convex optimization problem since $\mathcal{J}_2$ is not a convex set of $\tilde{Q}$, and computing E_B of (52) is not a convex optimization problem since $\mathcal{K}_3$ and $\mathcal{K}_4$ are not convex sets of $(\tilde{Q}, \overline{Q})$, and not even of $(\tilde{Q}_{Y|X}, \overline{Q}_{Y|X})$. An efficient algorithm their efficient computation is an important open problem. However, the expurgated exponent (121) is concave[14] in s and convex in $P_{X\tilde{X}}$. This promotes the importance of the lower bounds derived in Subsection V-D, which only require two-dimensional optimization problems, irrespective of the alphabet sizes.

3) *Choice of input distribution*: Thus far, the input distribution P_X was assumed fixed, but it can obviously be optimized. Nonetheless, there might be a tension between the optimal choice for channel coding versus the optimal choice for detection. For example, consider the detection problem between W, a Z-channel, i.e. $W(0|0) =$

[13] As its r.h.s. always increases, but its l.h.s. does not.

[14] The second derivative w.r.t. s of $d_s(x, \tilde{x})$ is the variance of $\log \frac{V(y|\tilde{x})}{W(y|x)}$ w.r.t. the distribution P_Y which satisfies $P_Y(y) \propto W^{1-s}(y|x)V^s(y|\tilde{x})$.

$1, W(0|1) = w$ for some $0 \leq w \leq 1$, and V, an S-channel, i.e. $V(1|0) = v, V(1|1) = 1$ for some $0 \leq v \leq 1$. Choosing $P_X(0) = 1$ will result an infinite FA and MD exponents (upon appropriate choice of α), but is useless from the channel coding perspective. One possible remedy is to define a Lagrangian that weighs, e.g. the FA and ordinary decoding exponents with some weight, and optimize it over the input type. However, still, the resulting optimization might be non-tractable.

4) *Simplified decoders*: Intuitively, the low-rate simplified detector/decoder ϕ_{L} has worse FA-MD trade-off than the optimal detector/decoder ϕ' since the effect of a non-typical codeword may be averaged out in $\frac{1}{M}\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)$, but may totally change $\max_{1\leq m\leq M} W(\mathbf{y}|\mathbf{x}_m)$. However, there exists a *critical rate* R_{cr} such that for all $R \leq R_{\text{cr}}$ the exponents of the two detectors/decoders coincide, when using the same parameter α. To see this, first let

$$\tilde{Q}_A \triangleq \arg\min_{\tilde{Q}\in\mathcal{J}_1} D(\tilde{Q}_{Y|X}\|W|P_X), \tag{148}$$

i.e. the exponent E_A for $R = 0$, and in fact, for all rates satisfying

$$R \leq \mathbf{s}\left(\tilde{Q}_Y; f_V(\tilde{Q}_A)\right) \triangleq R_{\text{cr,A}}. \tag{149}$$

Since from Remark 28 (Appendix B)

$$\mathbf{s}(\tilde{Q}_Y, \gamma) \leq \mathbf{t}(\tilde{Q}_Y, \gamma) \tag{150}$$

this is also the exponent $E_{A,\text{L}}$. Now, letting $R = 0$ in $\{\mathcal{K}_i\}_{i=3}^4$ and then solving

$$(\tilde{Q}_B, \overline{Q}_B) \triangleq \arg\min_{(\tilde{Q},\overline{Q})\in\cap_{i=1}^4\mathcal{K}_i} \left\{D(\tilde{Q}_{Y|X}\|W|P_X) + I(\overline{Q})\right\} \tag{151}$$

we get the exponent E_B for $R = 0$, and in fact, for all rates satisfying

$$R \leq \min\left\{I(\overline{Q}_B), \mathbf{s}\left(\tilde{Q}_Y; f_V(\overline{Q}_B)\right)\right\} \triangleq R_{\text{cr,B}}. \tag{152}$$

Similarly, this is also the exponent $E_{B,\text{L}}$. In conclusion, for all $R \leq R_{\text{cr}} \triangleq \min\{R_{\text{cr,A}}, R_{\text{cr,B}}\}$ it is assured that the FA exponents of ϕ' and ϕ_{L} are exactly the same. In the same manner, a critical rate can be found for the MD exponent. For the the high-rate simplified detector/decoder ϕ_{H} we only remark that in some cases, the output distributions $\tilde{W}$ and $\tilde{V}$ may be equal, and so this detector/decoder is useless, even though ϕ' achieves strictly positive exponents (cf. the example in Section VII).

5) *Continuous alphabet channels*: As previously mentioned, one of the advantages of the Gallager/Forney-Style bounds is their simple generalization to continuous channels with input constraints. We briefly describe this well known technique [21, Chapter 7]. For concreteness, let us focus on the power constraint $\mathbb{E}[X^2] \leq 1$. In this technique a one-dimensional input distribution is chosen, say with density $f_X(x)$, which satisfies the input constraint. Then, an n-dimensional distribution is defined as follows

$$f_n(\mathbf{x}) = \psi^{-1}\mathbb{I}\left\{n - \delta \leq \sum_{i=1}^{n} x_{m,i}^2 \leq n\right\}\prod_{i=1}^{n} P_X(x_i), \tag{153}$$

where ψ is a normalization factor. This distribution corresponds to a uniform distribution over a thin n-dimensional spherical shell, which is the surface of the n-dimensional 'ball' of sequences which satisfy the input constraint. While this input distribution is not memoryless, it is easily upper bounded by a memoryless distribution: by introducing a parameter $r \geq 0$, and using

$$\mathbb{I}\left\{n-\delta \leq \sum_{i=1}^{n} x_{m,i} \leq n\right\} \leq \exp\left[r \cdot \left(\sum_{i=1}^{n} x_{m,i}^2 - n + \delta\right)\right] \tag{154}$$

we get

$$f_n(\mathbf{x}) \leq \psi^{-1} e^{r\delta} \prod_{i=1}^{n} P_X(x_i) e^{r[x_i^2-1]}. \tag{155}$$

Now, e.g., in the derivation in (C.9) we may use

$$\mathbb{E}\left[W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)V^{s/\rho}(\mathbf{y}|\mathbf{X}_m)\right] = \int_{\mathbf{x}} f_n(\mathbf{x}) W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)V^{s/\rho}(\mathbf{y}|\mathbf{X}_m)d\mathbf{x} \tag{156}$$

$$\leq \psi^{-1} e^{r\delta}\left[\int_x f_X(x) e^{r[x^2-1]} W^{(1-s)/\rho}(y_i|x)V^{s/\rho}(y_i|x)dx\right]^n. \tag{157}$$

As discussed in [21, p. 341], the term $\psi^{-1}e^{r\delta}$ is sub-exponential, and can be disregarded. Now, the resulting exponential functions can be modified. For example, for a pair of power constrained AWGN channels W and V, we may define[15]

$$E_0'(s,\rho,r) \triangleq -\log \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f_X(x) e^{r[x^2-1]} W^{(1-s)/\rho}(y|x)V^{s/\rho}(y|x)dx\right)^{\rho} dy, \tag{158}$$

where the dependence in r was made explicit, and similarly,

$$E_0''(s,\rho,r_1,r_2) \triangleq -\log \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f_X(x) e^{r_1[x^2-1]} W^{(1-s)/\rho}(y|x)dx\right)^{\rho}\left(\int_{-\infty}^{\infty} f_X(x) e^{r_2[x^2-1]} V^{s/\rho}(y|x)dx\right)^{\rho} dy, \tag{159}$$

which requires two new parameters r_1, r_2. Then, the exponent in (140) can be computed exactly in the same way, with additional maximization over non-negative r, r_1, r_2. To obtain an explicit bound, it is required to choose an input distribution. The natural choice is the Gaussian distribution, which is appropriate from the channel coding perspective[16], and also enables to obtain analytic bounds. Of course, it might be very far from being optimal for the purpose of pure detection. Then, the integrals in (158) can be solved by 'completing the square' in the exponent of Gaussian distributions[17], and the optimal values of r and ρ can be found analytically [21, Section 7.4]. Here, since two channels are involved, and we also need to optimize over s, we have not been able to obtain simple expressions[18]. Nonetheless, the required optimization problem is only four-dimensional, and can be easily solved

[15]Since the additive noise has a density, the probability distributions in the bounds of subsection V-D can be simply replaced by densities, and the summations can be replaced by integrals.

[16]Nevertheless, it should be recalled that Gaussian input is optimal at high rates (above some critical rate). At low rates, the optimal input distribution is not known, even for pure channel coding.

[17]Namely, the identities $\int_{t=-\infty}^{\infty} \exp\left[-at^2-bt\right]dt = \sqrt{\frac{\pi}{a}} \cdot e^{\frac{b^2}{4a}}$ and $\int_{t=-\infty}^{\infty} \exp\left[-a\frac{t^2}{2}\right]dt = \sqrt{\frac{2\pi}{a}}$.

