Random Coding Error Exponents for the Two-User Interference Channel

Wasim Huleihel and Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering

Technion - Israel Institute of Technology

Haifa 3200003, ISRAEL

E-mail: {wh@campus, merhav@ee}.technion.ac.il

arXiv:1503.02389v2 [cs.IT] 20 Jun 2017

Abstract—This paper is about deriving lower bounds on the error exponents for the two-user interference channel under the random coding regime for several ensembles. Specifically, we first analyze the standard random coding ensemble, where the codebooks are comprised of independently and identically distributed (i.i.d.) codewords. For this ensemble, we focus on optimum decoding, which is in contrast to other, suboptimal decoding rules that have been used in the literature (e.g., joint typicality decoding, treating interference as noise, etc.). The fact that the interfering signal is a codeword, rather than an i.i.d. noise process, complicates the application of conventional techniques of performance analysis of the optimum decoder. Also, unfortunately, these conventional techniques result in loose bounds. Using analytical tools rooted in statistical physics, as well as advanced union bounds, we derive single-letter formulas for the random coding error exponents. We compare our results with the best known lower bound on the error exponent, and show that our exponents can be strictly better. Then, in the second part of this paper, we consider more complicated coding ensembles, and find a lower bound on the error exponent associated with the celebrated Han-Kobayashi (HK) random coding ensemble, which is based on superposition coding.

Index Terms—Random coding, error exponent, interference channels, superposition coding, Han-Kobayashi scheme, statistical physics, optimal decoding, multiuser communication.

I. INTRODUCTION

THE two-user interference channel (IFC) models a general scenario of communication between two transmitters and two receivers (with no cooperation at either side), where each receiver decodes its intended message from an observed signal, which is interfered by the other user, and corrupted by channel noise. The information-theoretic analysis of this model has begun over more than four decades ago and has recently witnessed a resurgence of interest. Most of the previous work on multiuser communication, and specifically on the IFC, has focused on obtaining inner and outer bounds to the capacity region (see, for example, [1, Ch. II.7]). In a nutshell, the study of this kind of channel started in [2] and continued in [3], where simple inner and outer bounds to the capacity region were given. Then, in [4], by using the well-known superposition coding technique, the inner bound of [3] was strictly improved. In [5], various inner and outer bounds were obtained by transforming the IFC model into

some multiple-access or broadcast channel. Unfortunately, the capacity region for the general interference channel is still unknown, although it has been solved for some special cases [6, 7]. The best known inner bound is the Han-Kobayashi (HK) region, established in [8], and which will also be considered in this paper.

To our knowledge, [9, 10] are the only previous works which treat the error exponents for the IFC under optimal decoding. Specifically, [9] derives lower bounds on error exponents of random codebooks comprised of i.i.d. codewords uniformly distributed over a given type class, under maximum likelihood (ML) decoding at each user, that is, optimal decoding. Contrary to the error exponent analysis of other multiuser communication systems, such as the multiple access channel [11], the difficulty in analyzing the error probability of the optimal decoder for the IFC is due to statistical dependencies induced by the interfering signal. Indeed, for the IFC, the marginal channel determining each receiver's ML decoding rule is induced also by the codebook of the interfering user. This extremely complicates the analysis, mostly because the interfering signal is a codeword and not an i.i.d. process. Another important observation, which was noticed in [9], is that the usual bounding techniques (e.g., Gallager's bounding technique) on the error probability fail to give tight results. To alleviate this problem, the authors of [9] combined some of the ideas from Gallager's bounding technique [12] to get an upper bound on the average probability of decoding error under ML decoding, the method of types [13], and used the method of type class enumerators, in the spirit of [14], which allows to avoid the use of Jensen's inequality in some steps.

The main purpose of this paper is to extend the study of achievability schemes to the more refined analysis of error exponents achieved by the two users, similarly as in [9]. Specifically, we derive single-letter expressions for the error exponents associated with the average error probability, for the finite-alphabet two-user IFC, under several random coding ensembles. The main contributions of this paper are as follows:

• Similarly as in recent works (see, e.g., [11, 15-18] and references therein) on the analysis of error exponents, we derive single-letter lower bounds for the random coding error exponents. For the standard random coding ensemble, considered in Subsection III-B, we analyze the optimal decoder for each receiver, which is interested solely in its intended message. This is in contrast to usual decoding techniques

analyzed for the IFC, in which each receiver decodes, in addition to its intended message, also part of (or all) the interfering codeword (that is, the other user's message), or other conventional achievability arguments [1, Ch. II.7], which are based on joint-typicality decoding, with restrictions on the decoder (such as, "treat interference as noise" or to "decode the interference"). This enables us to understand whether there is any significant degradation in performance due to the sub-optimality of the decoder. Also, since [9] analyzed the optimal decoder as well, we compare our formulas with those of [9], and show that our error exponent can be strictly better, which implies that the bounding technique in [9] is not tight. It is worthwhile to mention that the analytical formulas of our error exponents are simpler than the lower bound of [9].

• As was mentioned earlier, in [9] only random codebooks comprised of i.i.d. codewords (uniformly distributed over a type class) were considered. These ensembles are much simpler than the superposition codebooks of [8]. Unfortunately, it is very tedious to analyze superposition codebooks using the methods of [9]. In this paper, however, the new tools that we have derived enable us to analyze more involved random coding ensembles. Indeed, we can consider the coding ensemble used in the HK achievability scheme [8] and derive the respective error exponents. We also discuss an ensemble of hierarchical/tree codes [19].

• The analysis of the error exponents, carried out in this paper, turns out to be much more difficult than in previous works on point-to-point and multiuser communication problems, see, e.g., [11, 15-18]. Specifically, we encounter two main difficulties in our analysis: First, typically, when analyzing the probability of error, the first step is to apply the union bound. Usually, for point-to-point systems, under the random coding regime, the average error probability can be written as a union of pairwise independent error events. Accordingly, in this case, it is well known that the truncated union bound is exponentially tight [20, Lemma A.2]. This is no longer the case, however, when considering multiuser systems, and in particular, the IFC. For the IFC, the events comprising the union are strongly dependent, especially due to the fact that we are considering the optimal decoder. To alleviate this difficulty, following the ideas of [11], we derived new upper bounds on the probability of a union of events, which take into account the dependencies among the events. The second difficulty that we have encountered in our analysis is that in contrast to previous works, applying the type class enumerator method [14] is not simple, due to the reason mentioned above. Using some methods from large deviations theory, we were able to tackle this difficulty.

• Recently, in [21, 22], the authors independently suggested lower bounds on the error exponents of both standard and cognitive multiple-access channels (MACs), assuming suboptimal successive decoding scheme, and using the standard random coding ensemble (considered in Subsection III-B). Although the motivation in [21] is different, the codebook construction and the decoding rule are the same as in the first part of this paper, and thus, essentially, their results apply also for the IFC. It is important to emphasize that while we believe that our error exponent analysis is somewhat simpler, at least

conceptually, there is strong resemblance between our analysis and [21], as they both based on type enumeration techniques. Note, however, that while in [21] the standard union bound was used, here, the new upper bounds mentioned above, provide some potential gain over [21], even for the ordinary ensemble. Also, as was mentioned above, we consider also the more complicated ensemble pertaining to the HK scheme. The derivation of the lower bound on the error exponent of this ensemble is built upon the derivation of the lower bound on the error exponent of the standard random coding ensemble, and thus it makes useful and convenient to start with the analysis of the latter ensemble. We emphasize that the extension of [21] to the HK ensemble is non-trivial. Finally, we mention that the focus in [21] was on achievable rate region, rather than error exponents, and thus no comparison to [9] was provided.

• We believe that by using the techniques and tools derived in this paper, other multiuser systems, such as the IFC with mismatched decoding, the MAC [11], the broadcast channel, the relay channel, etc., and accordingly, other coding schemes, such as binning [15], and hierarchical codes [19], can be analyzed.

The paper is organized as follows. In Section II, we establish notation conventions. In Section III, we formalize the problem and assert the main theorems. Specifically, in Subsections III-B and III-C, we give the resulting error exponents under the standard random coding ensemble and the HK coding ensemble, respectively. Finally, Section IV is devoted to the proofs of our main results.

II. NOTATION CONVENTIONS

Throughout this paper, scalar random variables (RVs) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters, e.g. X, x, and $\mathcal{X}$, respectively. A similar convention will apply to random vectors of dimension n and their sample values, which will be denoted with the same symbols in the boldface font. We also use the notation X_i^j $(j > i)$ to designate the sequence of RVs $(X_i, X_{i+1}, \ldots, X_j)$. The set of all n-vectors with components taking values in a certain finite alphabet, will be denoted by the same alphabet superscripted by n, e.g., $\mathcal{X}^n$. Generic channels will be usually denoted by the letters P, Q, or W. We shall mainly consider joint distributions of two RVs (X, Y) over the Cartesian product of two finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. For brevity, we will denote any joint distribution, e.g. Q_{XY}, simply by Q, the marginals will be denoted by Q_X and Q_Y, and the conditional distributions will be denoted by $Q_{X|Y}$ and $Q_{Y|X}$. The joint distribution induced by Q_X and $Q_{Y|X}$ will be denoted by $Q_X \times Q_{Y|X}$, and a similar notation will be used when the roles of X and Y are switched.

The expectation operator will be denoted by $\mathbb{E}\{\cdot\}$, and when we wish to make the dependence on the underlying distribution Q clear, we denote it by $\mathbb{E}_Q\{\cdot\}$. Information measures induced by the generic joint distribution Q_{XY}, will be subscripted by Q, for example, $I_Q(X;Y)$ will denote the corresponding mutual information, etc. The divergence (or, Kullback-Liebler distance) between two probability measures Q and P will be

denoted by $D(Q||P)$. The weighted divergence between two channels, $Q_{Y|X}$ and $P_{Y|X}$, with weight P_X, is defined as

$$D(Q_{Y|X}||P_{Y|X}|P_X) \triangleq \sum_{x \in \mathcal{X}} P_X(x)$$
$$\sum_{y \in \mathcal{Y}} Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{P_{Y|X}(y|x)}. \quad (1)$$

For a given vector $\boldsymbol{x}$, let $\hat{Q}_{\boldsymbol{x}}$ denote the empirical distribution, that is, the vector $\{\hat{Q}_{\boldsymbol{x}}(x), \ x \in \mathcal{X}\}$, where $\hat{Q}_{\boldsymbol{x}}(x)$ is the relative frequency of the letter x in the vector $\boldsymbol{x}$. Let $T(P_X)$ denote the type class associated with P_X, that is, the set of all sequences $\boldsymbol{x}$ for which $\hat{Q}_{\boldsymbol{x}} = P_X$. Similarly, for a pair of vectors $(\boldsymbol{x}, \boldsymbol{y})$, the empirical joint distribution will be denoted by $\hat{Q}_{\boldsymbol{xy}}$, or simply by $\hat{Q}$, for short. All previously defined notation rules for regular distributions will also be used for empirical distributions.

The cardinality of a finite set $\mathcal{A}$ will be denoted by $|\mathcal{A}|$, its complement will be denoted by $\mathcal{A}^c$. The probability of an event $\mathcal{E}$ will be denoted by $\Pr\{\mathcal{E}\}$. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathcal{I}\{\mathcal{E}\}$. For two sequences of positive numbers, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ means that $\{a_n\}$ and $\{b_n\}$ are of the same exponential order, i.e., $n^{-1} \log a_n/b_n \to 0$ as $n \to \infty$, where logarithms are defined with respect to (w.r.t.) the natural basis, that is, $\log(\cdot) = \ln(\cdot)$. Finally, for a real number x, we denote $[x]_+ \triangleq \max\{0, x\}$.

III. PROBLEM FORMULATION AND MAIN RESULTS

We divide this section into three subsections. In the first, we present the model and formulate the problem. In the second, we present a lower bound on the IFC error exponent, assuming a simple random coding ensemble where random codebooks comprised of i.i.d. codewords are uniformly distributed over a type class. It is well-known [10] that this coding scheme can be improved by using superposition coding and introducing the notion of "private" and "common" messages (to be defined in the sequel). Accordingly, in the third subsection, we consider the HK coding scheme [8], and derive lower bounds on the error exponents. Finally, we discuss other ensembles that can be analyzed using the same methods.

A. The IFC Model

Consider a two-user interference channel of two senders, two receivers, and a discrete memoryless channel (DMC), defined by a set of single-letter transition probabilities, $W_{Y_1 Y_2 | X_1 X_2}(y_1, y_2 | x_1, x_2)$, with finite input alphabets, $\mathcal{X}_1, \mathcal{X}_2$, and finite output alphabets, $\mathcal{Y}_1, \mathcal{Y}_2$. Here, each sender, $k \in \{1, 2\}$, communicates an independent message $m_k \in \{1, 2, \ldots, M_k \triangleq 2^{nR_k}\}$ at rate R_k, and each receiver, $l \in \{1, 2\}$, decodes its respective message. Specifically, a $(2^{nR_1}, 2^{nR_2}, n)$ code $\mathcal{C}_n$ consists of:
• Two message sets $\mathcal{M}_1 \triangleq \{0, \ldots, 2^{nR_1} - 1\}$ and $\mathcal{M}_2 \triangleq \{0, \ldots, 2^{nR_2} - 1\}$ for the first and second users, respectively.
• Two encoders, where for each $k \in \{1, 2\}$, the k-th encoder assigns a codeword $\boldsymbol{x}_{k,i}$ to each message $i \in \mathcal{M}_k$.
• Two decoders, where each decoder $l \in \{1, 2\}$ assigns an estimate $\hat{m}_l$ to m_l.

We assume that the message pair (m_1, m_2) is uniformly distributed over $\mathcal{M}_1 \times \mathcal{M}_2$. It is clear that the *optimal decoder* of the first user, for this problem, is given by

$$\hat{m}_1 = \arg \max_{i \in \mathcal{M}_1} P(\boldsymbol{y}_1 | \boldsymbol{x}_{1,i}) \quad (2)$$

$$= \arg \max_{i \in \mathcal{M}_1} \frac{1}{M_2} \sum_{j=1}^{M_2-1} P(\boldsymbol{y}_1 | \boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j}) \quad (3)$$

where $P(\boldsymbol{y}_1 | \boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j})$ is the marginal channel defined as

$$P(\boldsymbol{y}_1 | \boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j}) \triangleq \prod_{k=1}^{n} W_{Y_1 | X_1 X_2}(y_{1k} | x_{1,i,k}, x_{2,j,k}), \quad (4)$$

and

$$W_{Y_1 | X_1 X_2}(y_{1,k} | x_{1,i,k}, x_{2,j,k})$$
$$\triangleq \sum_{y_{2,k} \in \mathcal{Y}_2} W_{Y_1 Y_2 | X_1 X_2}(y_{1,k}, y_{2,k} | x_{1,i,k}, x_{2,j,k}). \quad (5)$$

The optimal decoder of the second user is defined similarly. Since there is no cooperation between the two receivers, the error probabilities for the code $\mathcal{C}_n$, are defined as

$$P_{e,i}(\mathcal{C}_n) \triangleq 2^{-n(R_1+R_2)}$$
$$\cdot \sum_{\tilde{m}_1, \tilde{m}_2} \Pr\{\hat{m}_i(Y_i^n) \neq \tilde{m}_i | m_1 = \tilde{m}_1, m_2 = \tilde{m}_2\}, \quad (6)$$

for $i = 1, 2$.

B. The Ordinary Random Coding Ensemble

In this subsection, we consider the ordinary random coding ensemble: For each $k \in \{1, 2\}$, we select independently M_k codewords $\{\boldsymbol{x}_{k,i}\}$, for $i \in \mathcal{M}_k$, under the uniform distribution across the type class $T(P_{X_k})$, for a given distribution P_{X_k} on $\mathcal{X}_k$. Our goal is to assess the exponential rate of $\bar{P}_{e,1}^{(n)} \triangleq \mathbb{E}\{P_{e,1}(\mathcal{C}_n)\}$, where the average is over the code ensemble, that is,

$$E_1^*(R_1, R_2) \triangleq \liminf_{n \to \infty} -\frac{1}{n} \log \bar{P}_{e,1}^{(n)}, \quad (7)$$

and similarly for the second user. Before stating the main result, we define some quantities. Given a joint distribution $Q_{X_1 X_2 Y_1}$ over $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_1$, consider the definitions in (8), shown at the top of the next page. We devote Appendix C for a discussion on aspects of the computation of (8j). We have the following result.

Theorem 1 Let R_1 and R_2 be given, and let $E^*(R_1, R_2)$ be defined as in (7). Consider the ensemble of fixed composition codes of types P_{X_1} and P_{X_2}, for the first and second users, respectively. For a discrete memoryless two-user IFC, we have

$$E_1^*(R_1, R_2) \geq \tilde{E}_1(R_1, R_2), \quad (9)$$

for any $R_1, R_2 \geq 0$.