[18]Nonetheless, for a given s, the expression for $E_0'(s,\rho,r)$ is rather similar to the ordinary decoding exponent $E_0(\rho,r)$ and so the optimal ρ and r can be analytically found.

by an exhaustive search. Finally, it can be noticed that the computing the expurgated bounds is a similar problem as

$$E'_x(s,r) = E'_0(s,\rho=1,r) \tag{160}$$

and

$$E''_x(s,r) = E''_0(s,\rho=1,r). \tag{161}$$

6) *Comparison with* [10]: As mentioned in the introduction (Section I), the problem studied here is a generalization of [10]. Indeed, when the channel V does not depend on the input, i.e. $V(\mathbf{y}|\mathbf{x}) = Q_0(\mathbf{y})$, then the problem studied in [10] is obtained[19]. Of course, the detectors derived in Section IV can be used directly for this special case. Moreover, the exponent expressions can be slightly simplified as follows. A joint type $\tilde{Q}$ is feasible if and only if $f_W(P_X \times \tilde{Q}_Y) \leq -\alpha + f_V(P_X \times \tilde{Q}_Y)$, both in E_A of (47) and E_B of (52), as otherwise, the sets $\mathcal{J}_2$ and $\mathcal{K}_4$ are empty. For any such $\tilde{Q}$ which satisfies this condition, when utilizing the fact that $f_V(\overline{Q})$ depends only on $\overline{Q}_Y = \tilde{Q}_Y$, the optimal choice for E_B is $\overline{Q} = P_X \times \tilde{Q}_Y$, since it results $I(\overline{Q}) = 0$. Under this choice, we get $\mathcal{J}_1 \subset \mathcal{K}_3$ and $\mathcal{J}_2 \subset \mathcal{K}_4$ and so $E_A \geq E_B$. Thus, from (53)

$$E_{\text{TE}}^{\text{RC}}(R,\alpha,P_X,W,V) = \min_{\tilde{Q}\in\cap_{i=3}^{4}\mathcal{M}_i} D(\tilde{Q}_{Y|X}\|W|P_X) \tag{162}$$

where

$$\mathcal{M}_3 \triangleq \left\{\tilde{Q}:\ f_V(\tilde{Q}) \geq \alpha + f_W(\tilde{Q}) - R\right\}, \tag{163}$$

replaces $\mathcal{K}_3$, and

$$\mathcal{M}_4 \triangleq \left\{\tilde{Q}:\ \mathbf{s}\left(\tilde{Q}_Y, f_V(\tilde{Q}_Y) + R\right) \geq R\right\}, \tag{164}$$

replaces $\mathcal{K}_4$. Thus, the minimization in the exponent is only on $\tilde{Q}$.

VI. Composite Detection

Up until now, we have assumed that detection is performed between two simple hypotheses, namely W and V. In this section, we briefly discuss the generalization of the random coding analysis to composite hypotheses, to wit, a detection between a channel $W \in \mathcal{W}$ and a channel $V \in \mathcal{V}$, where $\mathcal{W}$ and $\mathcal{V}$ are disjoint. Due to the nature of the problems outlined in the introduction (Section I), we adopt a *worst case* approach. For a codebook $\mathcal{C}_n$ and a given detector/decoder ϕ, we generalize the FA probability to

$$P_{\text{FA}}(\mathcal{C}_n,\phi) \triangleq \max_{W\in\mathcal{W}} \frac{1}{M}\sum_{m=1}^{M} W(\mathcal{R}_0|\mathbf{x}_m), \tag{165}$$

and analogously, the MD and IE probabilities are obtained by maximizing over $V \in \mathcal{V}$ and $W \in \mathcal{W}$, respectively. Then, the trade-off between the IE probability and the FA and MD probabilities in (12) is defined exactly the same

[19]The meaning of FA and MD here is opposite to their respective meaning in [10], as sanctioned by the motivating applications.

way.

Just as we have seen in (22) (proof of Proposition 3), for any sequence of codebooks $\mathcal{C}_n$ and decoder ϕ

$$E_{\text{IE}}(\mathcal{C}_n, \phi) = \min\left\{E_{\text{O}}(\mathcal{C}_n, \phi), E_{\text{FA}}(\mathcal{C}_n, \phi)\right\} \tag{166}$$

where here, $E_{\text{O}}(\mathcal{C}_n, \phi)$ is the exponent achieved by an ordinary decoder, which is not aware of W. Thus, the asymptotic separation principle holds here too, in the sense that the optimal detector/decoder may first use a detector which achieves the optimal trade-off between the FA and MD exponents, and then a decoder which achieves the optimal ordinary exponent.

We next discuss the achievable random coding exponents. [20] As is well known, the *maximum mutual information* [31], [16, Chapter 10, p. 147] universally achieves the random for ordinary decoding. So, as in the simple hypotheses case, it remains to focus on the optimal trade-off between the FA and MD exponents, namely, solve

$$\begin{aligned} &\text{minimize} \quad P_{\text{FA}} \\ &\text{subject to} \quad P_{\text{MD}} \leq e^{-n\overline{E}_{\text{MD}}} \end{aligned} \tag{167}$$

for some given exponent $\overline{E}_{\text{MD}} > 0$. The next Lemma shows that the following *universal* detector/decoder ϕ^{U}, whose rejection region is

$$\mathcal{R}_0^{\text{U}} \triangleq \left\{\mathbf{y}:\ e^{n\alpha} \cdot \sum_{m=1}^{M} \max_{W\in\mathcal{W}} W(\mathbf{y}|\mathbf{x}_m) \leq \sum_{m=1}^{M} \max_{V\in\mathcal{V}} V(\mathbf{y}|\mathbf{x}_m)\right\}, \tag{168}$$

solves (167). The universality here is in the sense of (167), i.e., achieving the best worst-case (over W) FA exponent, under a worst case constraint (over V) on the MD exponent. There might be, however, a loss in exponents compared to a detector which is aware of the actual pair (W, V) (cf. Corollary 23).

Lemma 22. *Let $\mathcal{C} = \{\mathcal{C}_n\}$ be a given sequence of codebooks, let ϕ^{U} be as above, and let ϕ be any other partition of $\mathcal{Y}^n$ into $M+1$ regions. Then, if $E_{\text{FA}}(\mathcal{C}, \phi) \geq E_{\text{FA}}(\mathcal{C}, \phi^*)$ then $E_{\text{MD}}(\mathcal{C}, \phi) \leq E_{\text{MD}}(\mathcal{C}_n, \phi^*)$.*

Proof: The idea is that the maximum in (165) can be interchanged with the sum without affecting the exponential behavior. Specifically, let us define the sets of channels which maximize $f_W(Q)$ for some Q

$$\mathcal{W}_{\text{U}} \triangleq \left\{W \in \mathcal{W}:\ \exists Q \text{ such that } W = \arg\max_{W'\in\mathcal{W}} f_{W'}(Q)\right\}. \tag{169}$$

Clearly, since $f_W(Q)$ is only a function of the joint type, the cardinality of the sets $\mathcal{W}_{\text{U}}$ is not larger than the number of different joint types, and so their cardinality increases only polynomially with n. Then,

$$P_{\text{FA}}(\mathcal{C}_n, \phi) = \max_{W\in\mathcal{W}} \sum_{\mathbf{y}\in\mathcal{R}_0} \frac{1}{M} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) \tag{170}$$

$$\leq \sum_{\mathbf{y}\in\mathcal{R}_0} \frac{1}{M} \sum_{m=1}^{M} \max_{W\in\mathcal{W}} W(\mathbf{y}|\mathbf{x}_m) \tag{171}$$

[20] In universal decoding, typically only the random coding exponents are attempted to be achieved, cf. Remark 25.

$$= \sum_{\mathbf{y}\in\mathcal{R}_0} \frac{1}{M} \sum_{m=1}^{M} \max_{W\in\mathcal{W}_{\mathrm{U}}} W(\mathbf{y}|\mathbf{x}_m) \tag{172}$$

$$\triangleq \sum_{\mathbf{y}\in\mathcal{R}_0} g(\mathbf{y}) \tag{173}$$

$$\leq \sum_{\mathbf{y}\in\mathcal{R}_0} \frac{1}{M} \sum_{m=1}^{M} \sum_{W\in\mathcal{W}_{\mathrm{U}}} W(\mathbf{y}|\mathbf{x}_m) \tag{174}$$

$$= \sum_{W\in\mathcal{W}_{\mathrm{U}}} \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y}\in\mathcal{R}_0} W(\mathbf{y}|\mathbf{x}_m) \tag{175}$$

$$\doteq \max_{W\in\mathcal{W}_{\mathrm{U}}} \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y}\in\mathcal{R}_0} W(\mathbf{y}|\mathbf{x}_m) \tag{176}$$

$$\leq \max_{W\in\mathcal{W}} \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y}\in\mathcal{R}_0} W(\mathbf{y}|\mathbf{x}_m) \tag{177}$$

$$= P_{\text{FA}}(\mathcal{C}_n, \phi) \tag{178}$$

where the measure $g(\mathbf{y})$ was implicitly defined. Thus, up to a sub-exponential term which does not affect exponents,

$$P_{\text{FA}}(\mathcal{C}_n, \phi) \doteq \sum_{\mathbf{y}\in\mathcal{R}_0} g(\mathbf{y}). \tag{179}$$

Similarly, defining the measure

$$h(\mathbf{y}) \triangleq \frac{1}{M} \sum_{m=1}^{M} \max_{V\in\mathcal{V}} V(\mathbf{y}|\mathbf{x}_m) \tag{180}$$

we get

$$P_{\text{MD}}(\mathcal{C}_n, \phi) = \sum_{\mathbf{y}\in\overline{\mathcal{R}_0}} h(\mathbf{y}). \tag{181}$$

Now, the ordinary Neyman-Pearson lemma [18, Theorem 11.7.1] can be invoked[21] to show that the optimal detector is of the form (168), which completes the theorem. ■

It now remains to evaluate, for a given pair of channels $(W, V) \in \mathcal{W} \times \mathcal{V}$, the resulting random coding exponents when ϕ^{U} is used. Fortunately, this is an easy task given Theorem 6. Let us define the generalized normalized log-likelihood ratio of the set of channels $\mathcal{W}$ as

$$f_{\mathcal{W}}(Q) \triangleq \max_{W\in\mathcal{W}} \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} Q(x,y) \log W(y|x). \tag{182}$$

The following is easily verified.

Corollary 23 (to Theorem 6)**.** *Let a distribution P_X and a parameter $\alpha \in \mathbb{R}$ be given. Then, there exists a sequence*

[21]Note that the Neyman-Pearson lemma is also valid for general positive measures, not just for probability distributions. This can also be seen from the Lagrange formulation (28).

of codes $\mathcal{C} = \{\mathcal{C}_n\}_{n=1}^{\infty}$ *of rate* R*, such that for any* $\delta > 0$

$$E_{\text{FA}}\left(\mathcal{C}, \phi^{\text{U}}\right) \geq E_{\text{TE,U}}^{\text{RC}}\left(R, \alpha, P_X, W, V\right) - \delta, \tag{183}$$

$$E_{\text{MD}}\left(\mathcal{C}, \phi^{\text{U}}\right) \geq E_{\text{TE,U}}^{\text{RC}}\left(R, \alpha, P_X, W, V\right) - \alpha - \delta \tag{184}$$

where $E_{\text{TE,U}}^{\text{RC}}\left(R, \alpha, P_X, W, V\right)$ *is defined as* $E_{\text{TE}}^{\text{RC}}\left(R, \alpha, P_X, W, V\right)$ *of* (53)*, but replacing* $f_W(Q)$ *with* $f_{\mathcal{W}}(Q)$ *and* $f_V(Q)$ *with* $f_{\mathcal{V}}(Q)$ *in all the definitions preceding Theorem 6.*

We conclude with a few remarks.

Remark 24. The function $f_{\mathcal{W}}(Q)$ is a convex function of Q (as a pointwise maximum of linear functions), but not a linear function. This may harden the optimization problems involved in computing the exponents. Also, we implicitly assume that the set of channels $\mathcal{W}$ is sufficiently 'regular', so that $f_{\mathcal{W}}(Q)$ is a continuous function of Q.

Remark 25. The same technique works for the simplified low-rate detector/decoder. Unfortunately, since the bound (A.4) (Appendix A) utilizes the structure of the optimal detector/decoder, it is difficult to generalize the bounds which rely on it, namely, the expurgated exponents and the Gallager/Forney-style bounds. This is common to many other problem in universal decoding - for a non-exhaustive list of examples, see [32], [33], [34], [35], [36].

Remark 26. A different approach to composite hypothesis testing is the competitive minimax approach [37]. In this approach, a detector/decoder is sought which achieves the largest fraction of the error exponents achieved for a detection of only a pair of channels (W, V), uniformly over all possible pairs of channels (W, V). The application of this method on generalized decoders was exemplified for Forney's erasure/list decoder [17] in [38], [39], and the same techniques can work for this problem.

VII. An Example: A Detection of a Pair Binary Symmetric Channels

Let W and V be a pair of BSCs with crossover probabilities $w \in (0, 1)$ and $v \in (0, 1)$, respectively. In this case the exponent bounds of Section V can be greatly simplified, if the input distribution is uniform, i.e. $P_X = (\frac{1}{2}, \frac{1}{2})$. Indeed, in Appendix E we provide simplified expressions for the type-enumeration based exponents. Interestingly, while this input distribution is optimal from the channel coding perspective, the two output distributions $\tilde{W}$ and $\tilde{V}$ it induces are also uniform, and so the simple decoder which only uses the output statistics, namely ϕ_{H} of Subsection IV-B, is utterly useless. However, the optimal decoder ϕ' can produce strictly positive exponents.

We have plotted the FA exponent versus the MD exponent for the detection between two BSCs with $w = 0.1$ and $v = 0.4$. We have assumed the uniform input distribution $P_X = (\frac{1}{2}, \frac{1}{2})$, which results the capacity $C_W \triangleq I(P_X \times W) \approx 0.37$ (nats). Figure 1 shows that at zero rate, the expurgated bound which is based on type-enumeration significantly improves the random coding bound. In addition, the Gallager/Forney-style random coding exponent coincides with the exact exponent. By contrast, the Gallager/Forney-style expurgated exponent offers no improvement over the ordinary random coding bound (and thus not displayed). Figure 2 shows that at $R = 0.5 \cdot C_W$,

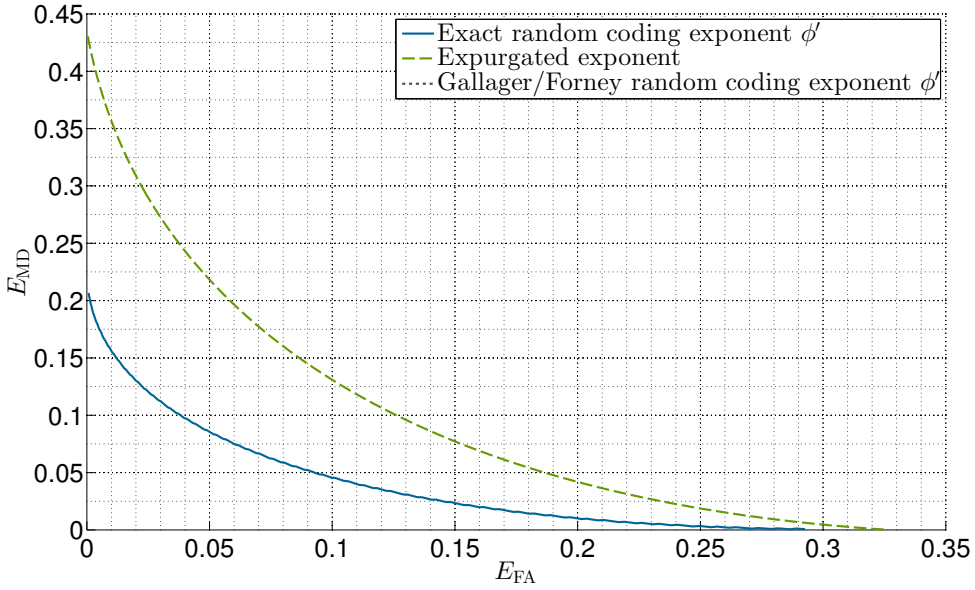

Figure 1. The trade-off between the FA exponent and the MD exponent at $R = 0$, for the detection of a BSC W with crossover probability 0.1, from a BSC V with crossover probability 0.4, when using the optimal detector ϕ'. The solid line corresponds to the exact random coding exponent, and also to the Gallager/Forney-style random coding exponent. The dashed line corresponds to the expurgated exponent.

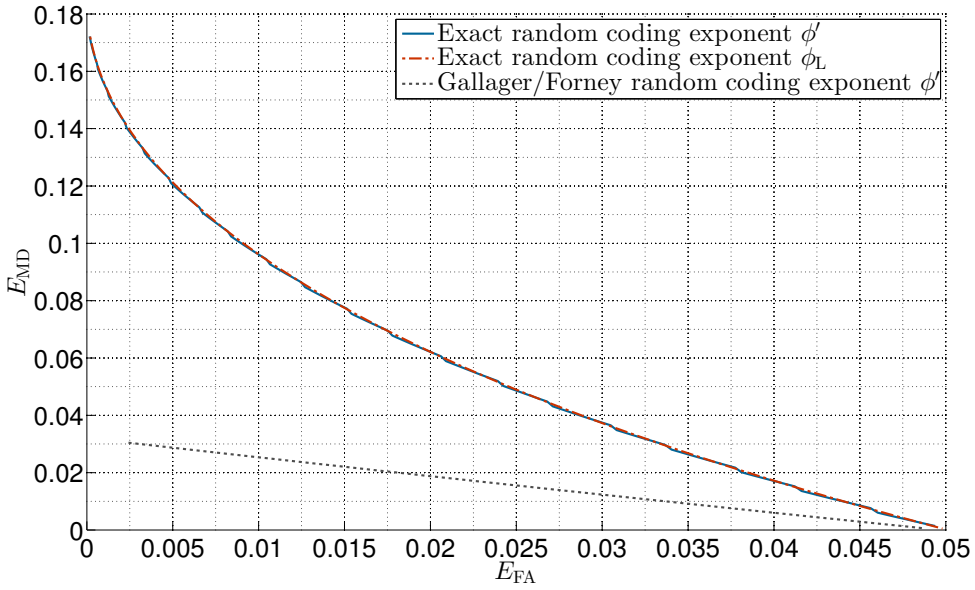

Figure 2. The trade-off between the FA exponent and the MD exponent at $R = 0.5 \cdot C_W$, for the detection of a BSC W with crossover probability 0.1, from a BSC V with crossover probability 0.4. The solid line corresponds to the exact random coding exponent of ϕ', and also to the exact random coding exponent of ϕ_{L}. The dotted line corresponds Gallager/Forney-style random coding exponent of ϕ'.

the simplified low-rate detector/decoder ϕ_{L} still performs as well as the optimal detector/decoder ϕ'. This, in fact continues to hold for all rates less than $R \approx 0.8 \cdot C_W$. In addition, it is evident that the Gallager/Forney-style random coding exponent is a poor bound, which exemplifies the importance of the ensemble-tight bounding technique of the type enumeration method.

APPENDIX A
PROOF OF THEOREM 13

Before getting into the proof, we derive a standard bound on the FA probability, which will also be used in Appendices C and D. For any given code and $s \geq 0$

$$P_{\text{FA}}(\mathcal{C}_n, \phi') = \sum_{\mathbf{y} \in \mathcal{R}_0'} \frac{1}{M} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) \tag{A.1}$$

$$= \sum_{\mathbf{y} \in \mathcal{R}_0'} \left[\frac{1}{M} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)\right]^{1-s} \left[\frac{1}{M} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)\right]^{s} \tag{A.2}$$

$$\overset{(a)}{\leq} e^{-n\alpha s} \sum_{\mathbf{y} \in \mathcal{R}_0'} \left[\frac{1}{M} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)\right]^{1-s} \left[\frac{1}{M} \sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m)\right]^{s} \tag{A.3}$$

$$\leq e^{-n\alpha s} \sum_{\mathbf{y} \in \mathcal{Y}^n} \left[\frac{1}{M} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)\right]^{1-s} \left[\frac{1}{M} \sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m)\right]^{s}, \tag{A.4}$$

where (a) is from (17).