Several remarks on Theorem 1 are in order.
• Due to symmetry, the error exponent for the second user, that is, $\tilde{E}_2(R_1, R_2)$ is simply obtained from Theorem 1 by

$$f\left(Q_{X_1 X_2 Y_1}\right) \triangleq \mathbb{E}_Q\left[\log W_{Y_1|X_1 X_2}(Y_1|X_1, X_2)\right], \tag{8a}$$

$$t_0(Q_{X_1 Y_1}) \triangleq R_2 + \max_{\substack{\hat{Q}:\ \hat{Q}_{X_2}=P_{X_2},\ \hat{Q}_{X_1 Y_1}=Q_{X_1 Y_1} \\ I_{\hat{Q}}(X_2;X_1,Y_1)\leq R_2}} \left[f(\hat{Q}) - I_{\hat{Q}}(X_2; X_1, Y_1)\right], \tag{8b}$$

$$E_1(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1}) \triangleq \min_{\substack{\hat{Q}:\ \hat{Q}_{X_2}=P_{X_2},\ \hat{Q}_{X_1 Y_1}=\tilde{Q}_{X_1 Y_1} \\ \hat{Q}\in\mathcal{L}(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1})}} \left[I_{\hat{Q}}(X_2; X_1, Y_1) - R_2\right]_+, \tag{8c}$$

$$E_2(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1}) \triangleq \min_{\substack{\hat{Q}:\ \hat{Q}_{X_2}=P_{X_2},\ \hat{Q}_{X_1 Y_1}=\tilde{Q}_{X_1 Y_1} \\ \hat{Q}\in\hat{\mathcal{L}}(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1})}} \left[I_{\hat{Q}}(X_2; Y_1) - R_2\right]_+, \tag{8d}$$

$$\mathcal{L}(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1}) \triangleq \left\{\hat{Q}:\ \max\left[t_0(Q_{X_1 X_2 Y_1}), f(Q_{X_1 X_2 Y_1})\right] \right.$$
$$\left. \leq \max\left[f(\tilde{Q}_{X_1 X_2 Y_1}), f(\hat{Q}) + \left[R_2 - I_{\hat{Q}}(X_2; X_1, Y_1)\right]_+\right]\right\}, \tag{8e}$$

$$\hat{\mathcal{L}}(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1}) \triangleq \left\{\hat{Q}:\ \max\left[t_0(Q_{X_1 X_2 Y_1}), f(Q_{X_1 X_2 Y_1})\right]\right.$$
$$\left. \leq \max\left[f(\tilde{Q}_{X_1 X_2 Y_1}), f(\hat{Q}) + \left[R_2 - I_{\hat{Q}}(X_2; Y_1)\right]_+\right]\right\}, \tag{8f}$$

$$\hat{E}_1(Q_{X_1 X_2 Y_1}, R_2) \triangleq \min_{\tilde{Q}:\ \tilde{Q}_{X_1}=P_{X_1},\ \tilde{Q}_{X_2 Y_1}=Q_{X_2 Y_1}} \left[I_{\tilde{Q}}(X_1; X_2, Y_1) + E_1(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1})\right], \tag{8g}$$

$$\hat{E}_2(Q_{X_1 X_2 Y_1}, R_2) \triangleq \min_{\tilde{Q}:\ \tilde{Q}_{X_1}=P_{X_1},\ \tilde{Q}_{X_2 Y_1}=Q_{X_2 Y_1}} E_2(\tilde{Q}_{X_1 X_2 Y_1}, Q_{X_1 X_2 Y_1}), \tag{8h}$$

$$E(Q_{X_1 X_2 Y_1}, R_1, R_2) \triangleq \max\left\{\left[\hat{E}_1(Q_{X_1 X_2 Y_1}, R_2) - R_1\right]_+, \hat{E}_2(Q_{X_1 X_2 Y_1}, R_2)\right\}, \tag{8i}$$

$$\tilde{E}_1(R_1, R_2) \triangleq \min_{\substack{Q_{Y_1|X_1 X_2}: \\ Q_{X_1}=P_{X_1}, Q_{X_2}=P_{X_2}}} \left[D(Q_{Y_1|X_1 X_2}||W_{Y_1|X_1 X_2}|P_{X_1}\times P_{X_2}) + E(Q_{X_1 X_2 Y_1}, R_1, R_2)\right]. \tag{8j}$$

swapping the roles of X_1, Y_1, and R_1, with those of X_2, Y_2, and R_2, respectively.

• An immediate byproduct of Theorem 1 is finding the set of rates (R_1, R_2) for which $\tilde{E}_1(R_1, R_2) > 0$, namely, the rates for which the probability of error vanishes exponentially as $n \to \infty$. We show in Appendix D, that this set is given by:

$$\mathcal{R}_{\text{ordinary},1} = \{R_1 < I(X_1; Y_1)\} \cup$$
$$\{\{R_1 + R_2 < I(X_1, X_2; Y_1)\} \cap \{R_1 < I(X_1; Y_1|X_2)\}\} \tag{10}$$

evaluated with $P_{X_1 X_2 Y_1} = P_{X_1} \times P_{X_2} \times W_{Y_1|X_1 X_2}$. Note that this region can be obtained also by using standard typicality-based achievability arguments (see, e.g., [23]). Fig. 1 demonstrates a qualitative description of this region. The interpretation is as follows: The corner point $(I(X_1; Y_1|X_2), I(X_2; Y_1))$ is achieved by first decoding the interference (the second user), canceling it, and then decoding the first user. The sum-rate constraint can be achieved by joint decoding the two users (similarly to MAC), and thus, obviously, also by our optimal decoder. Finally, the region $R_1 < I(X_1; Y_1)$ and $R_2 \geq I(X_2; Y_1|X_1)$ means that we decode the first user while treating the interference as noise. Evidently, from the perspective of the first decoder, which is interested only in the message transmitted by the first sender, the second sender can use any rate, and thus there is no bound on R_2 whenever $R_1 < I(X_1; Y_1)$. Now, it was shown in [10]
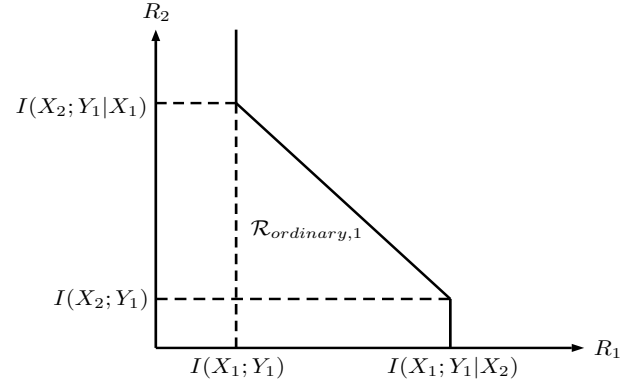

Fig. 1. Rate region $\mathcal{R}_{\text{ordinary},1}$ for which $\tilde{E}_1(R_1, R_2) > 0$.

that the error exponent achievable for the first user under the ordinary random coding regime is zero outside the closure of $\mathcal{R}_{\text{ordinary},1}$. Whence, this fact and the above conclusion, characterize the rate region where the attainable exponent with ordinary random coding is positive. Notice that $\mathcal{R}_{\text{ordinary},1}$ is well-known to be contained in the HK region [10, 23].

• *Existence of a single code:* our result holds true on the average, where the averaging is done over the random choice of codebooks. It can be shown (see, for example, [24, p. 2924]) that there exists deterministic sequence of fixed composition codebooks of increasing block length n for which the same

asymptotic error performance can be achieved for *both* users simultaneously.

• *About the proof:* it is instructive to discuss (in some more detail than earlier) one of the main difficulties in proving Theorem 1, which is customary to multiuser systems, such as the IFC. Without loss of generality, we assume throughout, that the transmitted codewords are $\boldsymbol{x}_{1,0}$ and $\boldsymbol{x}_{2,0}$. Accordingly, the average probability of error associated with the decoder (3) is given by (11), shown at the top of the next page, where $\mathcal{F}_0 \triangleq (\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,0}, \boldsymbol{Y}_1)$. In contrast to previous works, applying the type class enumerator[1] method [14], is not a simple task. Since we are interested in the optimal decoder, each event of the union in (11), depends on the whole codebook of the second user. One may speculate that this problem can be tackled by conditioning on the codebook of the second user, and then (12). However, the cost of this conditioning is a very complicated (if not intractable) large deviations analysis of some quantities. The consequence of this situation is that in order to analyze the probability of error, it is required to analyze the joint distribution of type class enumerators, and not just rely on their marginal distributions, as is usually done, e.g., [15-18].

Another difficulty is handling the union in (11). By the union bound and Shulman's inequality [20, Lemma A.2], we know that for a sequence of pairwise independent events, $\{\mathcal{A}_i\}_{i=1}^N$, the following holds

$$\frac{1}{2}\min\left\{1, \sum_{i=1}^N \Pr\{\mathcal{A}_i\}\right\} \leq \Pr\left\{\bigcup_{i=1}^N \mathcal{A}_i\right\}$$
$$\leq \min\left\{1, \sum_{i=1}^N \Pr\{\mathcal{A}_i\}\right\}, \quad (12)$$

which is a useful result when assessing the exponential behavior of such probabilities. Equation (12) is one of the building blocks of tight exponential analysis of previously considered point-to-point systems (see, e.g., [15-18], and many references therein). However, in our case the various events are not pairwise independent, and therefore this result cannot be applied. To alleviate this problem, following the techniques of [11], we derive new upper bounds on the probability of a union of events, which takes into account such dependencies among the events.

• As was mentioned in the Introduction, in [21], lower bounds on the error exponents of both standard and cognitive multiple-access channels (MACs) were suggested. Although the motivation in [21] is different, their results apply also for the IFC. Now, while in [21] the standard truncated union bound was used, here our new upper bound on the probability of a union of events, provides some potential gain over [21]. Specifically, the lower bound in [21] is the same as

(8j) but without the $\hat{E}_2(Q, R_2)$ term, i.e., it is given by $\min_Q \left\{D(Q\|W) + \left[\hat{E}_1(Q, R_2) - R_1\right]_+\right\}$, and thus, in general, our result may be tighter. It should be stressed, however, that we have not identified specific examples where the new term, namely, $\hat{E}_2(Q, R_2)$, dominated the maximum in (8i).

• The lower bound in [9] is extremely complicated, and it is very difficult to compare it analytically to the lower bound in Theorem 1. Nonetheless, we can still claim (in general) that our lower bound is at least as good as the lower bound in [9]. Indeed, the first step in the analysis of the error exponent in both our paper and in [9] is applying the union bound (actually, here, we employ a tighter union bound). However, it will be seen that every other passage in our analysis is exponentially *exact*, while in [9], some steps are associated with inequalities that may cause gaps in the exponential scale, and thus in general, $\tilde{E}(R_1, R_2) \geq E_{[9]}(R_1, R_2)$, for any $(R_1, R_2) \in \mathbb{R}_+^2$, where $E_{[9]}(R_1, R_2)$ is the lower bound in [9].

• *Comparison with [9]*: Similarly to [9], we present results for the following channel: $Y_1 = X_1 \cdot X_2 \oplus Z$ and $Y_2 = X_2$, where $X_1, X_2, Y_1, Y_2 \in \{0, 1\}$, $Z \sim \text{Bern}(p)$, "·" is multiplication, and "⊕" is modulo-2 addition. In the numerical calculations, we fix $p = 0.01$. Fig. 2 presents the lower bound on the error exponent under optimal decoding, derived in this paper, compared to the lower bound $E_{[9]}(R_1, R_2)$ of [9], as a function of R_1, for different values of P_{X_1}, P_{X_2}, and R_2. It can be seen that our exponents are strictly better than those of [9].

C. The Han-Kobayashi Coding Scheme

Consider the channel model of Subsection III-B. The best known inner bound on the capacity region is achieved by the HK coding scheme [8]. The idea of this scheme is to split the message m_1 into "private" and "common" messages, m_{11} and m_{12} at rates R_{11} and R_{12}, respectively, such that $R_1 = R_{11} + R_{12}$. Similarly, m_2 is split into m_{21} and m_{22} at rates R_{21} and R_{22}, with $R_2 = R_{21} + R_{22}$. The intuition behind this splitting is based on the receiver behavior at low and high signal-to-noise-ratio (SNR). Specifically, it is well-known [1] that: (1) when the SNR is low, treating the interference as noise is an optimal strategy, and (2) when the SNR is high, decoding and then canceling the interference is the optimal strategy. Accordingly, the above splitting captures the general intermediate situation where the first decoder, for example, is interested only in partial information from the second user, in addition to its own intended message.

Next, we describe explicitly the coding strategy of [8]. Fix a distribution $P_{Z_{11}} P_{Z_{12}} P_{Z_{21}} P_{Z_{22}} P_{X_1|Z_{11}Z_{12}} P_{X_2|Z_{21}Z_{22}}$, where the latter two conditional distributions represent deterministic mappings. For each $k, k' \in \{1, 2\}$, randomly and conditionally independently generate a sequence $\boldsymbol{z}_{k,k'}(m_{k,k'})$ under the uniform distribution across the type class $T(P_{Z_{kk'}})$ for a given $P_{Z_{k,k'}}$. To communicate a message pair (m_{11}, m_{12}), sender 1 transmits $\boldsymbol{x}_1(\boldsymbol{z}_{11}, \boldsymbol{z}_{12})$, and analogously for sender 2. All our results can be extended to the setting in which the codewords are generated conditionally on a time-sharing sequence $\boldsymbol{q}$. However, this leads to more complicated notation. Thus, we focus primarily on the case without time-sharing.

Let us now describe the operation of each receiver. Receiver $k = 1, 2$, recovers its intended message m_k and the common

---

[1]For a given $y^n \in \mathcal{Y}^n$, and a given joint probability distribution Q_{XY} on $\mathcal{X} \times \mathcal{Y}$, the *type class enumerator*, $N(Q_{XY})$, is the number of codewords $\{x_i^n\}$ in $\mathcal{C}_n$ whose conditional empirical joint distribution with y^n is Q_{XY}, namely, $N(Q_{XY}) = \left|x^n \in \mathcal{C}_n : \hat{Q}_{x^n y^n} = Q_{XY}\right|$, where $\hat{Q}_{x^n y^n}$ is the empirical joint distribution of x^n and y^n, and $|\mathcal{A}|$ designates the cardinality of a finite set $\mathcal{A}$. Type class enumeration method refers to the process of converting a sum of exponentially many terms (usually likelihood functions) into polynomial number of type class enumerators, which are easier to analyze.

$$\bar{P}_{e,1}^{(n)} = \Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\right]$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\middle|\mathcal{F}_0\right]\right\} \tag{11}$$
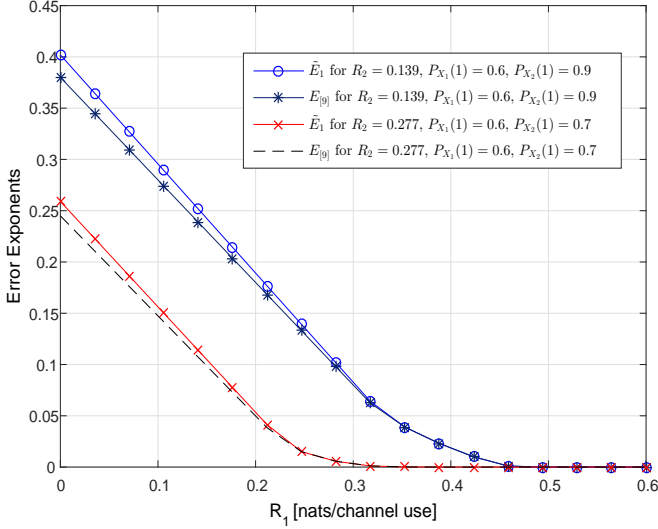

Fig. 2. Comparison between $\tilde{E}_1(R_1, R_2)$ and $E_{[9]}(R_1, R_2)$ of [9], as a function of R_1 for two different values of R_2 and fixed choices of P_{X_1} and P_{X_2}.

Fig. 3. Han-Kobayashi coding scheme.

$$\arg\max_{(i,j,k)\in\mathcal{M}_{11}\times\mathcal{M}_{12}\times\mathcal{M}_{21}}P\left(\boldsymbol{y}_1|\boldsymbol{z}_{11,i},\boldsymbol{z}_{12,j},\boldsymbol{z}_{21,k}\right) \tag{13}$$

$$= \arg\max_{(i,j,k)\in\mathcal{M}_{11}\times\mathcal{M}_{12}\times\mathcal{M}_{21}}$$

$$\frac{1}{M_{22}}\sum_{l=0}^{M_{22}-1}P\left(\boldsymbol{y}_1|\boldsymbol{z}_{11,i},\boldsymbol{z}_{12,j},\boldsymbol{z}_{21,k},\boldsymbol{z}_{22,l}\right). \tag{14}$$

Accordingly, the probability of error for the code $\mathcal{C}_n$ and for the first user, is defined as

$$P_{e,1}\left(\mathcal{C}_n\right) \triangleq \Pr\left\{(\hat{m}_{11}, \hat{m}_{12}) \neq (m_{11}, m_{12})\right\}, \tag{15}$$

and similarly for the second user. Our goal is to assess the exponential rate of $\bar{P}_{e,1}^{(n)} \triangleq \mathbb{E}\left\{P_{e,1}\left(\mathcal{C}_n\right)\right\}$, where the average is over the code ensemble, namely,

$$E_{\mathrm{HK}}^*(R_1, R_2) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log\bar{P}_{e,1}^{(n)}, \tag{16}$$

and similarly for the second user. In order to facilitate the presentation of the following result, we move the technical definitions to Appendix A. Our second main result is the following.

message from the other sender (although it is not required to). This scheme is illustrated in Fig. 3. Note that this decoding operation is the one that was used in [8], but there, the sub-optimal non-unique simultaneous joint typical decoder [1, Ch. II.7] was used. Here, by contrast, we use sub-optimal ML decoding (the sub-optimality is due to the fact that our decoder recovers also the common message from the other sender). As will be explained in the sequel, analyzing the optimal ML decoder is a challenging task, and therefore we will focus on sub-optimal ML decoding.