Proof of Theorem 13: For a given code $\mathcal{C}_n$, a codeword $1 \leq m \leq M$, and a joint type $P_{X\tilde{X}}$, define the *type class enumerator*

$$\acute{N}_m(P_{X\tilde{X}}, \mathcal{C}_n) \triangleq \left|\left\{\mathbf{x} \in \mathcal{C}_n \backslash \mathbf{x}_m : \ \hat{Q}_{\mathbf{x}_m \mathbf{x}} = P_{X\tilde{X}}\right\}\right|. \tag{A.5}$$

Upon restricting $0 \leq s \leq 1$ in (A.4), we obtain the bound

$$P_{\text{FA}}(\mathcal{C}_n, \phi') \leq e^{-n\alpha s} \sum_{\mathbf{y} \in \mathcal{Y}^n} \left[\frac{1}{M} \sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)\right]^{1-s} \left[\frac{1}{M} \sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m)\right]^{s} \tag{A.6}$$

$$\overset{(a)}{\leq} e^{-n\alpha s} \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{M} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^{1-s}(\mathbf{y}|\mathbf{x}_m) V^{s}(\mathbf{y}|\mathbf{x}_k) \tag{A.7}$$

$$\overset{(b)}{=} e^{-n\alpha s} \frac{1}{M} \sum_{m=1}^{M} \sum_{P_{X\tilde{X}}} \acute{N}_m(P_{X\tilde{X}}, \mathcal{C}_n) \exp\left[-n\left(\mathbb{E}_{P_{X\tilde{X}}}\left[d_s(X, \tilde{X})\right]\right)\right], \tag{A.8}$$

where (a) follows from $\sum_i a_i^{\nu} \geq \left(\sum_i a_i\right)^{\nu}$ for $\nu \leq 1$, and (b) is using (A.5) and (119). Now, the packing lemma [16, Problem 10.2] essentially shows (see also [29, Appendix]) that for any $\delta > 0$, there exists a code $\mathcal{C}_n^*$ (of rate R) such that

$$\acute{N}_m(P_{X\tilde{X}}, \mathcal{C}_n^*) \leq \begin{cases} \exp\left[n\left(R + \delta - I(P_{X\tilde{X}})\right)\right], & I(P_{X\tilde{X}}) \leq R + \delta \\ 0, & I(P_{X\tilde{X}}) > R + \delta \end{cases} \tag{A.9}$$

for all $1 \leq m \leq M$ and $P_{X\tilde{X}}$. This, along with Proposition 4 completes the proof of the theorem. ■

APPENDIX B
PROOF OF THEOREM 17

The proof is very similar to the proof of Theorem 6. We will use the following lemma, which is analogous to Lemma 9.

Lemma 27. *Under the conditions of Lemma 9,*

$$\mathbb{P}\left(\bigcap_{Q\in\mathcal{Q}:\,Q_Y=\tilde{Q}_Y}\left\{\mathbb{I}\left\{\hat{N}(Q|\mathbf{y})\geq 1\right\}<e^{nJ(Q)}\right\}\right)\begin{cases}=1-o(n), & \mathbf{T}(\tilde{Q}_Y;J,\mathcal{Q})>R\\ \doteq e^{-n\infty}, & \textit{otherwise}\end{cases}, \tag{B.1}$$

where $\mathbf{y}\in\mathcal{T}(\tilde{Q}_Y)$*, and*

$$\mathbf{T}(\tilde{Q}_Y;J,\mathcal{Q})\triangleq\min_{Q\in\mathcal{Q}:\,Q=\tilde{Q}_Y,J(Q)\leq 0}I(Q). \tag{B.2}$$

Proof: We have

$$\mathbb{P}\left(\bigcap_{Q\in\mathcal{Q}:\,Q_Y=\tilde{Q}_Y}\left\{\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}<e^{nJ(Q)}\right\}\right)=\mathbb{P}\left(\bigcap_{Q\in\mathcal{Q}:\,Q_Y=\tilde{Q}_Y,J(Q)\leq 0}\left\{\mathbb{I}\left\{N(Q|\mathbf{y})=0\right\}\right\}\right). \tag{B.3}$$

From this point onward, the proof follows the same lines of the proof of Lemma 9. ■

Remark 28. Remarks 10 and 11 are also valid here. If $J(Q)$ is convex in $Q_{Y|X}$ then Lagrange duality [26, Chapter 5] implies

$$\mathbf{T}(\tilde{Q}_Y;J,\mathcal{Q})=\min_{Q\in\mathcal{Q}:\,Q=\tilde{Q}_Y}\max_{\lambda\geq 0}\left[I(Q)+\lambda J(Q)\right] \tag{B.4}$$

$$=\max_{\lambda\geq 0}\min_{Q\in\mathcal{Q}:\,Q=\tilde{Q}_Y}\left[I(Q)+\lambda J(Q)\right]. \tag{B.5}$$

The only difference from $\mathbf{S}(\tilde{Q}_Y;J,\mathcal{Q})$ of (73) in this case is the maximization domain for λ. Note that the function $\mathbf{t}(\tilde{Q}_Y;\gamma)$ of (124) is a specific instance of $\mathbf{T}(\tilde{Q}_Y;\cdot,\cdot)$ defined in (B.2) with $\mathcal{Q}=\mathcal{Q}_W$ and $J(Q)=-\alpha-f_W(Q)+\gamma$ which is convex in $Q_{Y|X}$ (in fact, linear).

Proof of Theorem 17: In general, since

$$\sum_{m=2}^{M}W(\mathbf{y}|\mathbf{x}_m)=\sum_{Q}N(Q|\mathbf{y})e^{nf_W(Q)} \tag{B.6}$$

but

$$\max_{2\leq m\leq M}W(\mathbf{y}|\mathbf{x}_m)=\max_{Q}\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}e^{nf_W(Q)} \tag{B.7}$$

$$\doteq\sum_{Q}\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}e^{nf_W(Q)}, \tag{B.8}$$

then the analysis of the FA exponent of ϕ_{L} follows the same lines as the analysis in the proof of Theorem 6, when replacing $N(Q|\mathbf{y})$ with $\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}$. Thus, in the following we only highlight the main changes. Just as in the

derivations leading to (80),

$$\overline{P_{\text{FA}}}(\mathbf{x}_1,\mathbf{y}) \triangleq \mathbb{P}\left(\mathbf{y}\in\mathcal{R}_{0,\text{L}}|\mathbf{X}_1=\mathbf{x}_1,\mathbf{Y}=\mathbf{y}\right) \tag{B.9}$$

$$\doteq \max\left\{A_{\text{L}}(\tilde{Q}),B_{\text{L}}(\tilde{Q})\right\}, \tag{B.10}$$

where

$$A_{\text{L}}(\tilde{Q}) \triangleq \mathbb{P}\left(\sum_{Q}\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}e^{nf_W(Q)}\leq e^{-n\alpha}\cdot e^{nf_V(\tilde{Q})}\right)\cdot\mathbb{I}\left\{f_W(\tilde{Q})\leq -\alpha+f_V(\tilde{Q})\right\} \tag{B.11}$$

and

$$B_{\text{L}}(\tilde{Q}) \triangleq \mathbb{P}\left(e^{nf_W(\tilde{Q})}+\max_{Q}\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}e^{nf_W(Q)}\leq e^{-n\alpha}\cdot\max_{Q}\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}e^{nf_V(Q)}\right). \tag{B.12}$$

For the first term,

$$A_{\text{L}}(\tilde{Q}) \overset{(IR)}{\doteq} \mathbb{P}\left(\bigcap_{Q:\,f_W(Q)>-\infty}\left\{\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}<e^{n\left[-\alpha+f_V(\tilde{Q})-f_W(Q)\right]}\right\}\right)\cdot\mathbb{I}\left\{f_W(\tilde{Q})\leq -\alpha+f_V(\tilde{Q})\right\} \tag{B.13}$$

$$\overset{(a)}{\doteq} \mathbb{I}\left\{\mathbf{T}(\tilde{Q}_Y;-\alpha+f_V(\tilde{Q})-f_W(Q),\mathcal{Q}_W)>R\right\}\cdot\mathbb{I}\left\{f_W(\tilde{Q})\leq -\alpha+f_V(\tilde{Q})\right\}, \tag{B.14}$$

where (a) is by Lemma 27. Upon averaging over $(\mathbf{X}_1,\mathbf{Y})$, we obtain the exponent $E_{A,\text{L}}$ of (126) (utilizing the definition (124)).

Moving on to the second term, similarly as in the analysis leading to (87)

$$B_{\text{L}}(\tilde{Q}) \doteq \sum_{\overline{Q}:\,f_W(\overline{Q})\leq -\alpha+f_V(\overline{Q})} \mathbb{P}\left(\bigcap_{Q\neq\overline{Q}:\,f_W(Q)>-\infty}\left\{\mathbb{I}\left\{N(Q|\mathbf{y})\geq 1\right\}\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(Q)\right]}\cdot\mathbb{I}\left\{N(\overline{Q}|\mathbf{y})\geq 1\right\}\right\}\cap \left\{1\leq e^{n\left[-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})\right]}\cdot\mathbb{I}\left\{N(\overline{Q}|\mathbf{y})\geq 1\right\}\right\}\right) \tag{B.15}$$

$$\triangleq \sum_{\overline{Q}:\,f_W(\overline{Q})\leq -\alpha+f_V(\overline{Q})}\zeta_{\text{L}}(\overline{Q}). \tag{B.16}$$

We now split the analysis into three cases:

Cases 1 and 2: Assume $0\leq I(\overline{Q})<R$. An analysis similar to cases 1 and 2 in the proof of Theorem 6 shows that

$$\zeta_{\text{L}}(\overline{Q}) \doteq \mathbb{I}\left\{\mathbf{T}(\tilde{Q}_Y;-\alpha+f_V(\overline{Q})-f_W(Q),\mathcal{Q}_W)>R\right\}\mathbb{I}\left\{-\alpha+f_V(\overline{Q})-f_W(\tilde{Q})>0\right\}. \tag{B.17}$$

Case 3: Assume that $I(\overline{Q})>R$. An analysis similar to case 3 in the proof of Theorem 6 shows that the inner

probability in (B.16) is exponentially equal to

$$\zeta_{\text{L}}(\overline{Q}) \doteq \mathbb{I}\left\{\mathbf{T}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q), \mathcal{Q}_W) > R\right\} \mathbb{I}\left\{-\alpha + f_V(\overline{Q}) - f_W(\tilde{Q}) > 0\right\} e^{-n\left(I(\overline{Q})-R\right)}. \quad \text{(B.18)}$$

Returning to (B.16) we obtain that $B_{\text{L}}(\tilde{Q})$ is exponentially equal to the maximum between

$$\max_{\overline{Q}:\, f_W(\overline{Q}) < -\alpha + f_V(\overline{Q}),\, I(\overline{Q}) < R,\, f_V(\overline{Q}) > \alpha + f_W(\tilde{Q})} \mathbb{I}\left\{\mathbf{T}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q), \mathcal{Q}_W) > R\right\}, \quad \text{(B.19)}$$

and

$$\max_{\overline{Q}:\, f_W(\overline{Q}) < -\alpha + f_V(\overline{Q}),\, I(\overline{Q}) \geq R,\, f_V(\overline{Q}) > \alpha + f_W(\tilde{Q})} \mathbb{I}\left\{\mathbf{T}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q), \mathcal{Q}_W) > R\right\} e^{-n\left(I(\overline{Q})-R\right)}, \quad \text{(B.20)}$$

or, more succinctly,

$$B(\tilde{Q}) = \max_{\overline{Q}} \mathbb{I}\left\{\mathbf{T}(\tilde{Q}_Y; -\alpha + f_V(\overline{Q}) - f_W(Q), \mathcal{Q}_W) > R\right\} e^{-n\left[I(\overline{Q})-R\right]_+} \quad \text{(B.21)}$$

where the maximization is over

$$\left\{\overline{Q}:\ f_W(\overline{Q}) < -\alpha + f_V(\overline{Q}),\ f_V(\overline{Q}) \geq \alpha + f_W(\tilde{Q})\right\}. \quad \text{(B.22)}$$

Upon averaging over $(\mathbf{X}_1, \mathbf{Y})$, we obtain the exponent $E_{B,\text{L}}$ of (129) (utilizing again (124)), and the proof of the FA exponent (131) is proved using (B.10).

For the MD expression, since ϕ_{L} is not necessarily the optimal detector in the Neyman-Pearson sense, we cannot use Proposition 4. However, due to the symmetry in $\mathcal{R}_{0,\text{L}}$ of W and V, a similar observation as in Fact 7 holds, which leads directly to (132). The rest of the proof follows the same lines as the proof of theorem 6. ■

APPENDIX C
PROOF OF THEOREM 20

Proof of Theorem 20: As in the proof of Theorem 6, we only need to upper bound the FA probability as the MD probability can be easily evaluated from the FA bound, using Proposition 4. It remains to derive an upper bound on the average FA error probability. We assume the ensemble of randomly selected codes of size $M = \left\lceil e^{nR} \right\rceil$, where each codeword is selected independently at random, with i.i.d. components from the distribution P_X. Introducing a parameter $\rho \geq \max\{s, 1-s\}$, we continue the bound (A.4) as follows:

$$P_{\text{FA}}(\mathcal{C}_n, \phi') \leq e^{-n(\alpha s + R)} \sum_{\mathbf{y} \in \mathcal{Y}^n} \left[\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m)\right]^{\rho(1-s)/\rho} \left[\sum_{m=1}^{M} V(\mathbf{y}|\mathbf{x}_m)\right]^{\rho s/\rho} \quad \text{(C.1)}$$

$$\overset{(a)}{\leq} e^{-n(\alpha s + R)} \sum_{\mathbf{y} \in \mathcal{Y}^n} \left[\sum_{m=1}^{M} W^{(1-s)/\rho}(\mathbf{y}|\mathbf{x}_m)\right]^{\rho} \left[\sum_{m=1}^{M} V^{s/\rho}(\mathbf{y}|\mathbf{x}_m)\right]^{\rho} \quad \text{(C.2)}$$

$$= e^{-n(\alpha s + R)} \sum_{\mathbf{y} \in \mathcal{Y}^n} \left[\sum_{m=1}^{M} \sum_{k=1}^{M} W^{(1-s)/\rho}(\mathbf{y}|\mathbf{x}_m) V^{s/\rho}(\mathbf{y}|\mathbf{x}_k)\right]^{\rho}, \quad \text{(C.3)}$$

where (a) follows from $(\sum_i a_i)^\nu \leq \sum_i a_i^\nu$ for $\nu \leq 1$. Using now the fact that the codewords are selected at random, we obtain

$$\overline{P_{\text{FA}}}(\mathcal{C}_n, \phi') \leq e^{-n(\alpha s+R)} \sum_{\mathbf{y}\in\mathcal{Y}^n} \mathbb{E}\left\{\left[\sum_{m=1}^{M}\sum_{k=1}^{M} W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)V^{s/\rho}(\mathbf{y}|\mathbf{X}_k)\right]^\rho\right\} \tag{C.4}$$

$$\overset{(a)}{\leq} e^{-n(\alpha s+R)} \sum_{\mathbf{y}\in\mathcal{Y}^n} \left\{\sum_{m=1}^{M}\sum_{k=1}^{M} \mathbb{E}\left[W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)V^{s/\rho}(\mathbf{y}|\mathbf{X}_k)\right]\right\}^\rho, \tag{C.5}$$

where (a) is by restricting $\rho \leq 1$ and using Jensen Inequality. For a given $\mathbf{y}$, let us focus on the inner expectation. If $m = k$ then

$$\mathbb{E}\left[W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)V^{s/\rho}(\mathbf{y}|\mathbf{X}_m)\right] = \mathbb{E}\left[\prod_{i=1}^{n} W^{(1-s)/\rho}(y_i|X_{m,i})V^{s/\rho}(y_i|X_{m,i})\right] \tag{C.6}$$

$$= \prod_{i=1}^{n} \mathbb{E}\left[W^{(1-s)/\rho}(y_i|X_{m,i})V^{s/\rho}(y_i|X_{m,i})\right] \tag{C.7}$$

$$= \prod_{i=1}^{n} \left(\sum_{x\in\mathcal{X}} P_X(x)W^{(1-s)/\rho}(y_i|x)V^{s/\rho}(y_i|x)\right) \tag{C.8}$$

$$\triangleq \Psi_{s,\rho}(\mathbf{y}). \tag{C.9}$$

Otherwise, if $m \neq k$, then since the codewords are selected independently

$$\mathbb{E}\left[W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)V^{s/\rho}(\mathbf{y}|\mathbf{X}_k)\right] = \mathbb{E}\left[W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)\right]\mathbb{E}\left[V^{s/\rho}(\mathbf{y}|\mathbf{X}_k)\right] \tag{C.10}$$

$$= \mathbb{E}\left[\prod_{i=1}^{n} W^{(1-s)/\rho}(y_i|X_{m,i})\right]\mathbb{E}\left[\prod_{i=1}^{n} V^{s/\rho}(y_i|X_{k,i})\right] \tag{C.11}$$

$$= \prod_{i=1}^{n} \mathbb{E}\left[W^{(1-s)/\rho}(y_i|X_{m,i})\right]\mathbb{E}\left[V^{s/\rho}(y_i|X_{k,i})\right] \tag{C.12}$$

$$= \prod_{i=1}^{n} \left(\sum_{x\in\mathcal{X}} P_X(x)W^{(1-s)/\rho}(y_i|x)\right)\left(\sum_{x\in\mathcal{X}} P_X(x)V^{s/\rho}(y_i|x)\right) \tag{C.13}$$

$$\triangleq \Gamma_{s,\rho}(\mathbf{y}). \tag{C.14}$$

So, the double inner summand in (C.5) is bounded as

$$\left\{\sum_{m=1}^{M}\sum_{k=1}^{M} \mathbb{E}\left[W^{(1-s)/\rho}(\mathbf{y}|\mathbf{X}_m)V^{s/\rho}(\mathbf{y}|\mathbf{X}_k)\right]\right\}^\rho = \{M\Psi_{s,\rho}(\mathbf{y}) + M(M-1)\Gamma_{s,\rho}(\mathbf{y})\}^\rho \tag{C.15}$$

$$\leq 2^\rho \max\left\{M^\rho \Psi^\rho_{s,\rho}(\mathbf{y}), M^{2\rho}\Gamma^\rho_{s,\rho}(\mathbf{y})\right\}, \tag{C.16}$$

using $\{c+d\}^\rho \leq [2\max\{c,d\}]^\rho$ for any $c, d \geq 0$. Thus, we may continue the bound of (C.5) as

$$\overline{P_{\text{FA}}}(\mathcal{C}_n, \phi') \leq e^{-n(\alpha s+R)} 2^\rho \max\left\{\sum_{\mathbf{y}\in\mathcal{Y}^n} M^\rho \Psi^\rho_{s,\rho}(\mathbf{y}), \sum_{\mathbf{y}\in\mathcal{Y}^n} M^{2\rho}\Gamma^\rho_{s,\rho}(\mathbf{y})\right\}. \tag{C.17}$$