We wish to find a lower bound on the error exponent, achieved by the HK encoding functions, in conjunction with the above described decoding functions. To this end, note that by combining the channel and the deterministic mappings as indicated by the dashed box in Fig. 3, the channel $(Z_{11}, Z_{12}, Z_{21}, Z_{22}) \mapsto (Y_1, Y_2)$ is just a four-sender, two-receiver, DMC interference channel, with virtual inputs. Note that this formulation induces the Markovian structure $(Z_{11}, Z_{12}, Z_{21}, Z_{22}) \ominus (X_1, X_2) \ominus (Y_1, Y_2)$, where the (virtual) input distributions, i.e., $P_{Z_{k,k'}}$ for $k, k' \in \{1, 2\}$, can be optimized. We assume that the message quadruple $(m_{11}, m_{12}, m_{21}, m_{22})$ is uniformly distributed over $\mathcal{M}_{11} \times \mathcal{M}_{12} \times \mathcal{M}_{21} \times \mathcal{M}_{22}$. Following the above descriptions, our decoder for this problem is given by

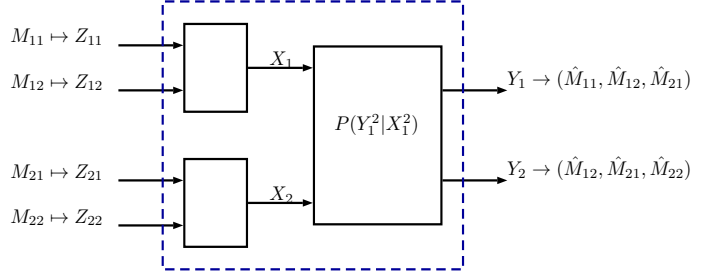$$(\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{21}) =$$

Theorem 2 Let $E_{\mathrm{HK}}^*(R_1, R_2)$ be defined as in (16). Consider the HK encoding scheme described above. For a discrete memoryless two-user IFC, we have:

$$E_{\mathrm{HK}}^*(R_1, R_2) \geq \max_{\substack{(R_{11},R_{12},R_{21},R_{22}):\\ R_{11}+R_{12}=R_1\\ R_{21}+R_{22}=R_2}} \tilde{E}_{\mathrm{HK}}(R_{11}, R_{12}, R_{21}, R_{22}),$$

$$\tag{17}$$

for any $R_1, R_2 \geq 0$, where $\tilde{E}_{\mathrm{HK}}(R_{11}, R_{12}, R_{21}, R_{22})$ is given in (A.30).

Several remarks on Theorem 2 are in order.

• As before, an immediate byproduct of Theorem 2 is finding the set of rates $(R_{11}, R_{12}, R_{21}, R_{22})$ for which

$\tilde{E}_{\text{HK}}(R_{11}, R_{12}, R_{21}, R_{22}) > 0$, namely, for which the probability of error vanishes exponentially as $n \to \infty$. It can be shown that this set is given by the HK region, that is,

$$R_{11} \leq I(Z_1; Y_1 | Z_2, Z_3), \qquad (18a)$$

$$R_{12} \leq I(Z_2; Y_1 | Z_1, Z_3), \qquad (18b)$$

$$R_{21} \leq I(Z_3; Y_1 | Z_1, Z_2), \qquad (18c)$$

$$R_{11} + R_{12} \leq I(Z_1, Z_2; Y_1 | Z_3), \qquad (18d)$$

$$R_{11} + R_{21} \leq I(Z_1, Z_3; Y_1 | Z_2), \qquad (18e)$$

$$R_{12} + R_{21} \leq I(Z_2, Z_3; Y_1 | Z_1), \qquad (18f)$$

$$R_{11} + R_{12} + R_{21} \leq I(Z_1, Z_2, Z_3; Y_1), \qquad (18g)$$

evaluated with

$$P_{Z_1^4 Y_1} = P_{Z_1} P_{Z_2} P_{Z_3} P_{Z_4} P_{X_1 | Z_1 Z_2} P_{X_2 | Z_3 Z_4} W_{Y_1 | X_1 X_2}$$

and similarly for the second user, where $P_{X_1 | Z_1 Z_2}$ and $P_{X_2 | Z_3 Z_4}$ represent deterministic mappings. As was mentioned earlier, it is possible to introduce a time-sharing sequence $\boldsymbol{q}$, and accordingly, (18) remains almost the same, but with a time-sharing RV Q (with alphabet size bounded by eight [8]), appearing at the conditioning of each the above mutual information terms. Finally, we mention that in [23] it was shown that by using the optimal ML decoder (given in (20)) instead of the non-unique simultaneous joint typical decoder [8], we cannot improve the achievable region. This observation do not for the error exponent.

• It can be shown[2] that the error exponent in Theorem 2 is no worse than the error exponent in Theorem 1, namely, $\tilde{E}_{\text{HK}}(R_{11}, R_{12}, R_{21}, R_{22}) \geq \tilde{E}_1(R_1, R_2)$ for any (R_1, R_2) such that $R_1 = R_{11} + R_{12}$ and $R_2 = R_{21} + R_{22}$. Moreover, it is well-known that upon optimizing the auxiliary RVs, $\{Z_{ij}\}$, the HK region in (18) is *strictly* better than $\mathcal{R}_{\text{ordinary},1}$. Therefore, this necessarily implies that for a certain region of high rates, the HK error exponent in Theorem 2 will be positive while the standard random coding error exponent in Theorem 1 will be zero. On the other extreme, it is easy to show that for $(R_1, R_2) = (0, 0)$ the error exponent in Theorem 2 equals to the error exponent in Theorem 1, so for small rates there is no improvement in the error exponents.

• Contrary to the ordinary ensemble, described in Subsection III-B, the HK ensemble depends on some auxiliary RVs which should be optimized. For a give pair of rates (R_1, R_2), our error exponent formula provides a criterion for the choice of the optimal auxiliary RVs: maximize the lower bound on the error exponent in Theorem 2, w.r.t. the auxiliaries, subject to some relevant constraints. As a matter of fact, for a *given* pair of rates (R_1, R_2), it is very likely that the optimal choice of these auxiliaries will be different from the optimal choice for the same pair in the achievable region. Indeed, even in the single-user case, the capacity achieving distribution is usually different from the optimal distribution in the error exponent sense.

---

[2]By definition, the ordinary ensemble is a simple instance of the HK ensemble, and thus the latter is indeed better upon optimization of the auxiliary RVs $\{Z_{ij}\}$. To see that the ordinary ensemble is a special case of the HK ensemble, we take $Z_{11} = X_1$, $Z_{12} = Z_{21} = \emptyset$, and $Z_{22} = X_2$.

• Using the same techniques and tools derived in this paper, we can consider other random coding ensembles. For example, we can analyze the error exponents resulting from the *hierarchical code ensemble*. Specifically, in this ensemble, the message m_1 is split into common and private messages m_{11}, m_{12} at rates R_{11} and R_{12}, respectively, such that $R_1 = R_{11} + R_{12}$. Similarly m_2 is split into m_{21}, m_{22} at rates R_{21} and R_{22}, respectively, such that $R_2 = R_{21} + R_{22}$. Then, we first randomly draw a rate R_{11} codebook of block length n according to a given distribution. Then, for each such codeword, we randomly and conditionally independently generate a rate R_{12} codebook of block length n. In other words, the code has a tree structure with two levels, where the first serves for "cloud centers", and the second for the "satellites". We do the same for the second user. Under this ensemble, we can analyze the optimal decoder. Note, however, that this ensemble is different from the product ensemble considered in Theorem 2. Indeed, while for the former for each first stage codeword (cloud center) we independently draw a new codebook (satellites), for the latter, for each cloud center we have the same satellite codebook. Loosely speaking, this means that the product ensemble is "less random". From the point of view of achievable region, however, the hierarchical ensemble is equivalent to the product ensemble used in HK scheme [1, Ch. II.7]. Nonetheless, the error exponents associated with these ensembles could be different.

• In Theorem 2 we assumed the sub-optimal decoder given in (14). Indeed, the optimal decoder for our problem is given by:

$$(\hat{m}_{11}, \hat{m}_{12}) = \arg \max_{i,j} P\left(\boldsymbol{y}_1 | \boldsymbol{z}_{11,i}, \boldsymbol{z}_{12,j}\right) \qquad (19)$$

$$= \arg \max_{i,j}$$

$$\frac{1}{M_{21} M_{22}} \sum_{k=0}^{M_{21}-1} \sum_{l=0}^{M_{22}-1} P\left(\boldsymbol{y}_1 | \boldsymbol{z}_{11,i}, \boldsymbol{z}_{12,j}, \boldsymbol{z}_{21,k}, \boldsymbol{z}_{22,l}\right). \quad (20)$$

Unfortunately, it turns out that analyzing the HK scheme (in conjunction with (20)) is much more difficult, and requires some more delicate tools from large deviations theory. Specifically, the main difficulty in the derivations, is to analyze the large deviations behavior of a two-dimensional sum (due to the double summation in (20)) involving binomial RVs which are strongly dependent (contrary to the standard one-dimensional version, see, e.g., [15, p. 6027-6028]). Nonetheless, we note that for the hierarchical code ensemble described above, the optimal decoder can be analyzed. Indeed, for this ensemble, it is clear that the optimal decoder is given by

$$(\hat{m}_{11}, \hat{m}_{12}) = \arg \max_{i,j} P\left(\boldsymbol{y}_1 | \boldsymbol{x}_1(i, j)\right) \qquad (21)$$

$$= \arg \max_{i,j} \frac{1}{M_{21} M_{22}} \sum_{k=0}^{M_{21}-1} \sum_{l=0}^{M_{22}-1} P\left(\boldsymbol{y}_1 | \boldsymbol{x}_1(i, j), \boldsymbol{x}_2(k, l)\right)$$

$$\qquad (22)$$

where $\boldsymbol{x}_1(i, j) \triangleq f_1(\boldsymbol{x}_1'(i), \boldsymbol{x}_1''(i, j))$ and $\boldsymbol{x}_2(i, j) \triangleq f_2(\boldsymbol{x}_2'(i), \boldsymbol{x}_2''(i, j))$ due to the hierarchical structure. Now, while here too, we will deal with two-dimensional summation, the summands will be independent, given the cloud centers codebook, and the proof can be carried out smoothly.

IV. PROOFS OF MAIN RESULTS

A. Proof of Theorem 1

Without loss of generality, we assume throughout, that the transmitted codewords are $\boldsymbol{x}_{1,0}$ and $\boldsymbol{x}_{2,0}$, and due to the fact that we analyze the first decoder, for convenience, we use $\boldsymbol{y}$ instead of $\boldsymbol{y}_1$. Accordingly, the average probability of error associated with the optimal decoder (3), is given by (24), shown at the top of the next page, where $\mathcal{F}_0 \triangleq (\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,0}, \boldsymbol{Y})$. In the following, we propose new upper bound on the probability of a union of events, which are suitable for some structured dependency between the events, as above.

In order to give some motivation for this new bound, we first rewrite (23) in another (equivalent) form. Specifically, we express (24) in terms of the joint types of $(\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,0}, \boldsymbol{Y})$ and $\{(\boldsymbol{Y}, \boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j})\}_{i,j}$. First, for a given joint distribution $Q_{X_1 X_2 Y}$ of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$, let

$$f(Q_{X_1 X_2 Y}) \triangleq \frac{1}{n} \log P(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \tag{25}$$

$$= \mathbb{E}_Q \left[\log W_{Y|X_1 X_2}(Y|X_1 X_2) \right]. \tag{26}$$

For a given joint type $Q_{X_{1,0} X_{2,0} Y}$ of the random vectors $(\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,0}, \boldsymbol{Y})$, define the set $T_I\left(Q_{X_{1,0} X_{2,0} Y}\right)$, given in (27), shown at the top of the next page. The set $T_I(Q_{X_{1,0} X_{2,0} Y})$ is the set of all possible types of $(\boldsymbol{X}_{1,i}, \mathcal{C}_2)$, where $\mathcal{C}_2$ denotes the codebook of the second user, which lead to a decoding error when $(\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,0}, \boldsymbol{Y}) \in T(Q_{X_{1,0} X_{2,0} Y})$ is transmitted. The various marginal constraints in (28) and (29) arise from the fact that we are assuming constant-composition random coding and, of course, fixed marginals due to the given fixed joint distribution $Q_{X_{1,0} X_{2,0} Y}$. Finally, the constraint

$$e^{nf(\tilde{Q}^0_{X_{1,0} X_{2,0} Y})} + \sum_{k=1}^{M_2-1} \left[e^{nf(\tilde{Q}^k_{X_1 X_2 Y})} - e^{nf(\hat{Q}^k_{X_1 X_2 Y})} \right]$$
$$\geq e^{nf(Q_{X_{1,0} X_{2,0} Y})} \tag{30}$$

in (27), represents a decoding error event, that is, it holds if and only if

$$\sum_{j=0}^{M_2-1} P(\boldsymbol{y}|\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j}) \geq \sum_{j=0}^{M_2-1} P(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}), \tag{31}$$

or, equivalently,

$$P(\boldsymbol{y}|\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,0}) + \sum_{j=1}^{M_2-1} \left[P(\boldsymbol{y}|\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j}) - P(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}) \right]$$
$$\geq P(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}), \tag{32}$$

for $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(Q_{X_{1,0} X_{2,0} Y})$, $(\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1 X_2 Y})$, $(\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1 X_2 Y})$, and $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_1 X_2 Y})$, for $j = 1, 2, \ldots, M_2 - 1$. Now, with these definitions, fixing $Q_{X_{1,0} X_{2,0} Y}$, and letting $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}, \boldsymbol{y})$ be an arbitrary triplet of sequences such that $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(Q_{X_{1,0} X_{2,0} Y})$, it follows, by definition, that the error event

$$\bigcup_{i=1}^{M_1-1} \left\{ \sum_{j=0}^{M_2-1} P(\boldsymbol{Y}|\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j}) \geq \sum_{j=0}^{M_2-1} P(\boldsymbol{Y}|\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,j}) \right\} \tag{33}$$

can be rewritten, in terms of types, as follows

$$\bigcup_{i=1}^{M_1-1} \bigcup_{\left\{ \tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y} \right\}_j \in T_I(Q_{X_{1,0} X_{2,0} Y})} \left\{ \begin{array}{c} (\boldsymbol{X}_{1,i}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1 X_2 Y}), \\ \left\{ (\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1 X_2 Y}) \right\}_{j=1}^{M_2-1}, \\ \left\{ (\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0} X_2 Y}) \right\}_{j=1}^{M_2-1} \end{array} \right\}. \tag{34}$$

We wish to analyze the probability of the event in (34), conditioned on $\mathcal{F}_0$. Note that the inner union in (34) is over vectors of types (an exponential number of them). Finally, for the sake of convenience, we simplify the notations of (34), and write it equivalently as

$$\bigcup_{i=1}^{M_1-1} \bigcup_{\boldsymbol{l}} \left\{ \begin{array}{c} \boldsymbol{X}_{1,i} \in \mathcal{A}_{\boldsymbol{l},0}, \\ (\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j}) \in \mathcal{A}_{\boldsymbol{l},j}, \text{ for } j = 1, \ldots, M_2 - 1, \\ \boldsymbol{X}_{2,j} \in \tilde{\mathcal{A}}_{\boldsymbol{l},j}, \text{ for } j = 1, \ldots, M_2 - 1 \end{array} \right\} \tag{35}$$

where, again, the index "$\boldsymbol{l}$" in the inner union runs over the combinations of types (namely, $\boldsymbol{l} = \{\tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y}\}_j$) that belong to $T_I(Q_{X_{1,0} X_{2,0} Y})$, and the various sets $\{\mathcal{A}_{\boldsymbol{l},j}, \tilde{\mathcal{A}}_{\boldsymbol{l},j}\}_{\boldsymbol{l},j}$ correspond to the typical sets in (34) (recall that $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}, \boldsymbol{y})$ are given at this stage). Next, following the ideas of [11], we provide a new upper bound on a generic probability which has the form of (35). The proof of this lemma is relegated to Appendix B.