The first term in the above maximization is given by

$$e^{-n\left(\alpha s-(\rho-1)R-\frac{\rho\log 2}{n}\right)}\sum_{\mathbf{y}\in\mathcal{Y}^n}\prod_{i=1}^{n}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{(1-s)/\rho}(y_i|x)V^{s/\rho}(y_i|x)\right)^{\rho} \tag{C.18}$$

$$=e^{-n\left(\alpha s-(\rho-1)R-\frac{\rho\log 2}{n}\right)}\prod_{i=1}^{n}\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{(1-s)/\rho}(y|x)V^{s/\rho}(y|x)\right)^{\rho} \tag{C.19}$$

$$=e^{-n\left(\alpha s-(\rho-1)R-\frac{\rho\log 2}{n}\right)}\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{(1-s)/\rho}(y|x)V^{s/\rho}(y|x)\right)^{\rho}\right]^{n} \tag{C.20}$$

$$=\exp\left[-n\cdot\left(\alpha s+E_0'(s,\rho)-(\rho-1)R-\frac{\rho\log 2}{n}\right)\right] \tag{C.21}$$

where $E_0'(s,\rho)$ was defined in (138). In a similar manner, the second term in the maximization is given by

$$e^{-n\left(\alpha s-(2\rho-1)R-\frac{\rho\log 2}{n}\right)}\sum_{\mathbf{y}\in\mathcal{Y}^n}\prod_{i=1}^{n}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{(1-s)/\rho}(y_i|x)\right)^{\rho}\left(\sum_{x\in\mathcal{X}}P_X(x)V^{s/\rho}(y_i|x)\right)^{\rho} \tag{C.22}$$

$$\le e^{-n\left(\alpha s-(2\rho-1)R-\frac{\rho\log 2}{n}\right)}\sum_{\mathbf{y}\in\mathcal{Y}^n}\prod_{i=1}^{n}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{(1-s)/\rho}(y_i|x)\right)^{\rho}\left(\sum_{x\in\mathcal{X}}P_X(x)V^{s/\rho}(y_i|x)\right)^{\rho} \tag{C.23}$$

$$= e^{-n\left(\alpha s-(2\rho-1)R-\frac{\rho\log 2}{n}\right)}\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{(1-s)/\rho}(y|x)\right)^{\rho}\left(\sum_{x\in\mathcal{X}}P_X(x)V^{s/\rho}(y|x)\right)^{\rho}\right]^{n} \tag{C.24}$$

$$= \exp\left[-n\cdot\left(\alpha s+E_0''(s,\rho)-(2\rho-1)R-\frac{\rho\log 2}{n}\right)\right] \tag{C.25}$$

where $E_0''(s,\rho)$ was defined in (139). Definition (140) then implies the achievability in (141). ■

APPENDIX D
PROOF OF THEOREM 21

Proof of Theorem 21: Let us begin with the FA probability. We start again from the bound (A.4) and restrict $s\le 1$

$$P_{\text{FA}}(\mathcal{C}_n,\phi')\le e^{-n\alpha s}\frac{1}{M}\sum_{\mathbf{y}\in\mathcal{Y}^n}\left[\sum_{m=1}^{M}W(\mathbf{y}|\mathbf{x}_m)\right]^{1-s}\left[\sum_{m=1}^{M}V(\mathbf{y}|\mathbf{x}_m)\right]^{s} \tag{D.1}$$

$$\overset{(a)}{\le} e^{-n\alpha s}\frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}^n}W^{1-s}(\mathbf{y}|\mathbf{x}_m)V^{s}(\mathbf{y}|\mathbf{x}_k) \tag{D.2}$$

where (a) follows from $\sum_i a_i^{\nu}\ge\left(\sum_i a_i\right)^{\nu}$ for $\nu\le 1$. Let us denote the random variable

$$Z_m\triangleq\sum_{k=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}^n}W^{1-s}(\mathbf{y}|\mathbf{X}_m)V^{s}(\mathbf{y}|\mathbf{X}_k) \tag{D.3}$$

over a random choice of codewords from i.i.d. distribution P_X. Introducing a parameter $\rho \geq 1$, for any given $B > 0$, we may use the classical variation of the Markov inequality, as e.g. in [17, Eqs. (96)-(98)],

$$\mathbb{P}\left(Z_m \geq B\right) \leq \mathbb{E}\left[\sum_{k=1}^{M} \frac{\left[\sum_{\mathbf{y}\in\mathcal{Y}^n} W^{1-s}(\mathbf{y}|\mathbf{X}_m)V^s(\mathbf{y}|\mathbf{X}_k)\right]^{1/\rho}}{B^{1/\rho}}\right] \tag{D.4}$$

$$= B^{-1/\rho}\sum_{k=1}^{M}\mathbb{E}\left\{\left[\sum_{\mathbf{y}\in\mathcal{Y}^n} W^{1-s}(\mathbf{y}|\mathbf{X}_m)V^s(\mathbf{y}|\mathbf{X}_k)\right]^{1/\rho}\right\} \tag{D.5}$$

$$\overset{(a)}{\leq} B^{-1/\rho}\sum_{k=1}^{M}\left\{\sum_{\mathbf{y}\in\mathcal{Y}^n}\mathbb{E}\left[W^{1-s}(\mathbf{y}|\mathbf{X}_m)V^s(\mathbf{y}|\mathbf{X}_k)\right]\right\}^{1/\rho} \tag{D.6}$$

$$= B^{-1/\rho}\left\{\left[\sum_{\mathbf{y}\in\mathcal{Y}^n}\Psi_{s,1}(\mathbf{y})\right]^{1/\rho} + (M-1)\left[\sum_{\mathbf{y}\in\mathcal{Y}^n}\Gamma_{s,1}(\mathbf{y})\right]^{1/\rho}\right\} \tag{D.7}$$

$$< B^{-1/\rho}2\cdot\max\left\{\left[\sum_{\mathbf{y}\in\mathcal{Y}^n}\Psi_{s,1}(\mathbf{y})\right]^{1/\rho}, M\left[\sum_{\mathbf{y}\in\mathcal{Y}^n}\Gamma_{s,1}(\mathbf{y})\right]^{1/\rho}\right\}, \tag{D.8}$$

where (a) follows from Jensen inequality, and we have used the definitions of $\Gamma_{s,\rho}(\mathbf{y})$ and $\Psi_{s,\rho}(\mathbf{y})$ from (C.14) and (C.9). Now, as

$$\sum_{\mathbf{y}\in\mathcal{Y}^n}\Psi_{s,1}(\mathbf{y}) = \sum_{\mathbf{y}\in\mathcal{Y}^n}\prod_{i=1}^{n}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{1-s}(y_i|x)V^s(y_i|x)\right) \tag{D.9}$$

$$= \left[\sum_{x\in\mathcal{X}}P_X(x)\sum_{y\in\mathcal{Y}}W^{1-s}(y|x)V^s(y|x)\right]^n, \tag{D.10}$$

and

$$\sum_{\mathbf{y}\in\mathcal{Y}^n}\Gamma_{s,1}(\mathbf{y}) = \sum_{\mathbf{y}\in\mathcal{Y}^n}\prod_{i=1}^{n}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{1-s}(y_i|x)\right)\left(\sum_{x\in\mathcal{X}}P_X(x)V^s(y_i|x)\right) \tag{D.11}$$

$$= \left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}}P_X(x)W^{1-s}(y|x)\right)\left(\sum_{x\in\mathcal{X}}P_X(x)V^s(y|x)\right)\right]^n, \tag{D.12}$$

then using the definition of $E'_x(s)$ and $E''_x(s)$ in (143) and (144), respectively, as well as

$$F_x(s,\rho,\alpha,P_X) \triangleq \min\left\{\frac{1}{\rho}E'_x(s), \frac{1}{\rho}E''_x(s) - R\right\}, \tag{D.13}$$

we get that (D.8) is

$$\mathbb{P}\left(Z_m \geq B\right) \leq 2B^{-1/\rho}\cdot\exp\left[-n\cdot F_x(s,\rho,\alpha)\right]. \tag{D.14}$$

For any given $\delta > 0$ let us choose

$$B^* = e^{n\delta/2}4^{\rho}\exp\left[-n\cdot\rho F_x(s,\rho,\alpha)\right] \tag{D.15}$$

we obtain

$$\mathbb{P}\left(Z_m \geq B^*\right) < \frac{1}{2} e^{-n\delta/2\rho}. \tag{D.16}$$

So, if we expurgate $\frac{1}{2}$ of the bad codewords in a randomly chosen codebook, then

$$\mathbb{P}\left(\bigcup_{m=1}^{M} \{Z_m \geq B^*\}\right) < e^{-n\delta/2\rho} \tag{D.17}$$

where the probability is over the random codebooks (note also that this expurgation only causes the sum over k in (D.3) to decrease). Indeed, to see this, define $\mathfrak{C}_n$ as the set of 'bad' codes which have $\{Z_m > B^*\}$ for more than half of the codewords. Assume by contradiction, that the probability of a 'bad' code is larger than $e^{-\frac{n\delta}{2\rho}}$. Hence, from the symmetry of the codewords

$$\mathbb{P}\left(Z_m \geq B^*\right) = \sum_{\mathcal{C}_n} \mathbb{P}\left(\mathcal{C}_n\right) \mathbb{I}\left\{Z_m > B^*\right\} \tag{D.18}$$

$$= \sum_{\mathcal{C}_n} \mathbb{P}\left(\mathcal{C}_n\right) \frac{1}{M} \sum_{m=1}^{m} \mathbb{I}\left\{Z_m > B^*\right\} \tag{D.19}$$

$$\geq \sum_{\mathcal{C}_n \in \mathfrak{C}_n} \mathbb{P}\left(\mathcal{C}\right) \frac{1}{2} \tag{D.20}$$

$$\geq \frac{1}{2} e^{-n\delta/2\rho}, \tag{D.21}$$

which contradicts (D.16). Namely, if we expurgate $\frac{1}{2}$ of the bad codewords of each codebook, then

$$\overline{P_{\text{FA}}}(\mathcal{C}_n, \phi') \leq \exp\left[-n \cdot \left(E_{\text{GF}}^{\text{EX}}\left(R, \alpha, P_X, W, V\right) - \delta\right)\right] \tag{D.22}$$

for all sufficiently large n, with probability tending exponentially fast to 1 (over the random ensemble). Then, Proposition 4 implies that also

$$\overline{P_{\text{MD}}}(\mathcal{C}_n, \phi') \leq \exp\left[-n \cdot \left(E_{\text{GF}}^{\text{EX}}\left(R, \alpha, P_X, W, V\right) - \alpha - \delta\right)\right]. \tag{D.23}$$

Thus, one can find a *single* sequence of codebooks, of size larger than $\frac{M}{2}$ which simultaneously achieves both upper bounds above. ■

APPENDIX E
SIMPLIFIED EXPRESSIONS FOR BSC

In Subsection V-A (respectively, V-C), the exponents (47) and (52) (respectively, (126) and (129)) are given as minimization problems over the joint types $\tilde{Q}, \overline{Q}$, and also over Q, via $\mathbf{s}(\tilde{Q}_Y, \gamma)$ (respectively, $\mathbf{t}(\tilde{Q}_Y, \gamma)$). These joint types are constrained to $\tilde{Q}_X = \overline{Q}_X = Q_X = P_X$ and $\tilde{Q}_Y = \overline{Q}_Y = Q_Y$. To obtain simplified expressions, we will show that the optimal joint types are symmetric, to wit, they result from an input distributed according to P_X which undergoes a BSC. Thus, as both the input and output distributions for such symmetric joint types are uniform, it is only remains to optimize over the crossover probabilities $\tilde{q}, \overline{q}, q$.