Lemma 1 Let $\{V_1(i)\}_{i=1}^{L_1}, V_2, V_3, \ldots, V_K$ be independent sequences of independently and identically distributed (i.i.d.) RVs on the alphabets $\mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K$, respectively, with $V_1(i) \sim P_{V_1}, V_2 \sim P_{V_2}, \ldots, V_K \sim P_{V_K}$. Fix a sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^N, \{\mathcal{A}_{i,2}\}_{i=1}^N, \ldots, \{\mathcal{A}_{i,K-1}\}_{i=1}^N$, where $\mathcal{A}_{i,j} \subseteq \mathcal{V}_1 \times \mathcal{V}_{j+1}$, for $1 \leq j \leq K-1$ and for all $1 \leq i \leq N$. Also, fix a set $\{\mathcal{A}_{i,0}\}_{i=1}^N$ where $\mathcal{A}_{i,0} \subseteq \mathcal{V}_1$ for all $1 \leq i \leq N$, and another sequence of sets $\{\mathcal{G}_{i,2}\}_{i=1}^N, \{\mathcal{G}_{i,3}\}_{i=1}^N, \ldots, \{\mathcal{G}_{i,K}\}_{i=1}^N$, where $\mathcal{G}_{i,j} \subseteq \mathcal{V}_j$, for $2 \leq j \leq K$ and for all $1 \leq i \leq N$. Define

$$\mathcal{B}_{m,1} \triangleq \left\{ v_1 : v_1 \in \mathcal{A}_{m,0}, \bigcap_{j=1}^{K-1} (v_1, v_{j+1}) \in \mathcal{A}_{m,j}, \right. $$
$$\left. \bigcap_{j=2}^{K} v_j \in \mathcal{G}_{m,j} \text{ for some } \{v_j\}_{j=2}^K \right\}, \tag{36}$$

and

$$\mathcal{B}_{m,2} \triangleq \left\{ \{v_j\}_{j=2}^K : v_1 \in \mathcal{A}_{m,0}, \bigcap_{j=1}^{K-1} (v_1, v_{j+1}) \in \mathcal{A}_{m,j}, \right. $$
$$\left. \bigcap_{j=2}^{K} v_j \in \mathcal{G}_{m,j} \text{ for some } v_1 \right\}, \tag{37}$$

for $m = 1, 2, \ldots, N$. Then, a general upper bound is given in (38), shown at the top of the next page, with $(V_1, \ldots, V_K) \sim P_{V_1} \cdots \times P_{V_K}$.

$$\bar{P}_{e,1}^{(n)} = \Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\right] \tag{23}$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\,\middle|\,\mathcal{F}_0\right]\right\} \tag{24}$$

$$T_I\left(Q_{X_{1,0}X_{2,0}Y}\right) \triangleq \left\{\tilde{Q}_{X_1X_{2,0}Y}^0 \in \mathcal{S}_0, \left(\left\{\tilde{Q}_{X_1X_2Y}^k\right\}_{k=1}^{M_2-1}, \left\{\hat{Q}_{X_{1,0}X_2Y}^k\right\}_{k=1}^{M_2-1}\right) \in \mathcal{S}_1 : \right.$$

$$\left. e^{nf(\tilde{Q}_{X_1X_{2,0}Y}^0)} + \sum_{k=1}^{M_2-1}\left[e^{nf(\tilde{Q}_{X_1X_2Y}^k)} - e^{nf(\hat{Q}_{X_1X_2Y}^k)}\right] \geq e^{nf(Q_{X_{1,0}X_{2,0}Y})}\right\}, \tag{27}$$

$$\mathcal{S}_0(Q_{X_{1,0}X_{2,0}Y}) \triangleq \left\{\tilde{Q}_{X_1X_{2,0}Y}^0 : \tilde{Q}_{X_1}^0 = P_{X_1}, \tilde{Q}_{X_2}^0 = P_{X_2}, \tilde{Q}_{X_{2,0}Y}^0 = Q_{X_{2,0}Y}\right\}, \tag{28}$$

$$\mathcal{S}_1(Q_{X_{1,0}X_{2,0}Y}) \triangleq \left\{\left\{\tilde{Q}_{X_1X_2Y}^k\right\}_{k=1}^{M_2-1}, \left\{\hat{Q}_{X_{1,0}X_2Y}^k\right\}_{k=1}^{M_2-1} : \tilde{Q}_{X_1}^k = P_{X_1}, \tilde{Q}_{X_2}^k = P_{X_2}, \tilde{Q}_Y^k = Q_Y, \right.$$

$$\hat{Q}_{X_{1,0}}^k = P_{X_1}, \hat{Q}_{X_2}^k = P_{X_2}, \hat{Q}_{X_{1,0}Y}^k = Q_{X_{1,0}Y},\ \forall 1 \leq k \leq M_2-1$$

$$\left. \tilde{Q}_{X_2Y}^k = \hat{Q}_{X_2Y}^k, \tilde{Q}_{X_1Y}^k = \tilde{Q}_{X_1Y}^m, \forall k, m\right\}. \tag{29}$$

$$\Pr\left\{\bigcup_i\left\{\bigcup_{m=1}^{N}\left\{V_1(i) \in \mathcal{A}_{m,0}, \bigcap_{k=1}^{K-1}(V_1(i),V_{k+1}) \in \mathcal{A}_{m,k}, \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{m,k}\right\}\right\}\right\}$$

$$\leq \min\left\{1, L_1\Pr\left\{\bigcup_{m=1}^{N}\{V_1 \in \mathcal{B}_{m,1}\}\right\}, \Pr\left\{\bigcup_{m=1}^{N}\left\{\{V_j\}_{k=2}^K \in \mathcal{B}_{m,2}\right\}\right\},\right.$$

$$\left. L_1\Pr\left\{\bigcup_{m=1}^{N}\left\{V_1 \in \mathcal{A}_{m,0}, \bigcap_{k=1}^{K-1}(V_1,V_{k+1}) \in \mathcal{A}_{m,k}, \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{m,k}\right\}\right\}\right\} \tag{38}$$

Next, we apply Lemma 1 to the problem at hand. To this end, we choose the following parameters in accordance to the notations used in Lemma 1. Recall that we deal with

$$\bigcup_{i=1}^{M_1-1}\bigcup_{\boldsymbol{l}}\left\{\begin{array}{c}\boldsymbol{X}_{1,i} \in \mathcal{A}_{\boldsymbol{l},0}, \\ (\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}) \in \mathcal{A}_{\boldsymbol{l},j},\ \text{for } j=1,\ldots,M_2-1, \\ \boldsymbol{X}_{2,j} \in \tilde{\mathcal{A}}_{\boldsymbol{l},j},\quad \text{for } j=1,\ldots,M_2-1\end{array}\right\} \tag{39}$$

and in Lemma 1 we have considered:

$$\bigcup_{i=1}^{L_1}\bigcup_{m=1}^{N}\left\{\begin{array}{c}V_1(i) \in \mathcal{A}_{m,0}, \\ (V_1(i),V_{j+1}) \in \mathcal{A}_{m,j},\ \text{for } j=1,\ldots,K-1, \\ V_2 \in \mathcal{G}_{m,2},\ldots,V_K \in \mathcal{G}_{m,K}\end{array}\right\}. \tag{40}$$

Thus, comparing (39) and (40), we readily notice the following parallels:

• The numbers of events in the unions over i is $L_1 = M_1-1$. Also, we have $K = M_2$ independent random vectors $V_1(i) = \boldsymbol{X}_{1,i}$ and $V_{l+1} = \boldsymbol{X}_{2,l}$, for $1 \leq i \leq M_1-1$ and $1 \leq l \leq M_2-1$.

• The union over m corresponds to a union over $\boldsymbol{l}$, which as was mentioned before, is actually a union over a vector of types. Accordingly, we have:

1) $\mathcal{A}_{m,i} = \mathcal{A}_{\boldsymbol{l},i}$, for $0 \leq i \leq M_2-1$,
2) $\mathcal{G}_{m,i} = \tilde{\mathcal{A}}_{\boldsymbol{l},i-1}$, for $2 \leq i \leq M_2$.

These sets correspond to each of the typical sets $T(\tilde{Q}_{X_1X_{2,0}Y}^0)$, $\{T(\tilde{Q}_{X_1X_2Y}^k)\}_{k=1}^{M_2-1}$, and $\{T(\hat{Q}_{X_{1,0}X_2Y}^k)\}_{k=1}^{M_2-1}$.

• According to (36) and (37) we need to define $\mathcal{B}_{m,1} = \mathcal{B}_1(\tilde{Q}_{X_1X_{2,0}Y}^0, \{\tilde{Q}_{X_1X_2Y}^j, \hat{Q}_{X_1X_2Y}^j\}_j)$ and $\mathcal{B}_{m,2} = \mathcal{B}_2(\tilde{Q}_{X_1X_{2,0}Y}^0, \{\tilde{Q}_{X_1X_2Y}^j, \hat{Q}_{X_1X_2Y}^j\}_j)$. Using (36) and (37), we get (41) and (42), given at the top of the next page.

Thus, invoking Lemma 1, we have (44), where each of the probabilities at the r.h.s. of (44) are conditioned on $\mathcal{F}_0$. Therefore, we were able to simplify the problematic union over the codebook of the first user. Note, however, that we cannot (directly) apply here the method of types due to the fact that the union is over an exponential number of types, and thus a more refined analysis is needed. We start by analyzing the last term at the r.h.s. of (44). To this end, we will invoke the type enumeration method, but first, the main observation here is that similarly to the passage from (33) to (34), the last term at the r.h.s. of (44) can be rewritten as (46), shown at the

$$\mathcal{B}_{m,1} = \left\{ \boldsymbol{x}_1 : \begin{array}{c} (\boldsymbol{x}_1, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1 X_{2,0} Y}), \\ \left\{ (\boldsymbol{x}_1, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1 X_2 Y}) \right\}_{j=1}^{M_2-1}, \\ \left\{ (\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0} X_2 Y}) \right\}_{j=1}^{M_2-1} \text{ for some } \{\boldsymbol{x}_{2,j}\}_j \end{array} \right\} \tag{41}$$

$$\mathcal{B}_{m,2} = \left\{ \{\boldsymbol{x}_{2,j}\}_{j\geq 1} : \begin{array}{c} (\boldsymbol{x}_1, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1 X_{2,0} Y}), \\ \left\{ (\boldsymbol{x}_1, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1 X_2 Y}) \right\}_{j=1}^{M_2-1}, \\ \left\{ (\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0} X_2 Y}) \right\}_{j=1}^{M_2-1} \text{ for some } \boldsymbol{x}_1 \end{array} \right\} \tag{42}$$

$$\tilde{P}_{e,1}^{(n)} \triangleq \Pr \left[\bigcup_{i=1}^{M_1-1} \left\{ \sum_{j=0}^{M_2-1} P(\boldsymbol{Y}|\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j}) \geq \sum_{j=0}^{M_2-1} P(\boldsymbol{Y}|\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,j}) \right\} \Bigg| \mathcal{F}_0 \right] \tag{43}$$

$$\leq \min \left\{ 1, M_1 \cdot \Pr \left[\bigcup_{\left\{ \tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y} \right\}_j \in T_I(Q_{X_{1,0} X_{2,0} Y})} \boldsymbol{X}_{1,1} \in \mathcal{B}_1 \left(\tilde{Q}^0_{X_1 X_{2,0} Y}, (\tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y})_j) \right) \right], \right.$$

$$\Pr \left[\bigcup_{\left\{ \tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y} \right\}_j \in T_I(Q_{X_{1,0} X_{2,0} Y})} \{\boldsymbol{X}_{2,j}\}_{j\geq 1} \in \mathcal{B}_2 \left(\tilde{Q}^0_{X_1 X_{2,0} Y}, (\tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y})_j) \right) \right],$$

$$\left. M_1 \cdot \Pr \left[\bigcup_{\left\{ \tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y} \right\}_j \in T_I(Q_{X_{1,0} X_{2,0} Y})} \left\{ \begin{array}{c} (\boldsymbol{X}_{1,1}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1 X_{2,0} Y}), \\ \left\{ (\boldsymbol{X}_{1,1}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1 X_2 Y}) \right\}_{j=1}^{M_2-1}, \\ \left\{ (\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0} X_2 Y}) \right\}_{j=1}^{M_2-1} \end{array} \right\} \right] \right\} \tag{44}$$

next page. That is, we returned back to the structure of the original probability in (24), but now, without the union over the codebook of the first user. Note that the conditioning on the random vector $\boldsymbol{X}_{1,1}$ in (46), is due to the fact that $\boldsymbol{X}_{1,1}$ is common to all the summands in the inner summation over the codebook of the second user. We next evaluate the exponential behavior of the probability in (46). For a given realization of $\boldsymbol{Y} = \boldsymbol{y}$, $\boldsymbol{X}_{1,0} = \boldsymbol{x}_{1,0}$, $\boldsymbol{X}_{1,1} = \boldsymbol{x}_{1,1}$, and $\boldsymbol{X}_{2,0} = \boldsymbol{x}_{2,0}$, let us define

$$s \triangleq \frac{1}{n} \log P(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}), \tag{47}$$

and

$$r \triangleq \frac{1}{n} \log P(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{x}_{2,0}). \tag{48}$$

For a given $(\boldsymbol{y}, \boldsymbol{x}_{1,0}, \boldsymbol{x}_{1,1}, \boldsymbol{x}_{2,0})$, and a given joint probability distribution $Q_{X_1 X_2 Y}$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$, let $N_1(Q_{X_1 X_2 Y})$ designate the number of codewords $\{\boldsymbol{X}_{2,j}\}_j$ (excluding $\boldsymbol{x}_{2,0}$) whose conditional empirical distribution with $\boldsymbol{y}$ and $\boldsymbol{x}_{1,1}$ is $Q_{X_1 X_2 Y}$, that is,

$$N_1(Q_{X_1 X_2 Y}) \triangleq \sum_{j=1}^{M_2-1} \mathcal{I} \{ (\boldsymbol{x}_{1,1}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(Q_{X_1 X_2 Y}) \}, \tag{49}$$

and let $N_2(Q_{X_1 X_2 Y})$ designate the number of codewords $\{\boldsymbol{X}_{2,j}\}_j$ (excluding $\boldsymbol{x}_{2,0}$) whose conditional empirical distribution with $\boldsymbol{y}$ and $\boldsymbol{x}_{1,0}$ is $Q_{X_1 X_2 Y}$, that is

$$N_2(Q_{X_1 X_2 Y}) \triangleq \sum_{j=1}^{M_2-1} \mathcal{I} \{ (\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(Q_{X_1 X_2 Y}) \}. \tag{50}$$

Also, recall that

$$f(Q_{X_1 X_2 Y}) = \frac{1}{n} \log P(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \tag{51}$$

$$= \mathbb{E}_Q \left[\log W_{Y|X_1 X_2}(Y|X_1, X_2) \right] \tag{52}$$

where $Q_{X_1 X_2 Y}$ is understood to be the joint empirical distribution of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n$. Thus, in terms of the above notations, we may write:

$$\sum_{j=0}^{M_2-1} P(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{X}_{2,j}) = e^{nr}$$

$$+ \sum_{Q_{X_2|X_1 Y} \in \mathcal{S}(Q_{X_1 Y})} N_1(Q_{X_1 X_2 Y}) e^{nf(Q_{X_1 X_2 Y})} \tag{53}$$

$$\triangleq e^{nr} + \mathcal{N}_1(Q_{X_1 Y}). \tag{54}$$

where for a given $Q_{X_1 Y}$, $\mathcal{S}(Q_{X_1 Y})$ is defined as the set of all distributions $\{Q_{X_2|X_1 Y}\}$, such that

$$\Pr\left[\bigcup_{\left\{\tilde{Q}_{X_1X_2Y}^j,\hat{Q}_{X_1X_2Y}^j\right\}_j\in T_I(Q_{X_{1,0}X_{2,0}Y})}\left\{\begin{array}{l}(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}_{X_1X_{2,0}Y}^0),\\[2mm]\left\{(\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\tilde{Q}_{X_1X_2Y}^j)\right\}_{j=1}^{M_2-1},\\[2mm]\left\{(\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\hat{Q}_{X_{1,0}X_2Y}^j)\right\}_{j=1}^{M_2-1}\end{array}\right\}\right]$$

$$=\Pr\left[\left\{\sum_{j=0}^{M_2-1}P(\boldsymbol{Y}|\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j})\geq\sum_{j=0}^{M_2-1}P(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j})\right\}\Bigg|\mathcal{F}_0\right]\tag{45}$$

$$=\mathbb{E}\left\{\Pr\left[\left\{\sum_{j=0}^{M_2-1}P(\boldsymbol{Y}|\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j})\geq\sum_{j=0}^{M_2-1}P(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j})\right\}\Bigg|\mathcal{F}_0,\boldsymbol{X}_{1,1}\right]\Bigg|\mathcal{F}_0\right\}.\tag{46}$$

$\sum_{(x_1,y)\in\mathcal{X}_1\times\mathcal{Y}}Q_{X_1Y}(x_1,y)Q_{X_2|X_1Y}(x_2|x_1,y)=P_{X_2}(x_2)$ for all $x_2\in\mathcal{X}_2$, namely,

$$\mathcal{S}(Q_{X_1Y})=\left\{Q_{X_1X_2Y}':Q_{X_1Y}'=Q_{X_1Y},\ Q_{X_2}'=P_{X_2}\right\}.\tag{55}$$

Similarly,

$$\sum_{j=0}^{M_2-1}P(\boldsymbol{y}|\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j})=e^{ns}$$

$$+\sum_{Q_{X_2|X_{1,0}Y}\in\mathcal{S}(Q_{X_{1,0}Y})}N_2\left(Q_{X_{1,0}X_2Y}\right)e^{nf(Q_{X_{1,0}X_2Y})}\tag{56}$$

$$\triangleq e^{ns}+\mathcal{N}_2(Q_{X_{1,0}Y}).\tag{57}$$

where for a given $Q_{X_{1,0}Y}$, $\mathcal{S}(Q_{X_{1,0}Y})$ is defined as the set of all distributions $\{Q_{X_2|X_{1,0}Y}\}$, such that $\sum_{(x_1,y)\in\mathcal{X}_1\times\mathcal{Y}}Q_{X_{1,0}Y}(x_1,y)Q_{X_2|X_{1,0}Y}(x_2|x_1,y)=P_{X_2}(x_2)$ for all $x_2\in\mathcal{X}_2$ (similarly as in (55)). For simplicity of notation, in the following, we use Q and $\tilde{Q}$ to denote $Q_{X_1X_2Y}$ and $Q_{X_{1,0}X_2Y}$, respectively. Therefore, with these definitions in mind, we wish to calculate (given $(\mathcal{F}_0,\boldsymbol{X}_{1,1})$)

$$\Pr\left[\sum_{j=0}^{M_2-1}P(\boldsymbol{Y}|\boldsymbol{x}_{1,1},\boldsymbol{X}_{2,j})\geq\sum_{j=0}^{M_2-1}P(\boldsymbol{Y}|\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j})\right]$$

$$=\Pr\left[\mathcal{N}_1(Q_{X_1Y})-\mathcal{N}_2(Q_{X_{1,0}Y})\geq e^{ns}-e^{nr}\right]\tag{58}$$

where s, r, $\mathcal{N}_1(Q)$ and $\mathcal{N}_2(Q)$ are given in (47), (48), (54), and (57), respectively. Let $\varepsilon>0$ be arbitrarily small, and define $i_1\triangleq\left\lfloor\frac{1}{n\epsilon}\log P(\boldsymbol{y}|\boldsymbol{x}_{1,0},\boldsymbol{x}_{2,0})\right\rfloor$. Then,

$$\Pr\left[\mathcal{N}_1(Q_{X_1Y})-\mathcal{N}_2(Q_{X_{1,0}Y})\geq e^{ns}-e^{nr}\right]$$

$$=\sum_{i=i_1}^{\lceil R_2/\varepsilon\rceil}\Pr\left\{e^{ni\varepsilon}\leq\mathcal{N}_2(Q_{X_{1,0}Y})\leq e^{n(i+1)\varepsilon},\right.$$
$$\left.\mathcal{N}_1(Q_{X_1Y})-\mathcal{N}_2(Q_{X_{1,0}Y})\geq e^{ns}-e^{nr}\right\}$$

$$\leq\sum_{i=i_1}^{\lceil R_2/\varepsilon\rceil}\Pr\left\{e^{ni\varepsilon}\leq\mathcal{N}_2(Q_{X_{1,0}Y})\leq e^{n(i+1)\varepsilon},\right.$$
$$\left.\mathcal{N}_1(Q_{X_1Y})\geq e^{ni\varepsilon}+e^{ns}-e^{nr}\right\}\tag{59}$$

$$=\sum_{i=i_1}^{\lceil R_2/\varepsilon\rceil}\Pr\left\{e^{ni\varepsilon}\leq\mathcal{N}_2(Q_{X_{1,0}Y})\leq e^{n(i+1)\varepsilon}\right\}$$

$$\times\Pr\left\{\mathcal{N}_1(Q_{X_1Y})\geq e^{ni\varepsilon}+e^{ns}-e^{nr}\right.$$
$$\left.|\ e^{ni\varepsilon}\leq\mathcal{N}_2(Q_{X_{1,0}Y})\leq e^{n(i+1)\varepsilon}\right\}.\tag{60}$$

It is not difficult to show that (see, e.g., [15, p. 6028])

$$\Pr\left\{e^{nt}\leq\mathcal{N}_2(Q_{X_{1,0}Y})\leq e^{n(t+\varepsilon)}\right\}$$
$$\doteq\begin{cases}0&t<t_0(Q_{X_{1,0}Y})-\varepsilon\\1&t_0(Q_{X_{1,0}Y})-\varepsilon\leq t\leq t_0(Q_{X_{1,0}Y})\\\exp\left[-nE(t,Q_{X_{1,0}Y})\right]&t>t_0(Q_{X_{1,0}Y})\end{cases}\tag{61}$$

where

$$t_0(Q_{X_{1,0}Y})\triangleq R_2+$$
$$\max_{\tilde{Q}\in\mathcal{S}(Q_{X_{1,0}Y}):\ I_{\tilde{Q}}(X_2;X_{1,0},Y)\leq R_2}\left[f(\tilde{Q})-I_{\tilde{Q}}(X_2;X_{1,0},Y)\right],\tag{62}$$

in which $\mathcal{S}(Q)$ is defined in (55), $f(Q)$ is given in (52), and

$$E(t,Q_{X_{1,0}Y})\triangleq\min\left\{\left[I_{\tilde{Q}}(X_2;X_{1,0},Y)-R_2\right]_+:\right.$$
$$\left.f(\tilde{Q})+\left[R_2-I_{\tilde{Q}}(X_2;X_{1,0},Y)\right]_+\geq t\right\}.\tag{63}$$

Now, in the exponential scale, the term at the r.h.s. of (60) is dominated by one of the summands, and we claim that the dominant contribution to the sum over i is due to the first term[3], $i=t_0(Q_{X_{1,0}Y})/\varepsilon$. Indeed, let $\mathscr{A}_k\triangleq\left\{e^{nk\varepsilon}\leq\mathcal{N}_2(Q_{X_{1,0}Y})\leq e^{n(k+1)\varepsilon}\right\}$ and $\mathscr{B}_k\triangleq\left\{\mathcal{N}_1(Q_{X_1Y})\geq e^{nk\varepsilon}+e^{ns}-e^{nr}\right\}$, and notice that the summands in (60) correspond to $\Pr\{\mathscr{A}_k\cap\mathscr{B}_k\}$. According to (61), $\Pr\{\mathscr{A}_{t_0}\}\to1$ (the exponent $E(k\varepsilon,Q_{X_{1,0}Y})$ vanishes), and note that $\Pr\{\mathscr{B}_k\}$ is monotonically decreasing with k. Therefore,

$$\Pr\{\mathscr{A}_{t_0}\cap\mathscr{B}_{t_0}\}\leq\max_{k\geq t_0}\Pr\{\mathscr{A}_k\cap\mathscr{B}_k\}$$
$$\leq\max_{k\geq t_0}\Pr\{\mathscr{B}_k\}=\Pr\{\mathscr{B}_{t_0}\}.\tag{64}$$

[3]Note that according to (61), $\Pr\{e^{ni\varepsilon}\leq\mathcal{N}_2(Q_{X_{1,0}Y})\leq e^{n(i+1)\varepsilon}\}$ vanishes (in the exponential scale) for $i<t_0(Q_{X_{1,0}Y})/\varepsilon$. Thus, to asses the exponential scale of (60) we consider only the indices correspond to $i\geq t_0(Q_{X_{1,0}Y})/\varepsilon$.

On the other hand,

$$\Pr\left\{\mathscr{A}_{t_0} \cap \mathscr{B}_{t_0}\right\} = \Pr\left\{\mathscr{B}_{t_0}\right\} - \Pr\left\{\mathscr{A}_{t_0}^c \cap \mathscr{B}_{t_0}\right\}$$
$$\geq \Pr\left\{\mathscr{B}_{t_0}\right\} - \Pr\left\{\mathscr{A}_{t_0}^c\right\}. \tag{65}$$

Thus, due to the fact that $\Pr\left\{\mathscr{A}_{t_0}\right\} \to 1$ super-exponentially fast [15, p. 6028], we may conclude that

$$\Pr\left\{\mathscr{A}_{t_0} \cap \mathscr{B}_{t_0}\right\} \doteq \Pr\left\{\mathscr{B}_{t_0}\right\}. \tag{66}$$

Combining (60), (64) and (66), and the fact that ε is arbitrarily small, we get (68) (shown at the top of the next page) by using standard large deviations techniques (see, e.g., [15, p. 6027]), where $N_1(Q)$ and $\mathcal{S}(Q)$ are defined in (49) and (55), respectively, and

$$\tilde{\mathcal{L}} \triangleq \left\{Q: \ \max\left[t_0, s\right] - f(Q) \leq \left[R_2 - I_Q(X_2; X_1, Y)\right]_+\right\}. \tag{69}$$

Thus,

$$\Pr\left[\mathcal{N}_1(Q_{X_1 Y}) - \mathcal{N}_2(Q_{X_{1,0} Y}) \geq e^{ns} - e^{nr}\right]$$
$$\doteq \exp\left\{-nE_1(Q_{X_1 X_{2,0} Y}, Q_{X_{1,0} X_{2,0} Y})\right\} \tag{70}$$

where $E_1(\cdot, \cdot)$ is defined in (8c). Note that when $r > \max[t_0, s]$, the r.h.s. term of the inequality in the probability in (67) is negative, and due to the fact that the enumerator is nonnegative, the overall probability is unity. Finally, we average over $\boldsymbol{X}_{1,1}$ given $\mathcal{F}_0$. Using the method of types, we readily obtain (73), given at the top of the next page, where

$$\hat{\mathcal{S}}(Q_{X_2 Y}) \triangleq \left\{Q'_{X_1 X_2 Y}: \ Q'_{X_2 Y} = Q_{X_2 Y}, \ Q'_{X_1} = P_{X_1}\right\}, \tag{74}$$

and $\hat{E}_1(\cdot, \cdot)$ is defined in (8g). This completes the analysis of the last term at the r.h.s. of (44).

Next, we analyze the second and third terms at the r.h.s. of (44). Recall that the latter is given by (75). Accordingly, in the spirit of (46), we note that $P_{e,3}$ can be equivalently rewritten as (78), shown at the top of the next page, where s, r, $\mathcal{N}_1(Q)$, $\mathcal{N}_2(Q)$, and $\hat{\mathcal{S}}(Q)$, are given in (47), (48), (54), (57), and (74), respectively, and the second passage follows by using the method of types. Now, due to the fact that only $\mathcal{N}_1$ (and not $\mathcal{N}_2$) in (78) depends on $\boldsymbol{x}_{1,1}$, and since the analysis in (58)-(67) is independent of $\boldsymbol{x}_{1,1}$, it can be repeated here, and we obtain (79), shown at the top of page 14, where $N_1(Q)$ is given in (49), and we have defined $e^{n\gamma} \triangleq e^{n[t_0(Q_{X_{1,0} Y}) - f(Q)]} + e^{n[s - f(Q)]} - e^{n[r - f(Q)]}$. Recall (49), and let $\tilde{N}_1(Q) \triangleq \sum_{j=1}^{M_2 - 1} \mathcal{I}\left\{(\boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(Q_{X_2 Y})\right\}$.

We claim that (79) can be rewritten as[4]

$$P_{e,3} \doteq \max_{Q_{X_1 | X_{2,0} Y} \in \hat{\mathcal{S}}(Q_{X_{2,0} Y})} \max_{Q \in \mathcal{S}(Q_{X_1 Y})} \Pr\left\{\tilde{N}_1(Q) \geq e^{n\gamma}\right\}, \tag{80}$$

which follows from the fact that the set $\{\boldsymbol{x}_2: \ (\boldsymbol{x}_{1,1}, \boldsymbol{x}_2, \boldsymbol{y}) \in T(Q_{X_1 X_2 Y}), \text{ for some } \boldsymbol{x}_{1,1}\}$ equals $T(Q_{X_2 Y})$ (see, e.g., [11, eqs. (24)-(25)]). Thus, by using standard large deviations techniques (see, e.g., [15, p. 6027])

$$P_{e,3} \doteq \max_{Q_{X_1 | X_{2,0} Y} \in \hat{\mathcal{S}}(Q_{X_{2,0} Y})} \max_{Q \in \mathcal{S}(Q_{X_1 Y})} \tag{81}$$

$$\begin{cases} 1 & r > \max[t_0, s] \\ e^{-n[I_Q(X_2; Y) - R_2]_+} & r \leq \max[t_0, s], \ Q \in \hat{\mathcal{L}} \quad (82) \\ 0 & r \leq \max[t_0, s], \ Q \in \hat{\mathcal{L}}^c \end{cases}$$

where

$$\hat{\mathcal{L}} \triangleq \left\{Q: \ \max[t_0, s] - f(Q) \leq \left[R_2 - I_Q(X_2; Y)\right]_+\right\}. \tag{83}$$

Therefore,

$$P_{e,3} \doteq \exp\left\{-n\hat{E}_2(Q_{X_{1,0} X_{2,0} Y}, R_2)\right\} \tag{84}$$

where $\hat{E}_2(Q_{X_{1,0} X_{2,0} Y}, R_2)$ is defined in (8h). This completes the analysis of the third term at the r.h.s. of (44). Finally, recall that the second term at the r.h.s. of (44) is given by

$$A \triangleq M_1 \cdot \Pr\left[\bigcup_{T_I(Q_{X_{1,0} X_{2,0} Y})}\right.$$

$$\left.\boldsymbol{X}_{1,1} \in \mathcal{B}_1\left(\tilde{Q}^0_{X_1 X_{2,0} Y}, (\tilde{Q}^j_{X_1 X_2 Y}, \hat{Q}^j_{X_1 X_2 Y})_j)\right)\right] \tag{85}$$

and is equivalent to (86), given at the top of page 14. This term can be analyzed as before, but, we claim that it is actually larger than the fourth term at the r.h.s. of (44), and thus, essentially, does not affect the minimum in (44). Indeed, recall that the fourth term is given by (87), shown at the top of page 14, and since the factor M_1 is common to both A and B, we just need to compare the probabilities in these terms. However, it is obvious that the probability term in B is smaller than the probability in A, due to the fact that events in the former are contained in the events in the latter. Indeed, this is equivalent to comparing $\Pr\left\{(Z_1, Z_2) \in \mathcal{Z}\right\}$ and $\Pr\left\{(Z_1, z_2) \in \mathcal{Z}, \text{ for some } z_2 \in \mathcal{Z}_2\right\}$, where Z_1 and Z_2

---

[4]It is easy to see that (80) is an upper bound on (79). The other direction follows from:

$$\max_{Q, \hat{Q} \in \mathcal{S}(Q)} \Pr\left\{\tilde{N}_1(\hat{Q}) \geq e^{n\gamma(Q)}\right\} = \max_{\hat{Q} \in \mathcal{S}(Q^*)} \Pr\left\{\tilde{N}_1(\hat{Q}) \geq e^{n\gamma(Q^*)}\right\}$$

$$= \max_{\hat{Q} \in \mathcal{S}(Q^*)} \Pr\left\{N_1(\hat{Q}) \geq e^{n\gamma(Q^*)}, \text{ for some } \boldsymbol{x}_{1,1} \in T(Q^*)\right\}$$

$$\leq \max_{Q} \max_{\hat{Q} \in \mathcal{S}(Q)} \Pr\left\{N_1(\hat{Q}) \geq e^{n\gamma(Q)}, \text{ for some } \boldsymbol{x}_{1,1} \in T(Q)\right\},$$

where in the first equality we designate Q^* as the maximizer, and the second equality follows from the fact that $T(Q_{X_2 Y}) = \{\boldsymbol{x}_2: (\boldsymbol{x}_{1,1}, \boldsymbol{x}_2, \boldsymbol{y}) \in T(Q_{X_1 X_2 Y}), \text{ for some } \boldsymbol{x}_{1,1}\}$.