To prove the above claim, we introduce some new notation of previously defined quantities, but specified for the binary symmetric case. For $q, q_1, q_2 \in [0,1]$, the *binary normalized log likelihood ratio* is defined as

$$f_{w,\mathrm{B}}(q) \triangleq \frac{1}{n}\log\left[w^{qn}(1-w)^{(1-q)n}\right] \tag{E.1}$$

$$= \log(1-w) - q\rho_w, \tag{E.2}$$

where $\rho_w \triangleq \log\frac{1-w}{w}$, the *binary entropy* is denoted by

$$h_{\mathrm{B}}(q) \triangleq -q\log q - (1-q)\log(1-q), \tag{E.3}$$

and the *binary information divergence* is denoted by

$$D_{\mathrm{B}}(q_1||q_2) \triangleq q_1\log\frac{q_1}{q_2} + (1-q_1)\log\frac{(1-q_1)}{(1-q_2)}. \tag{E.4}$$

For a given type Q, let us define the *average crossover probability*

$$\hat{q}(Q) \triangleq \frac{1}{2}[Q_{Y|X}(0|1) + Q_{Y|X}(1|0)], \tag{E.5}$$

and let $\mathcal{Q}$ be a set of joint types, for which the inclusion of Q in $\mathcal{Q}$ depends on Q only via $\hat{q}(Q)$. It is easy to verify the following facts:

1) The information divergence satisfies

$$\min_{Q_{Y|X}\in\mathcal{Q}} D(Q_{Y|X}||W|P_X) = \min_{0\le q\le 1} D_{\mathrm{B}}(q||w). \tag{E.6}$$

from the convexity of the information divergence in $Q_{Y|X}$ and symmetry of P_X and W.

2) The normalized log likelihood ratio $f_W(Q)$ depends on Q only via $\hat{q}(Q)$, and so

$$f_W(Q) = \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} Q(x,y)\log W(y|x) \tag{E.7}$$

$$= (1-\hat{q}(Q))\log(1-w) + \hat{q}(Q)\log(w) \tag{E.8}$$

$$= f_{w,\mathrm{B}}\left(\hat{q}(Q)\right). \tag{E.9}$$

3) Let $L(q)$ be a linear function of q. Then

$$\max_{\tilde{Q}_Y}\min_{Q:\, Q_Y=\tilde{Q}_Y}\left\{I(Q) + L\left[\hat{q}(Q)\right]\right\} = \min_{0\le q\le 1}\left\{\log 2 - h_{\mathrm{B}}(q) + L(q)\right\}. \tag{E.10}$$

To see this, note that $I(Q)$ is concave in $\tilde{Q}_Y$ (as the input distribution to the reverse channel $Q_{X|Y}$), and $L\left[\hat{q}(Q)\right]$ is linear in $\tilde{Q}_Y$. So,

$$\min_{Q:\, Q_Y=\tilde{Q}_Y}\left\{I(Q) + L(\hat{q}(Q))\right\} = \min_{Q_{X|Y}}\left\{I(\tilde{Q}_Y\times Q_{X|Y}) + L\left[\hat{q}(\tilde{Q}_Y\times Q_{X|Y})\right]\right\} \tag{E.11}$$

is a pointwise minimum of concave functions in $\tilde{Q}_Y$ and thus a concave function. Moreover, it is symmetric in the sense that if $\tilde{Q}_Y(0)$ is replaced with $\tilde{Q}_Y(1)$, and $Q_{X|Y}(\cdot|0)$ is replaced with $Q_{X|Y}(\cdot|1)$, then the same value for the objective function is obtained. This fact along with convexity implies that the maximizing $\tilde{Q}_Y$ is uniform. Since P_X is also uniform, the minimizing $Q_{X|Y}$ is also symmetric.

We are now ready to provide the various bounds for detection of two BSCs under uniform input using the facts above.

A. Exact Random Coding Exponents

Let us begin with E_A of (47). Assume by contradiction that the optimal $\tilde{Q}^*$ is not symmetric. Fact 1 implies that if the inputs are permuted, $\tilde{Q}^*(\cdot|0) \leftrightarrow \tilde{Q}^*(\cdot|1)$ and this joint type is averaged with $\tilde{Q}^*$ with weight $\frac{1}{2}$ to result a new type $\tilde{Q}^{**}$ then

$$D(\tilde{Q}^{**}_{Y|X}||W|P_X) \leq D(\tilde{Q}^{*}_{Y|X}||W|P_X). \tag{E.12}$$

Also, Fact 2 implies that $\tilde{Q}^{**} \in \mathcal{J}_1$. In addition, since the function $J(Q) \triangleq -\alpha + f_V(\tilde{Q}) - f_W(Q)$ is linear in Q and depends on Q only via $\hat{q}(Q)$, then Remark 11 and Fact 3 above implies that $\tilde{Q}^{**} \in \mathcal{J}_2$. Consequently, the optimal $\tilde{Q}^*$ must be symmetric, and the minimization problem involved in computing E_A (47) may be reduced to optimizing only over crossover probabilities, rather than joint types. The result is as follows. Let $\gamma_{wv} \triangleq \log \frac{1-v}{1-w}$. Then,

$$\mathcal{J}_{1,\mathrm{B}} \triangleq \{\tilde{q}:\ f_{w,\mathrm{B}}(\tilde{q}) + \alpha - f_{v,\mathrm{B}}(\tilde{q}) \leq 0\} \tag{E.13}$$

$$= \{\tilde{q}:\ \tilde{q}(\rho_v - \rho_w) \leq -\alpha + \gamma_{wv}\} \tag{E.14}$$

and

$$\mathcal{J}_{2,\mathrm{B}} \triangleq \left\{\tilde{q}:\ \max_{0\leq\lambda\leq 1} \min_{0\leq q\leq 1} \{\log 2 - h_{\mathrm{B}}(q) + \lambda\left[-\alpha + f_{v,\mathrm{B}}(\tilde{q}) - f_{w,\mathrm{B}}(q)\right]\} > R\right\} \tag{E.15}$$

$$\stackrel{(a)}{=} \left\{\tilde{q}:\ \max_{0\leq\lambda\leq 1} \{\log 2 - h_{\mathrm{B}}(q^*) + \lambda\left[-\alpha + f_{v,\mathrm{B}}(\tilde{q}) - f_{w,\mathrm{B}}(q^*)\right]\} > R\right\} \tag{E.16}$$

where (a) is obtained by simple differentiation and $q^* = \frac{w^\lambda}{(1-w)^\lambda + w^\lambda}$. Then,

$$E_{A,\mathrm{B}} \triangleq \min_{\tilde{q}\in\cap_{i=1}^2 \mathcal{J}_{i,\mathrm{B}}} D_{\mathrm{B}}(\tilde{q}\|w). \tag{E.17}$$

Let us now inspect E_B of (52). The same reasoning as above shows that the optimal $(\tilde{Q}, \overline{Q})$ must be symmetric. Now, let

$$\mathcal{K}_{2,\mathrm{B}} \triangleq \{(\tilde{q}, \overline{q}):\ \overline{q}(\rho_v - \rho_w) \leq -\alpha + \gamma_{wv}\} \tag{E.18}$$

$$\mathcal{K}_{3,\mathrm{B}} \triangleq \left\{(\tilde{q}, \overline{q}):\ f_{v,\mathrm{B}}(\overline{q}) \geq \alpha + f_{w,\mathrm{B}}(\tilde{q}) - \left[R - \log 2 + h_{\mathrm{B}}(\overline{q})\right]_+\right\} \tag{E.19}$$

and

$$\mathcal{K}_{4,\mathrm{B}} \triangleq \left\{(\tilde{q},\overline{q}):\ \max_{0\le\lambda\le 1}\min_{0\le q\le 1}\left\{\log 2 - h_{\mathrm{B}}(q) + \lambda\left[-\alpha + f_{v,\mathrm{B}}(\overline{q}) - f_{w,\mathrm{B}}(q) + [R - \log 2 + h_{\mathrm{B}}(\overline{q})]_+\right]\right\} > R\right\} \tag{E.20}$$

$$= \left\{(\tilde{q},\overline{q}):\ \max_{0\le\lambda\le 1}\left\{\log 2 - h_{\mathrm{B}}(q^*) + \lambda\left[-\alpha + f_{v,\mathrm{B}}(\overline{q}) - f_{w,\mathrm{B}}(q^*) + [R - \log 2 + h_{\mathrm{B}}(\overline{q})]_+\right]\right\} > R\right\} \tag{E.21}$$

we obtain

$$E_{B,\mathrm{B}} \triangleq \min_{(\tilde{q},\overline{q})\in\cap_{i=2}^{4}\mathcal{K}_{i,\mathrm{B}}} D_{\mathrm{B}}(\tilde{q}\|w) + [\log 2 - h_{\mathrm{B}}(\overline{q}) - R]_+ . \tag{E.22}$$

The most difficult optimization problem to solve, namely $E_{B,\mathrm{B}}$, is only two-dimensional.