$$\Pr\left[\mathcal{N}_1(Q_{X_1Y}) - \mathcal{N}_2(Q_{X_{1,0}Y}) \geq e^{ns} - e^{nr}\right] \doteq \Pr\left\{\mathscr{B}_{t_0}\right\}$$

$$= \Pr\left\{\mathcal{N}_1(Q_{X_1Y}) \geq e^{nt_0(Q_{X_{1,0}Y})} + e^{ns} - e^{nr}\right\}$$

$$\doteq \max_{Q \in \mathcal{S}(Q_{X_1Y})} \Pr\left\{N_1(Q) \geq e^{n\left[t_0(Q_{X_{1,0}Y}) - f(Q)\right]} + e^{n[s - f(Q)]} - e^{n[r - f(Q)]}\right\} \tag{67}$$

$$\doteq \max_{Q \in \mathcal{S}(Q_{X_1Y})} \begin{cases} 1 & r > \max\left[t_0, s\right] \\ e^{-n[I_Q(X_2;X_1,Y) - R_2]_+} & r \leq \max\left[t_0, s\right], \ Q \in \tilde{\mathcal{L}} \\ 0 & r \leq \max\left[t_0, s\right], \ Q \in \tilde{\mathcal{L}}^c \end{cases} \tag{68}$$

$$\mathbb{E}\left\{\Pr\left[\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,1}, \boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,j}\right)\right\}\bigg|\mathcal{F}_0, \boldsymbol{X}_{1,1}\right]\bigg|\mathcal{F}_0\right\} \tag{71}$$

$$\doteq \exp\left\{-n \min_{Q_{X_1|X_{2,0}Y} \in \hat{\mathcal{S}}(Q_{X_{2,0}Y})} \left[I_Q(X_1; X_{2,0}, Y) + E_1(Q_{X_1X_{2,0}Y}, Q_{X_{1,0}X_{2,0}Y})\right]\right\} \tag{72}$$

$$\triangleq \exp\left\{-n\hat{E}_1(Q_{X_{1,0}X_{2,0}Y}, R_2)\right\} \tag{73}$$

$$P_{e,3} \triangleq \Pr\left[\bigcup_{\left\{\tilde{Q}_{X_1X_2Y}^j, \hat{Q}_{X_1X_2Y}^j\right\}_j \in T_I(Q_{X_{1,0}X_{2,0}Y})} \{\boldsymbol{X}_{2,j}\}_{j\geq 1} \in \mathcal{B}_2\left(\tilde{Q}_{X_1X_{2,0}Y}^0, (\tilde{Q}_{X_1X_2Y}^j, \hat{Q}_{X_1X_2Y}^j)_j)\right)\right]. \tag{75}$$

$$P_{e,3} = \Pr\left[\bigcup_{Q_{X_1|X_{2,0}Y}} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{x}_{2,0}\right) + \sum_{j=1}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{X}_{2,j}\right) \geq P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}\right)\right.$$

$$\left. + \sum_{j=1}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}\right), \text{ for some } \boldsymbol{x}_{1,1} \in T(Q_{X_1X_{2,0}Y})\bigg|\mathcal{F}_0\right], \tag{76}$$

$$\doteq \max_{Q_{X_1|X_{2,0}Y} \in \hat{\mathcal{S}}(Q_{X_{2,0}Y})} \Pr\left[P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{x}_{2,0}\right) + \sum_{j=1}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{X}_{2,j}\right) \geq P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}\right)\right.$$

$$\left. + \sum_{j=1}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}\right), \text{ for some } \boldsymbol{x}_{1,1} \in T(Q_{X_1X_{2,0}Y})\bigg|\mathcal{F}_0\right] \tag{77}$$

$$= \max_{Q_{X_1|X_{2,0}Y} \in \hat{\mathcal{S}}(Q_{X_{2,0}Y})} \Pr\left[\mathcal{N}_1(Q_{X_1Y}) - \mathcal{N}_2(Q_{X_{1,0}Y}) \geq e^{ns} - e^{nr}, \text{ for some } \boldsymbol{x}_{1,1} \in T(Q_{X_1X_{2,0}Y})|\mathcal{F}_0\right] \tag{78}$$

---

are RVs that are defined over the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$, respectively, and $\mathcal{Z} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$. Let $\mathcal{V} \triangleq \tilde{\mathcal{V}} \times \mathcal{Z}_2$, in which

$$\tilde{\mathcal{V}} \triangleq \{z_1 \in \mathcal{Z}_1 : (z_1, z_2) \in \mathcal{Z}, \text{ for some } z_2 \in \mathcal{Z}_2\}. \tag{88}$$

Then, it is obvious that $\mathcal{Z} \subseteq \mathcal{V}$, and thus

$$\Pr\{(Z_1, Z_2) \in \mathcal{Z}\} = \sum_{(z_1, z_2) \in \mathcal{Z}} P(z_1, z_2) \tag{89}$$

$$\leq \sum_{(z_1, z_2) \in \mathcal{V}} P(z_1, z_2) \tag{90}$$

$$= \sum_{z_1 \in \tilde{\mathcal{V}}} P(z_1) = \Pr\{(Z_1, z_2) \in \mathcal{Z}, \text{ for some } z_2\}. \tag{91}$$

Wrapping up, using (24), (44), and the last results, after averaging w.r.t. $\mathcal{F}_0$, we get (96), shown at the top of the next page, as required.

B. Proof of Theorem 2

Without loss of generality, we assume throughout, that the transmitted codewords are $\boldsymbol{x}_{1,0}$ and $\boldsymbol{x}_{2,0}$ which correspond

$$P_{e,3} \doteq \max_{Q_{X_1|X_{2,0}Y} \in \hat{\mathcal{S}}(Q_{X_{2,0}Y})} \sum_{i=i_1}^{\lceil R_2/\varepsilon \rceil} \Pr\left\{ e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(i+1)\varepsilon} \right\}$$

$$\times \Pr\left\{ \bigcup_{\boldsymbol{x}_{1,1}} \mathcal{N}_1(Q_{X_1Y}) \geq e^{ni\varepsilon} + e^{ns} - e^{nr} \,\middle|\, e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(i+1)\varepsilon} \right\}$$

$$\doteq \max_{Q_{X_1|X_{2,0}Y} \in \hat{\mathcal{S}}(Q_{X_{2,0}Y})} \Pr\left\{ \mathcal{N}_1(Q_{X_1Y}) \geq e^{nt_0(Q_{X_{1,0}Y})} + e^{ns} - e^{nr}, \text{ for some } \boldsymbol{x}_{1,1} \right\}$$

$$\doteq \max_{Q_{X_1|X_{2,0}Y} \in \hat{\mathcal{S}}(Q_{X_{2,0}Y})} \max_{Q \in \mathcal{S}(Q_{X_1Y})} \Pr\left\{ N_1(Q) \geq e^{n\gamma}, \text{ for some } \boldsymbol{x}_{1,1} \right\}, \tag{79}$$

$$A = M_1 \cdot \Pr\left[\bigcup_{T_I(Q_{X_{1,0}X_{2,0}Y})} \left\{ \begin{array}{c} (\boldsymbol{X}_{1,1}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1X_{2,0}Y}), \\ \left\{ (\boldsymbol{X}_{1,1}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1X_2Y}) \right\}_{j=1}^{M_2-1}, \text{ for some } \{\boldsymbol{x}_{2,j}\} \\ \left\{ (\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0}X_2Y}) \right\}_{j=1}^{M_2-1} \end{array} \right\} \right]. \tag{86}$$

$$B \triangleq M_1 \cdot \Pr\left[\bigcup_{T_I(Q_{X_{1,0}X_{2,0}Y})} \left\{ \begin{array}{c} (\boldsymbol{X}_{1,1}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1X_{2,0}Y}), \\ \left\{ (\boldsymbol{X}_{1,1}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1X_2Y}) \right\}_{j=1}^{M_2-1}, \\ \left\{ (\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0}X_2Y}) \right\}_{j=1}^{M_2-1} \end{array} \right\} \right], \tag{87}$$

$$\bar{P}_{e,1}^{(n)} \doteq \mathbb{E}\left\{ \min\left\{ 1, e^{-n(\hat{E}_1(Q_{X_{1,0}X_{2,0}Y},R_2)-R_1)}, e^{-n\hat{E}_2(Q_{X_{1,0}X_{2,0}Y},R_2)} \right\} \right\} \tag{93}$$

$$= \mathbb{E}\left\{ \min\left\{ e^{-n\left[\hat{E}_1(Q_{X_{1,0}X_{2,0}Y},R_2)-R_1\right]_+}, e^{-n\hat{E}_2(Q_{X_{1,0}X_{2,0}Y},R_2)} \right\} \right\} \tag{94}$$

$$= \mathbb{E}\left\{ \exp\left[-n \max\left\{ \left[\hat{E}_1(Q_{X_{1,0}X_{2,0}Y},R_2)-R_1\right]_+, \hat{E}_2(Q_{X_{1,0}X_{2,0}Y},R_2) \right\} \right] \right\} \tag{95}$$

$$\doteq \exp\left\{ -n\left[\min_{Q_{Y|X_{1,0}X_{2,0}}} \left[D(Q_{Y|X_{1,0}X_{2,0}} \| W_{Y|X_{1,0}X_{2,0}} | P_{X_{1,0}} \times P_{X_{2,0}}) + E(Q,R_1,R_2) \right] \right] \right\} \tag{96}$$

$$E(Q,R_1,R_2) \triangleq \max\left\{ \left[\hat{E}_1(Q_{X_{1,0}X_{2,0}Y},R_2)-R_1\right]_+, \hat{E}_2(Q_{X_{1,0}X_{2,0}Y},R_2) \right\}. \tag{97}$$

to $\boldsymbol{z}_{11,0}, \boldsymbol{z}_{12,0}, \boldsymbol{z}_{21,0}$ and $\boldsymbol{z}_{22,0}$. Here, we distinguish between several types of errors. Recall that the overall error probability is given by

$$\bar{P}_{e,1}^{(n)} = \Pr\left\{ (\hat{m}_{11}, \hat{m}_{12}) \neq (0,0) \right\}, \tag{98}$$

which can be divided into six possible types of errors: $(\hat{m}_{11} \neq 0, \hat{m}_{12} = 0, \hat{m}_{21} = 0)$, $(\hat{m}_{11} = 0, \hat{m}_{12} \neq 0, \hat{m}_{21} = 0)$, $(\hat{m}_{11} \neq 0, \hat{m}_{12} \neq 0, \hat{m}_{21} = 0)$, $(\hat{m}_{11} \neq 0, \hat{m}_{12} = 0, \hat{m}_{21} \neq 0)$, $(\hat{m}_{11} = 0, \hat{m}_{12} \neq 0, \hat{m}_{21} \neq 0)$, and $(\hat{m}_{11} \neq 0, \hat{m}_{12} \neq 0, \hat{m}_{21} \neq 0)$. Note that the event $(\hat{m}_{11} = 0, \hat{m}_{12} = 0, \hat{m}_{21} \neq 0)$ will not result in an error, and thus ignored. Obviously, the exponent of the overall error probability in (98) is given by the minimum between the error exponents corresponding to each type of error individually. We start with analyzing the last error event, which is also the most involved one. For this event,

the average probability of error, associated with the decoder in (14), is given by (100), given at the top of the next page, where $\tilde{\boldsymbol{Z}}_{ijk} \triangleq (\boldsymbol{Z}_{11,i}, \boldsymbol{Z}_{12,j}, \boldsymbol{Z}_{21,k})$, $\tilde{\boldsymbol{Z}}_0 \triangleq (\boldsymbol{Z}_{11,0}, \boldsymbol{Z}_{12,0}, \boldsymbol{Z}_{21,0})$, and $\mathcal{F}_0 \triangleq (\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{22,0}, \boldsymbol{Y})$. We will assess the exponential behavior of (100) in the same manner as we did for (24). Specifically, we start with expressing (100) in terms of types. First, for a given joint distribution $Q_{Z_1^4 Y}$, we let

$$f(Q_{Z_1^4 Y}) \triangleq \frac{1}{n} \log P\left(\boldsymbol{y} | \boldsymbol{x}_1(\boldsymbol{z}_1, \boldsymbol{z}_2), \boldsymbol{x}_2(\boldsymbol{z}_3, \boldsymbol{z}_4)\right). \tag{101}$$

Now, for a given joint type $Q_{Z_{1,0}^4 Y}$ of the random vectors $(\boldsymbol{Z}_{1,0}, \boldsymbol{Z}_{2,0}, \boldsymbol{Z}_{3,0}, \boldsymbol{Z}_{4,0}, \boldsymbol{Y})$, we define the set $T_I(Q_{Z_{1,0}^4 Y})$ in (102) given at the top of the next page. Now, with these definitions, fixing $Q_{Z_{1,0}^4 Y}$, it follows, by definition, that the

$$P_e^{(7)} \triangleq \Pr\left[\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{22,l}) \geq \sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{22,l})\right\}\right] \tag{99}$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{22,l}) \geq \sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{22,l})\right\}\Bigg|\mathcal{F}_0\right]\right\} \tag{100}$$

$$T_I(Q_{Z_{1,0}^4 Y}) \triangleq \left\{\tilde{Q}^0_{Z_1^3 Z_{4,0} Y} \in \mathcal{S}_0, \left(\left\{\tilde{Q}^l_{Z_1^4 Y}\right\}_{l=1}^{M_{22}-1}, \left\{\hat{Q}^l_{Z_{1,0}^3 Z_4 Y}\right\}_{l=1}^{M_{22}-1}\right) \in \mathcal{S}_1 : \right.$$

$$\left. e^{nf(\tilde{Q}^0_{Z_1^3 Z_{4,0} Y})} + \sum_{l=1}^{M_{22}-1}\left[e^{nf(\tilde{Q}^l_{Z_1^4 Y})} - e^{nf(\hat{Q}^l_{Z_{1,0}^3 Z_4 Y})}\right] \geq e^{nf(Q_{Z_{1,0}^4 Y})}\right\} \tag{102}$$

$$\mathcal{S}_0(Q_{Z_{1,0}^4 Y}) \triangleq \left\{\tilde{Q}^0_{Z_1^3 Z_{4,0} Y} : \tilde{Q}^0_{Z_i} = P_{Z_i}, \tilde{Q}^0_{Z_{4,0}Y} = Q_{Z_{4,0}Y}, \forall 1 \leq i \leq 4\right\}, \tag{103}$$

$$\mathcal{S}_1(Q_{Z_{1,0}^4 Y}) \triangleq \left\{\left\{\tilde{Q}^l_{Z_1^4 Y}\right\}_{l=1}^{M_{22}-1}, \left\{\hat{Q}^l_{Z_{1,0}^3 Z_4 Y}\right\}_{l=1}^{M_{22}-1} : \tilde{Q}^l_{Z_i} = P_{Z_i}, \tilde{Q}^l_Y = Q_Y, \right.$$

$$\hat{Q}^l_{Z_i} = P_{Z_i}, \hat{Q}^l_{Z_{1,0}^3 Y} = Q_{Z_{1,0}^3 Y}, \forall 1 \leq i \leq 4, \forall 1 \leq l \leq M_{22}-1$$

$$\left. \tilde{Q}^l_{Z_4 Y} = \hat{Q}^l_{Z_4 Y}, \tilde{Q}^l_{Z_1^3 Y} = \tilde{Q}^m_{Z_1^3 Y}, \forall l, m\right\}. \tag{104}$$

---

error event

$$\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}$$

$$\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{4,l}) \geq \sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{4,l})\right\} \tag{105}$$

can be rewritten, in terms of types, as follows

$$\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\bigcup_{T_I(Q_{Z_{1,0}^4 Y})} \tag{106}$$

$$\left\{\begin{array}{c}(\tilde{\boldsymbol{Z}}_{ijk}, \boldsymbol{z}_{4,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{Z_1^3 Z_{4,0} Y}), \\ \left\{(\tilde{\boldsymbol{Z}}_{ijk}, \boldsymbol{Z}_{4,l}, \boldsymbol{y}) \in T(\tilde{Q}^l_{Z_1^4 Y})\right\}_{l=1}^{M_{22}-1}, \\ \left\{(\tilde{\boldsymbol{z}}_0, \boldsymbol{Z}_{4,l}, \boldsymbol{y}) \in T(\hat{Q}^l_{Z_{1,0}^3 Z_4 Y})\right\}_{l=1}^{M_{22}-1}\end{array}\right\}. \tag{107}$$

We next analyze the probability of (107), conditioned on $\mathcal{F}_0$. Note that the inner union in (107) is over vectors of types (an exponential number of them). Finally, as before, we simplify the notations of (107), and write it equivalently as

$$\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\bigcup_{\boldsymbol{l}} \tag{108}$$

$$\left\{\begin{array}{c}\tilde{\boldsymbol{Z}}_{ijk} \in \mathcal{A}_{\boldsymbol{l},0}, \\ (\tilde{\boldsymbol{Z}}_{ijk}, \boldsymbol{Z}_{4,m}) \in \mathcal{A}_{\boldsymbol{l},m}, \quad \text{for } m = 1, \ldots, M_{22}-1 \\ \boldsymbol{Z}_{4,m} \in \tilde{\mathcal{A}}_{\boldsymbol{l},m}, \quad \text{for } m = 1, \ldots, M_{22}-1\end{array}\right\} \tag{109}$$

where, again, the index "$\boldsymbol{l}$" in the inner union runs over the combinations of types (namely, $\boldsymbol{l} = \{\tilde{Q}^l_{Z_1^4 Y}, \hat{Q}^l_{Z_{1,0}^3 Z_4 Y}\}_l$) that belong to $T_I(Q_{Z_{1,0}^4 Y})$, and the various sets $\{\mathcal{A}_{\boldsymbol{l},j}, \tilde{\mathcal{A}}_{\boldsymbol{l},j}\}_{\boldsymbol{l},j}$

correspond to the typical sets in (107) (recall that $(\boldsymbol{z}_{1,0}, \boldsymbol{z}_{2,0}, \boldsymbol{z}_{3,0}, \boldsymbol{z}_{4,0}, \boldsymbol{y})$ are given in this stage). Similarly as in the proof of Theorem 1, we derive upper bound on a generic probability which have the form of (109). In the following, we give a generalization of Lemma 1 to the probability of a union indexed by K values. The proof is very similar to the proof of Lemma 1, and thus omitted for brevity. For a given subset $\mathcal{J} = \{j_1, \ldots, k_{|\mathcal{J}|}\}$ of $\{1, \ldots, J\}$ we write $\boldsymbol{Z}_{\mathcal{J}}$ as a shorthand for $(Z_{j_1}, \ldots, Z_{j_{|\mathcal{J}|}})$.

Lemma 2 Let $\{Z_1(i)\}_{i=1}^{N_1}, \ldots, \{Z_J(i)\}_{i=1}^{N_J}$ and $\{V_1(i)\}_{i=1}^{N_{J+1}}, \{V_2(i)\}_{i=1}^{N_{J+1}}, \ldots, \{V_K(i)\}_{i=1}^{N_{J+1}}$ be independent sequences of independently and identically distributed (i.i.d.) RVs on the alphabets $\mathcal{Z}_1 \times \ldots \times \mathcal{Z}_J \times \mathcal{V}_1 \times \ldots \times \mathcal{V}_K$, respectively, with $Z_1(i) \sim P_{Z_1}, \ldots, Z_J(i) \sim P_{Z_J}, V_1(i) \sim P_{V_1}, \ldots, V_K(i) \sim P_{V_K}$. Fix a sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^N, \{\mathcal{A}_{i,2}\}_{i=1}^N, \ldots, \{\mathcal{A}_{i,K}\}_{i=1}^N$, where $\mathcal{A}_{i,j} \subseteq \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_J \times \mathcal{V}_j$, for $1 \leq j \leq K$ and for all $1 \leq i \leq N$. Also, fix a set $\{\mathcal{A}_{i,0}\}_{i=1}^N$ where $\mathcal{A}_{i,0} \subseteq \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_J$ for all $1 \leq i \leq N$, and another sequence of sets $\{\mathcal{G}_{i,1}\}_{i=1}^N, \{\mathcal{G}_{i,2}\}_{i=1}^N, \ldots, \{\mathcal{G}_{i,K}\}_{i=1}^N$, where $\mathcal{G}_{i,j} \subseteq \mathcal{V}_j$, for $1 \leq j \leq K$ and for all $1 \leq i \leq N$. Let $\boldsymbol{U} = (Z_1, Z_2, \ldots, Z_J, U_{J+1})$ with $U_{J+1} \triangleq (V_1, \ldots, V_K)$. Finally, define $\mathcal{B}_{l,\mathcal{J}}$ given in (110), for $l = 1, 2, \ldots, N$, and $\mathcal{Z}(i_1^J) = (Z_1(i_1), \ldots, Z_J(i_J))$. Then, a general upper bound is given by (111), shown at the top of the next page.

Applying Lemma 2 on (107) (or, (109)) we obtain (112), shown at the top of the next page, where $N_1 = M_{11}, N_2 = M_{12}, N_3 = M_{21}, N_4 = 1$, and

$$\boldsymbol{U} = (\boldsymbol{Z}_{11}, \boldsymbol{Z}_{12}, \boldsymbol{Z}_{21}, \boldsymbol{U}_4) \tag{113}$$

in which $\boldsymbol{U}_4 = (\boldsymbol{Z}_{4,1}, \ldots, \boldsymbol{Z}_{4,M_{22}-1})$, and $\mathcal{B}_{l,\mathcal{J}}$ is given in (114), also shown at the top of the next page. The various

$$\mathcal{B}_{l,\mathcal{J}} \triangleq \left\{ \boldsymbol{u}_{\mathcal{J}} : \ z_1^J \in \mathcal{A}_{l,0}, \ \bigcap_{j=1}^{K} \left(z_1^J, v_j\right) \in \mathcal{A}_{l,j}, \ \bigcap_{j=1}^{K} v_j \in \mathcal{G}_{l,j} \ \text{ for some } \boldsymbol{u}_{\mathcal{J}^c} \right\}, \tag{110}$$

$$\Pr\left\{ \bigcup_{i_1^J,j} \left\{ \bigcup_{l=1}^{N} \left\{ \mathcal{Z}(i_1^J) \in \mathcal{A}_{l,0}, \ \bigcap_{k=1}^{K} \left(\mathcal{Z}(i_1^J), V_k(j)\right) \in \mathcal{A}_{l,k}, \ \bigcap_{k=1}^{K} V_k(j) \in \mathcal{G}_{l,k} \right\} \right\} \right\}$$
$$\leq \min\left\{ 1, \ \min_{\mathcal{J}\subseteq\{1,\ldots,J+1\}\mathcal{J}\neq\emptyset} \left(\prod_{j\in\mathcal{J}} N_j\right) \Pr\left\{ \bigcup_{l=1}^{N} \boldsymbol{U}_{\mathcal{J}} \in \mathcal{B}_{l,\mathcal{J}} \right\} \right\}. \tag{111}$$

$$\Pr\left\{ \bigcup_{i=1}^{M_{11}-1} \bigcup_{j=1}^{M_{12}-1} \bigcup_{k=1}^{M_{21}-1} \left\{ \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk}, \boldsymbol{Z}_{4,l}) \geq \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{4,l}) \right\} \ \middle| \ \mathcal{F}_0 \right\}$$
$$\overset{\cdot}{\leq} \min\left\{ 1, \ \min_{\mathcal{J}\subseteq\{1,\ldots,4\}\mathcal{J}\neq\emptyset} \left(\prod_{j\in\mathcal{J}} N_j\right) \Pr\left\{ \bigcup_{l} \boldsymbol{U}_{\mathcal{J}} \in \mathcal{B}_{l,\mathcal{J}} \right\} \right\} \tag{112}$$

$$\mathcal{B}_{\boldsymbol{l},\mathcal{J}} = \left\{ \boldsymbol{u}_{\mathcal{J}} : \ \begin{array}{c} (\tilde{\boldsymbol{z}}_{111}, \boldsymbol{z}_{4,0}, \boldsymbol{y}) \in T(\tilde{Q}_{X_1 X_{2,0} Y}^0), \\[2mm] \left\{ (\tilde{\boldsymbol{z}}_{111}, \boldsymbol{z}_{4,l}, \boldsymbol{y}) \in T(\tilde{Q}_{Z_1^4 Y}^l) \right\}_{l=1}^{M_{22}-1}, \\[2mm] \left\{ (\tilde{\boldsymbol{z}}_0, \boldsymbol{z}_{4,l}, \boldsymbol{y}) \in T(\hat{Q}_{X_{1,0} X_2 Y}^j) \right\}_{l=1}^{M_{22}-1}, \ \text{ for some } \boldsymbol{u}_{\mathcal{J}^c} \end{array} \right\} \tag{114}$$

possibilities for the set $\mathcal{J}$ are,

$$\left\{ \begin{array}{c} 1; 2; 3; 4; \\ 12; 13; 14; 23; 24; 34; \\ 123; 124; 134; 234; \\ 1234 \end{array} \right\}, \tag{115}$$

namely, we have 15 possibilities. We claim that possibilities $\{1, 2, 3, 12, 13, 23, 123\}$ do not affect the outer minimum in (112), and so we left with possibilities $\{4, 14, 24, 34, 124, 134, 234, 1234\}$. This observation follows from the same arguments used in (86)-(91) for the second term at the r.h.s. of (44). For example, possibilities $\{1, 2, 3\}$ do not affect the outer minimum due to the fact that the probabilities that correspond to possibilities $\{14, 24, 34\}$, respectively, are smaller. Indeed, the multiplicative factors in (112) for each of the pairs $(1, 14)$, $(2, 24)$, and $(3, 34)$, are the same, but the respective probabilities in (112) are smaller for $\{14, 24, 34\}$ (due to the same reason used in (91)). Similarly, possibilities $\{12, 13, 23, 123\}$ do not affect the outer minimum due to possibilities $\{124, 134, 234, 1234\}$, respectively.

In the following, we analyze the remaining terms. For example, the term that corresponds to possibility "1234", is given by

$$P_{e,1234} \triangleq M_{11} M_{12} M_{21} \Pr\left\{ \bigcup_l \boldsymbol{U} \in \mathcal{B}_{\boldsymbol{l},1234} \right\}, \tag{116}$$

which similarly to the passage from (105) to (107), can be rewritten as in (117), shown at the top of the next page. Equation (117) has the same form of the probability in (71), which we already analyzed. Accordingly, we similarly obtain (120), where $E_7(\cdot, \cdot)$ is defined in (A.4), and $\mathcal{S}_{\{4\}}(Q)$ is given in (A.1).

The other terms are handled in a similar fashion. Specifically, let $\hat{\boldsymbol{Z}} \triangleq \{Z_1, Z_2, Z_3\}$, and define the sets $\mathcal{U} = \{1, 2, 3, 12, 13, 23, 123\}$, and $\tilde{\mathcal{U}} = \{14, 24, 34, 124, 134, 234, 1234\}$. Then, define for any[5] $u \in \{1, 2, \ldots, 7\}$:

$$P_{e,u}^{(6)} \triangleq M_{\mathcal{U}(u)} \cdot \Pr\left\{ \bigcup_l \boldsymbol{U}_{\tilde{\mathcal{U}}(u)} \in \mathcal{B}_{\boldsymbol{l},\tilde{\mathcal{U}}(u)} \right\}, \tag{121}$$

where

$$M_{\mathcal{U}(1)} \triangleq M_{11}; \ M_{\mathcal{U}(2)} = M_{12}; \ M_{\mathcal{U}(3)} = M_{21};$$
$$M_{\mathcal{U}(4)} = M_{11} M_{12}; \ M_{\mathcal{U}(5)} \triangleq M_{11} M_{21};$$
$$M_{\mathcal{U}(6)} = M_{12} M_{21} \ M_{\mathcal{U}(7)} = M_{11} M_{12} M_{21}. \tag{122}$$

Accordingly, following (49)-(84), we get (123), shown at the top of the next page, where $E_u(\cdot, \cdot)$ is defined in (A.4). Note that the mutual information term in the above exponent is due to the averaging over $\hat{\boldsymbol{Z}}_{\mathcal{U}(u)}$, and it is resulted by using the method of types as in (72). This concludes the analysis for

---

[5]Note that $P_{e,7}^{(6)}$ correspond to $P_{e,1234}$ in (117).

$$P_{e,1234} = M_{11}M_{12}M_{21}\Pr\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{111},\boldsymbol{Z}_{4,l}) \geq \sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{4,l})\,\middle|\,\mathcal{F}_0\right\}$$

$$= M_{11}M_{12}M_{21}\mathbb{E}\left\{\Pr\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{111},\boldsymbol{Z}_{4,l}) \geq \sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{4,l})\,\middle|\,\mathcal{F}_0,\tilde{\boldsymbol{Z}}_{111}\right\}\middle|\,\mathcal{F}_0\right\} \quad (117)$$

$$\mathbb{E}\left\{\Pr\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{111},\boldsymbol{Z}_{4,l}) \geq \sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{4,l})\,\middle|\,\mathcal{F}_0,\tilde{\boldsymbol{Z}}_{111}\right\}\middle|\,\mathcal{F}_0\right\} \quad (118)$$

$$\dot{=} \exp\left\{-n\min_{Q_{Z_1^3|Z_{4,0}Y}\in\mathcal{S}_{\{4\}}(Q_{Z_{1,0}^4Y})}\left[I_Q(Z_1^3;Z_{4,0},Y) + E_7(Q_{Z_1^3Z_{4,0}Y},Q_{Z_{1,0}^4Y})\right]\right\} \quad (119)$$

$$\triangleq \exp\left\{-n\hat{E}_7^{(6)}(Q_{Z_{1,0}^4Y},R_{22})\right\} \quad (120)$$

$$P_{e,u}^{(6)} \dot{=} \exp\left\{-n\min_{Q_{Z_1^3|Z_{4,0}Y}\in\mathcal{S}_{\{4\}}(Q_{Z_{1,0}^4Y})}\left[I_Q(\hat{\boldsymbol{Z}}_{\mathcal{U}(u)};Z_{4,0},Y|\hat{\boldsymbol{Z}}_{123\setminus\mathcal{U}(u)}) + E_u(Q_{Z_1^3Z_{4,0}Y},Q_{Z_{1,0}^4Y})\right]\right\}$$

$$\triangleq \exp\left\{-n\hat{E}_u^{(6)}(Q_{Z_{1,0}^4Y},R_{22})\right\} \quad (123)$$

possibilities $\{14,24,34,124,134,234,1234\}$, and we left with possibility $\{4\}$, which is very similar to (75). Accordingly, using the same arguments in (78)-(84), we obtain

$$P_{e,8}^{(6)} \triangleq \Pr\left\{\bigcup_l \boldsymbol{U}_4 \in \mathcal{B}_{\boldsymbol{l},4}\right\}$$

$$\dot{=} \exp\left\{-n\min_{Q_{Z_1^3|Z_{4,0}Y}\in\mathcal{S}_{\{4\}}(Q_{Z_{1,0}^4Y})}E_0(Q_{Z_1^3Z_{4,0}Y},Q_{Z_{1,0}^4Y})\right\}$$

$$\triangleq \exp\left\{-n\hat{E}_8^{(6)}(Q_{Z_{1,0}^4Y},R_{22})\right\} \quad (124)$$

where $E_0(\cdot,\cdot)$ is, again, defined in (A.4). Wrapping up, using (100), (112), and the last results, after averaging w.r.t. $\mathcal{F}_0$, we get (129), shown at the top of the next page, where $\mathcal{R}_u$ for $u = 1,2,\ldots,7$ is defined in (A.6).

This concludes the analysis of the error event ($\hat{m}_{11} \neq 0, \hat{m}_{12} \neq 0, \hat{m}_{21} \neq 0$) in (98). The other types of errors are analyzed in a similar manner. For ($\hat{m}_{11} \neq 0, \hat{m}_{12} = 0, \hat{m}_{21} = 0$), the average probability of error, associated with the decoder in (14) is given in (131), given at the top of the next page, where $\mathcal{F}_0 \triangleq (\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{22,0}, \boldsymbol{Y})$. Thus, due to the fact that $(\boldsymbol{Z}_{12,0}, \boldsymbol{Z}_{21,0})$ are now fixed, they play a same role as $\boldsymbol{Y}$ and $\boldsymbol{Z}_{22,0}$. Accordingly, following the same steps as in (78)-(84), we get (132), presented at the top of the next page, where $E_6(\cdot,\cdot)$ and $E_7(\cdot,\cdot)$ are defined in (A.4), and $\mathcal{S}_{\{2,3,4\}}(Q)$ is given in (A.1). Again, since $(\boldsymbol{Z}_{12,0}, \boldsymbol{Z}_{21,0})$, which correspond to $(\boldsymbol{Z}_{2,0}, \boldsymbol{Z}_{3,0})$), are fixed, they are conjugated to $(\boldsymbol{Z}_{4,0}, \boldsymbol{Y})$. The error exponent of $P_e^{(2)}$ which corresponds to ($\hat{m}_{11} = 0, \hat{m}_{12} \neq 0, \hat{m}_{21} = 0$) can be derived in the same way. We get that the exponent of $P_e^{(2)}$ is obtained by replacing the role of Z_1 with Z_2 and

$\mathcal{R}_1$ with $\mathcal{R}_2$, in (132)-(134). Similarly, $P_e^{(3)}$, corresponding to ($\hat{m}_{11} \neq 0, \hat{m}_{12} \neq 0, \hat{m}_{21} = 0$), is upper bounded by (136), presented at the top of the next page, where $\mathcal{S}_{\{3,4\}}(Q)$ is defined in (A.1).

Finally, the error exponents of $P_e^{(4)}$ and $P_e^{(5)}$, corresponding to ($\hat{m}_{11} \neq 0, \hat{m}_{12} = 0, \hat{m}_{21} \neq 0$) and ($\hat{m}_{11} = 0, \hat{m}_{12} \neq 0, \hat{m}_{21} \neq 0$), respectively, are obtained in the same way. The exponent of $P_e^{(4)}$ is obtained by replacing the role of Z_2 with Z_3, and changing the minimization in (137) to over the indexes $\{1,3,5\}$, and the exponent of $P_e^{(5)}$ is obtained by replacing the role of Z_1 with Z_3, and changing the minimization in (137) to over the indexes $\{2,3,6\}$.