B. Expurgated Exponents

The Chernoff distance (119) for a pair of BSCs with crossover probabilities w and v is

$$d_s(x,\tilde{x}) = \begin{cases} -\log\left[(1-w)^s v^{1-s} + w^s(1-v)^{1-s}\right], & x \ne \tilde{x} \\ -\log\left[(1-w)^s(1-v)^{1-s} + w^s v^{1-s}\right], & x = \tilde{x} \end{cases}. \tag{E.23}$$

Now, let us analyze (121). Since P_X is uniform, then the definition of the set $\mathcal{L}$ in (120) implies that $P_{X\tilde{X}}$ is symmetric. So,

$$E_{\mathrm{TE}}^{\mathrm{EX}}(R,\alpha,P_X,W,V) = \max_{0\le s\le 1}\min_{q:\,\log 2 - h_{\mathrm{B}}(q)\le R}\{\alpha s + (1-q)d_s(1,0) + q d_s(0,0) + \log 2 - h_{\mathrm{B}}(q) - R\} \tag{E.24}$$

$$= \max_{0\le s\le 1}\{\alpha s + (1-q^*)d_s(1,0) + q^* d_s(0,0) + \log 2 - h_{\mathrm{B}}(q^*) - R\} \tag{E.25}$$

where

$$q^* = \frac{\exp\left[\frac{1}{\mu}\left(d_s(1,0) - d_s(0,0)\right)\right]}{1 + \exp\left[\frac{1}{\mu}\left(d_s(1,0) - d_s(0,0)\right)\right]} \tag{E.26}$$

and $\mu \ge 1$ is either chosen to satisfy $h_{\mathrm{B}}(q^*) = \log 2 - R$ or $\mu = 1$.

C. Exact Random Coding Exponents of Simplified Detectors/Decoders

As was previously mentioned, the simplified detector/decoder for high rates is useless in this case. For the simplified detector/decoder for low rates, we may use the same reasoning as for the optimal detector/decoder. Let $\mathcal{J}_{1,\mathrm{L,B}} \triangleq \mathcal{J}_{1,\mathrm{B}}$ and

$$\mathcal{J}_{2,\mathrm{L,B}} \triangleq \left\{\tilde{q}:\ \max_{\lambda\ge 0}\min_{0\le q\le 1}\{\log 2 - h_{\mathrm{B}}(q) + \lambda[-\alpha + f_{v,\mathrm{B}}(\tilde{q}) - f_{w,\mathrm{B}}(q)]\} > R\right\} \tag{E.27}$$

$$= \left\{\tilde{q}:\ \max_{\lambda\ge 0}\{\log 2 - h_{\mathrm{B}}(q^*) + \lambda[-\alpha + f_{v,\mathrm{B}}(\tilde{q}) - f_{w,\mathrm{B}}(q^*)]\} > R\right\} \tag{E.28}$$

where $q^* = \frac{w^\lambda}{(1-w)^\lambda + w^\lambda}$. Then,

$$E_{A,\mathrm{L,B}} \triangleq \min_{\tilde{q} \in \cap_{i=1}^{2} \mathcal{J}_{i,\mathrm{L,B}}} D_{\mathrm{B}}(\tilde{q} \| w). \tag{E.29}$$

Let $\mathcal{K}_{2,\mathrm{L,B}} \triangleq \mathcal{K}_{2,\mathrm{B}}$ and

$$\mathcal{K}_{3,\mathrm{L,B}} \triangleq \left\{ (\tilde{q}, \overline{q}) : \ f_{v,\mathrm{B}}(\overline{q}) \geq \alpha + f_{w,\mathrm{B}}(\tilde{q}) \right\}, \tag{E.30}$$

and

$$\mathcal{K}_{4,\mathrm{L,B}} \triangleq \left\{ (\tilde{q}, \overline{q}) : \ \max_{\lambda \geq 0} \min_{0 \leq q \leq 1} \left\{ \log 2 - h_{\mathrm{B}}(q) + \lambda \left[-\alpha + f_{v,\mathrm{B}}(\overline{q}) - f_{w,\mathrm{B}}(q) \right] \right\} > R \right\} \tag{E.31}$$

$$= \left\{ (\tilde{q}, \overline{q}) : \ \max_{\lambda \geq 0} \left\{ \log 2 - h_{\mathrm{B}}(q^*) + \lambda \left[-\alpha + f_{v,\mathrm{B}}(\overline{q}) - f_{w,\mathrm{B}}(q^*) \right] \right\} > R \right\}, \tag{E.32}$$

then

$$E_{B,\mathrm{L,B}} \triangleq \min_{(\tilde{q}, \overline{q}) \in \cap_{i=2}^{4} \mathcal{K}_{i,\mathrm{L,B}}} D_{\mathrm{B}}(\tilde{q} \| w) + \left[\log 2 - h_{\mathrm{B}}(\overline{q}) - R \right]_+ . \tag{E.33}$$

REFERENCES

[1] J. R. Barry, D. G. Messerschmitt, and E. A. Lee, *Digital communication: Third edition*. Norwell, MA, USA: Kluwer Academic Publishers, 2003.

[2] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, pp. 1897–1908, 1971.

[3] I. Nikiforov, "A generalized change detection problem," *Information Theory, IEEE Transactions on*, vol. 41, no. 1, pp. 171–187, Jan 1995.

[4] V. Chandar and A. Tchamkerten, "Quickest transient-change detection under a sampling constraint," *Submitted to Information Theory, IEEE Transactions on*, January 2015, available online: http://arxiv.org/pdf/1501.05930v2.pdf.

[5] D. N. C. T. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[6] N. Merhav, "Exact random coding error exponents of optimal bin index decoding," *Information Theory, IEEE Transactions on*, vol. 60, no. 10, pp. 6024–6031, October 2014.

[7] G. V. Moustakides, "Optimum joint detection and estimation," in *Proc. 2011 IEEE International Symposium on Information Theory*, July 2011, pp. 2984–2988.

[8] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *Information Theory, IEEE Transactions on*, vol. 58, no. 7, pp. 4215–4229, July 2012.

[9] N. Merhav, "Asymptotically optimal decision rules for joint detection and source coding," *Information Theory, IEEE Transactions on*, vol. 60, no. 11, pp. 6787–6795, Nov 2014.

[10] N. Weinberger and N. Merhav, "Codeword or noise? exact random coding exponents for joint detection and decoding," *Information Theory, IEEE Transactions on*, vol. 60, no. 9, pp. 5077–5094, Sept 2014.

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

[12] A. Tchamkerten, V. Chandar, and G. W. Wornell, "On the capacity region of asynchronous channels," in *Proc. 2008 IEEE International Symposium on Information Theory*, July 2008, pp. 1213–1217.

[13] ——, "Communication under strong asynchronism," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4508–4528, Oct 2009.

[14] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 1-2, pp. 1–212, 2009.

[15] A. Somekh-Baruch and N. Merhav, "Exact random coding exponents for erasure decoding," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6444–6454, 2011.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[17] G. D. Forney Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *Information Theory, IEEE Transactions on*, vol. 14, no. 2, pp. 206–220, 1968.

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

[19] N. Weinberger and N. Merhav, "Optimum trade-offs between the error exponent and the excess-rate exponent of variable-rate Slepian-Wolf coding," *Information Theory, IEEE Transactions on*, vol. 61, no. 4, pp. 2165–2190, April 2015, extended version available online: http://arxiv.org/pdf/1401.0892v3.pdf.

[20] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 471–480, 1973.

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

[22] N. Weinberger and N. Merhav, "Simplified erasure/list decoding," *Submitted to Information Theory, IEEE Transactions on*, December 2014, available online: http://arxiv.org/pdf/1412.1964v1.pdf.

[23] H. L. Van Trees and K. L. Bell, *Detection estimation and modulation theory, pt. I*. Wiley, 2013.

[24] S. Shamai and S. Verdu, "The empirical distribution of good codes," *Information Theory, IEEE Transactions on*, vol. 43, no. 3, pp. 836–846, May 1997.

[25] M. Sion, "On general minimax theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, pp. 171–176, 1958.

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

[27] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. ii," *Information and Control*, vol. 10, no. 5, pp. 522–552, 1967.

[28] N. Merhav, "On zero-rate error exponents of finite-state channels with input-dependent states," *Information Theory, IEEE Transactions on*, vol. 61, no. 2, pp. 741–750, February 2015.

[29] ——, "List decoding - random coding exponents and expurgated exponents," *Information Theory, IEEE Transactions on*, vol. 60, no. 11, pp. 6749–6759, Nov 2014.

[30] R. Blahut, "Hypothesis testing and information theory," *Information Theory, IEEE Transactions on*, vol. 20, no. 4, pp. 405–417, 1974.

[31] V. D. Goppa, "Nonprobabilistic mutual information without memory," *Probl. Contr. Information Theory*, vol. 4, pp. 97–102, 1975.

[32] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of discrete memoryless channels," in *Proc. of International Symposium on Information Theory*, 1977, p. 107 (abstract).

[33] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *Information Theory, IEEE Transactions on*, vol. 28, no. 3, pp. 430–443, May 1982.

[34] J. Ziv, "Universal decoding for finite-state channels," *Information Theory, IEEE Transactions on*, vol. 31, no. 4, pp. 453–460, July 1985.

[35] N. Merhav, "Universal decoding for memoryless gaussian channels with a deterministic interference," *Information Theory, IEEE Transactions on*, vol. 39, no. 4, pp. 1261–1269, July 1993.

[36] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," *Information Theory, IEEE Transactions on*, vol. 44, no. 5, pp. 1726–1745, September 1998.

[37] M. Feder and N. Merhav, "Universal composite hypothesis testing: a competitive minimax approach," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1504–1517, Jun 2002.

[38] N. Merhav and M. Feder, "Minimax universal decoding with an erasure option," *Information Theory, IEEE Transactions on*, vol. 53, no. 5, pp. 1664–1675, 2007.

[39] W. Huleihel, N. Weinberger, and N. Merhav, "Erasure/list random coding error exponents are not universally achievable," *Submitted to Information Theory, IEEE Transactions on*, October 2014, available online: http://arxiv.org/pdf/1410.7005v1.pdf.