APPENDIX A
DEFINITIONS FOR THEOREM 2

In this appendix, we give the definitions of the various parameters appearing in Theorem 2. For simplicity of notation, in the following, we use the indexes $\{1,2,3,4\}$ instead of $\{11,12,21,22\}$, respectively. Let $\boldsymbol{Z} \triangleq (Z_1,Z_2,Z_3)$, and $\mathcal{U} \triangleq \{1,2,3,12,13,23,123\}$. For $u \in \{0,1,2,\ldots,7\}$, $\boldsymbol{Z}_{\mathcal{U}(u)}$ is a random vector consisting of the RVs corresponding to the indexes in $\mathcal{U}(u)$, for example, $\boldsymbol{Z}_1 \triangleq \boldsymbol{Z}_{\mathcal{U}(1)} = Z_1$, $\boldsymbol{Z}_{12} \triangleq \boldsymbol{Z}_{\mathcal{U}(4)} = (Z_1,Z_2)$, $\boldsymbol{Z}_{123} \triangleq \boldsymbol{Z}_{\mathcal{U}(7)} = (Z_1,Z_2,Z_3)$, and so on, where we define $\boldsymbol{Z}_{\mathcal{U}(0)} = \emptyset$. Let also $\tilde{\boldsymbol{Z}} \triangleq \{Z_1,Z_2,Z_3,Z_4\}$, $\mathcal{I} \subseteq \{1,2,3,4\}$, and $\tilde{\boldsymbol{Z}}_{\mathcal{I}}$ be the restriction of the entries of $\tilde{\boldsymbol{Z}}$ on the set $\mathcal{I}$. Then, let

$$\mathcal{S}_{\mathcal{I}}(Q) \triangleq \left\{\tilde{Q}: \tilde{Q}_{\tilde{\boldsymbol{z}}_{\mathcal{I}}Y_1} = Q_{\tilde{\boldsymbol{z}}_{\mathcal{I}}Y_1}, \tilde{Q}_{Z_i} = P_{Z_i}, \text{ for } i \in \mathcal{I}^c\right\}. \quad (A.1)$$

$$P_e^{(6)} \dot{\leq} \mathbb{E}\left\{ \min\left\{ 1, \min_{u\in\{1:7\}} e^{-n\left[\hat{E}_u^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22}) - n^{-1}\log M_{\tilde{\mathcal{U}}(u)}\right]}, e^{-n\hat{E}_8^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22})} \right\} \right\} \tag{125}$$

$$= \mathbb{E}\left\{ \min\left\{ \min_{u\in\{1:7\}} e^{-n\left[\hat{E}_u^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22}) - n^{-1}\log M_{\tilde{\mathcal{U}}(u)}\right]_+}, e^{-n\hat{E}_8^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22})} \right\} \right\} \tag{126}$$

$$= \mathbb{E}\left\{ \exp\left[-n\max\left\{ \max_u \left[\hat{E}_u^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \frac{1}{n}\log M_{\tilde{\mathcal{U}}(u)}\right]_+, \hat{E}_8^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22}) \right\} \right] \right\} \tag{127}$$

$$\dot{=} \exp\left\{ -n\left[\min_{Q_{Y|Z_{1,0}^4}} \left[D(Q_{Y|Z_{1,0}^4}||W_{Y|Z_{1,0}^4}|P_{Z_{1,0}^4}) + E_{\mathrm{HK}}^{(6)}(Q_{Z_{1,0}^4 Y}) \right] \right] \right\} \tag{128}$$

$$E_{\mathrm{HK}}^{(6)}(Q_{Z_{1,0}^4 Y}) \triangleq \max\left\{ \max_{u\in\{1:7\}} \left[\hat{E}_u^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(6)}(Q_{Z_{1,0}^4 Y}, R_{22}) \right\} \tag{129}$$

$$P_e^{(1)} = \Pr\left[\bigcup_{i=1}^{M_{11}-1} \left\{ \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{i00}, \boldsymbol{Z}_{22,l}) \geq \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{22,l}) \right\} \right] \tag{130}$$

$$= \mathbb{E}\left\{ \Pr\left[\bigcup_{i=1}^{M_{11}-1} \left\{ \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{i00}, \boldsymbol{Z}_{22,l}) \geq \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{22,l}) \right\} \Bigg| \mathcal{F}_0 \right] \right\} \tag{131}$$

$$P_e^{(1)} \dot{\leq} \exp\left\{ -n\left[\min_{Q_{Y|Z_{1,0}^4}} \left[D(Q_{Y|Z_{1,0}^4}||W_{Y|Z_{1,0}^4}|P_{Z_{1,0}^4}) + E_{\mathrm{HK}}^{(1)}(Q_{Z_{1,0}^4 Y}) \right] \right] \right\} \tag{132}$$

$$E_{\mathrm{HK}}^{(1)}(Q_{Z_{1,0}^4 Y}) \triangleq \max\left\{ \left[\hat{E}^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \mathcal{R}_1\right]_+, \hat{E}_8^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22}) \right\} \tag{133}$$

$$\hat{E}^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1|Z_{2,0}^4 Y}\in\mathcal{S}_{\{2,3,4\}}(Q_{Z_{1,0}^4 Y})} \left[I_Q(Z_1; Z_{2,0}^4, Y) + E_7(Q_{Z_1 Z_{2,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \right] \tag{134}$$

$$\hat{E}_8^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1|Z_{2,0}^4 Y}\in\mathcal{S}_{\{2,3,4\}}(Q_{Z_{1,0}^4 Y})} E_6(Q_{Z_1 Z_{2,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \tag{135}$$

$$P_e^{(3)} \dot{\leq} \exp\left\{ -n\left[\min_{Q_{Y|Z_{1,0}^4}} \left[D(Q_{Y|Z_{1,0}^4}||W_{Y|Z_{1,0}^4}|P_{Z_{1,0}^4}) + E_{\mathrm{HK}}^{(3)}(Q_{Z_{1,0}^4 Y}) \right] \right] \right\} \tag{136}$$

$$E_{\mathrm{HK}}^{(3)}(Q_{Z_{1,0}^4 Y}) \triangleq \max\left\{ \max_{u\in\{1,2,4\}} \left[\hat{E}_u^{(3)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(3)}(Q_{Z_{1,0}^4 Y}, R_{22}) \right\} \tag{137}$$

$$\hat{E}_1^{(3)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1^2|Z_{3,0}^4 Y}\in\mathcal{S}_{\{3,4\}}(Q_{Z_{1,0}^4 Y})} \left[I_Q(Z_1; Z_{3,0}^4, Y|Z_2) + E_5(Q_{Z_1^2 Z_{3,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \right] \tag{138}$$

$$\hat{E}_2^{(3)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1^2|Z_{3,0}^4 Y}\in\mathcal{S}_{\{3,4\}}(Q_{Z_{1,0}^4 Y})} \left[I_Q(Z_2; Z_{3,0}^4, Y|Z_1) + E_6(Q_{Z_1^2 Z_{3,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \right] \tag{139}$$

$$\hat{E}_4^{(3)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1^2|Z_{3,0}^4 Y}\in\mathcal{S}_{\{3,4\}}(Q_{Z_{1,0}^4 Y})} \left[I_Q(Z_1, Z_2; Z_{3,0}^4, Y) + E_7(Q_{Z_1^2 Z_{3,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \right] \tag{140}$$

$$\hat{E}_8^{(3)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1^2|Z_{3,0}^4 Y}\in\mathcal{S}_{\{3,4\}}(Q_{Z_{1,0}^4 Y})} E_3(Q_{Z_1^2 Z_{3,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \tag{141}$$

Define

$$f(Q_{Z_1^4 Y_1}) \triangleq \mathbb{E}_Q\left[\log W_{Y_1|X_1 X_2}(Y_1|X_1(Z_1, Z_2), X_2(Z_3, Z_4)) \right] \tag{A.2}$$

and let

$$r_0(Q_{Z_1^3 Y_1}) \triangleq R_{22} + \max_{\substack{\hat{Q}:\ \hat{Q}\in\mathcal{S}_{\{1,2,3\}}(Q) \\ I_{\hat{Q}}(Z_4; Z_1^3, Y_1)\leq R_{22}}} f(\hat{Q}) - I_{\hat{Q}}(Z_4; Z_1^3, Y_1). \tag{A.3}$$

For $u \in \{0, 1, 2, \ldots, 7\}$, define

$$E_u(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \triangleq$$
$$\min_{\substack{\hat{Q}: \ \hat{Q} \in \mathcal{S}_{\{1,2,3\}}(\tilde{Q}) \\ \hat{Q} \in \mathcal{D}_u(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1})}} \left[I_{\hat{Q}}(Z_4; \boldsymbol{Z}_{\mathcal{U}(u)}, Y_1) - R_{22} \right]_+,$$

(A.4)

where

$$\mathcal{D}_u(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \triangleq \left\{ \hat{Q}: \ \max \left[r_0(Q_{Z_1^4 Y_1}), f(Q_{Z_1^4 Y_1}) \right] \right.$$
$$\left. \leq \max \left\{ f(\tilde{Q}_{Z_1^4 Y_1}), f(\hat{Q}) + \left[R_{22} - I_{\hat{Q}}(Z_4; \boldsymbol{Z}_{\mathcal{U}(u)}, Y_1) \right]_+ \right\} \right\}$$

(A.5)

Finally, we let

$$\mathcal{R}_1 \triangleq R_{11}; \ \mathcal{R}_2 \triangleq R_{12}; \ \mathcal{R}_3 \triangleq R_{21}; \ \mathcal{R}_4 \triangleq R_{11} + R_{12};$$
$$\mathcal{R}_5 \triangleq R_{11} + R_{21}; \ \mathcal{R}_6 \triangleq R_{12} + R_{21};$$
$$\mathcal{R}_7 \triangleq R_{11} + R_{12} + R_{21}.$$

(A.6)

Using all the above definition, we define (A.7)-(A.30), shown at the top of page 20.

APPENDIX B
PROOF OF LEMMA 1

We prove a generalized version of Lemma 1, where we consider random sequences, $\{V_2(i)\}_{i=1}^{L_2}, \ldots, \{V_K(i)\}_{i=1}^{L_2}$, rather than single RVs $V_2, \ldots, V_K$. Lemma 1 is then obtained on substituting $L_2 = 1$. We start with the following result which can be thought of as an extension of [11, Lemma 2].

Lemma 3 Let $\{V_1(i)\}_{i=1}^{L_1}, \{V_2(i)\}_{i=1}^{L_2}, \ldots, \{V_K(i)\}_{i=1}^{L_2}$ be independent sequences of independently and identically distributed (i.i.d.) RVs on the alphabets $\mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K$, respectively, with $V_1(i) \sim P_{V_1}, V_2(i) \sim P_{V_2}, \ldots, V_K(i) \sim P_{V_K}$. Fix a sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^N, \{\mathcal{A}_{i,2}\}_{i=1}^N, \ldots, \{\mathcal{A}_{i,K-1}\}_{i=1}^N$, where $\mathcal{A}_{i,j} \subseteq \mathcal{V}_1 \times \mathcal{V}_{j+1}$, for $1 \leq j \leq K-1$ and for all $1 \leq i \leq N$. Also, fix a set $\{\mathcal{A}_{i,0}\}_{i=1}^N$ where $\mathcal{A}_{i,0} \subseteq \mathcal{V}_1$ for all $1 \leq i \leq N$, and another sequence of sets $\{\mathcal{G}_{i,2}\}_{i=1}^N, \{\mathcal{G}_{i,3}\}_{i=1}^N, \ldots, \{\mathcal{G}_{i,K}\}_{i=1}^N$, where $\mathcal{G}_{i,j} \subseteq \mathcal{V}_j$, for $2 \leq j \leq K$ and for all $1 \leq i \leq N$. We have (B.1), shown at the top of page 21, with $(V_1, \ldots, V_K) \sim P_{V_1} \cdots \times P_{V_K}$.

Proof of Lemma 3: The second term in (B.1) follows by first applying the union bound over i as in (B.2), shown at the top of page 21, and then we apply the truncated union bound to the union over j, and obtain (B.3). The third term is obtained similarly by applying the union bounds in the opposite order, and the upper bound of 1 is trivial. ∎

We are now in a position to prove Lemma 1.

Proof of Lemma 1: To obtain (38) we weaken (B.1) as follows. Let $\mathcal{F} \triangleq \bigcup_{l=1}^N \{V_1 \in \mathcal{B}_{l,1}\}$. The second term in (38) follows from (B.4), shown at the top of page 21, where the second equality follows from the fact that the inner term in the expectation vanishes over $\bigcap_{l=1}^N \{V_1 \notin \mathcal{B}_{l,1}\}$, and the third

inequality follows from the fact that $\min\{1, x\} \leq 1$. The third term in (38) follows in a similar fashion, and the forth term follows from the fact that $\min\{1, x\} \leq x$, and thus we get (B.5), which concludes the proof. ∎

APPENDIX C
COMPUTATIONAL ASPECTS OF THE EXPONENTS

In this appendix, we discuss the computation of (8j), similarly as in [21]. We start with an alternative formulation of (8j). Recall that

$$\tilde{E}_1(R_1, R_2) = \min_Q \{D(Q||W) + E(Q, R_1, R_2)\}, \quad \text{(C.1)}$$

where

$$E(Q, R_1, R_2) = \max \left\{ \left[\hat{E}_1(Q, R_2) - R_1 \right]_+, \hat{E}_2(Q, R_2) \right\}. \quad \text{(C.2)}$$

In the following, for a given $Q_{Y_1|X_1 X_2}$, we show that $\hat{E}_1(Q, R_2)$ and $\hat{E}_2(Q, R_2)$ can be calculated efficiently. For brevity, we let $\tilde{I}(\tilde{Q}) \equiv I_{\tilde{Q}}(X_1; X_2, Y_1)$ and $\hat{I}(\hat{Q}) \equiv I_{\hat{Q}}(X_2; X_1, Y_1)$. Recall that

$$\hat{E}_1(Q, R_2) = \min_{\substack{\tilde{Q} \in \mathcal{S}(Q), \hat{Q} \in \hat{S}(\tilde{Q}), \\ \hat{Q} \in \mathcal{L}(\tilde{Q}, Q)}} \left\{ \tilde{I}(\tilde{Q}) + \left[\hat{I}(\hat{Q}) - R_2 \right]_+ \right\},$$

(C.3)

$$\hat{E}_2(Q, R_2) = \min_{\substack{\tilde{Q} \in \mathcal{S}(Q), \hat{Q} \in \hat{S}(\tilde{Q}), \\ \hat{Q} \in \hat{\mathcal{L}}(\tilde{Q}, Q)}} \left[\hat{I}_{\hat{Q}}(X_2; Y_1) - R_2 \right]_+, \quad \text{(C.4)}$$

where $\mathcal{L}$ and $\hat{\mathcal{L}}$ are defined in (8e) and (8f), respectively, $\mathcal{S}(Q) = \{\tilde{Q}: \ \tilde{Q}_{X_1} = P_{X_1}, \ \tilde{Q}_{X_2 Y_1} = Q_{X_2 Y_1}\}$, and $\hat{S}(\tilde{Q}) = \{\hat{Q}: \ \hat{Q}_{X_2} = P_{X_2}, \ \hat{Q}_{X_1 Y_1} = \tilde{Q}_{X_1 Y_1}\}$. In [21], it was shown that $\hat{E}_1(Q, R_2)$ can be equivalently expressed in terms of the minimum between the following terms:

$$\hat{E}_1'(Q, R_2) \triangleq \min_{\tilde{Q} \in \mathcal{S}(Q), \max[t_0(Q), f(Q)] \leq f(\tilde{Q})} \tilde{I}(\tilde{Q}), \quad \text{(C.5)}$$

$$\hat{E}_1''(Q, R_2) \triangleq \min_{\substack{\tilde{Q} \in \mathcal{S}(Q), \hat{Q} \in \hat{S}(\tilde{Q}), \hat{I}(\hat{Q}) \leq R_2 \\ \max[t_0(Q), f(Q)] \leq f(\hat{Q}) + R_2 - \hat{I}(\hat{Q})}} \tilde{I}(\tilde{Q}), \quad \text{(C.6)}$$

$$\hat{E}_1'''(Q, R_2) \triangleq \min_{\substack{\tilde{Q} \in \mathcal{S}(Q), \hat{Q} \in \hat{S}(\tilde{Q}) \\ t_0(Q), f(Q) \leq f(\hat{Q})}} \left\{ \tilde{I}(\tilde{Q}) + \left[\hat{I}(\hat{Q}) - R_2 \right]_+ \right\}.$$

(C.7)

Using the same arguments as in [21], it can be shown that $\hat{E}_2(Q, R_2)$ can be equivalently be expressed as

$$\hat{E}_2''(Q, R_2) \triangleq \min_{\substack{\tilde{Q} \in \mathcal{S}(Q), \hat{Q} \in \hat{S}(\tilde{Q}) \\ \max[t_0(Q), f(Q)] \leq f(\hat{Q})}} \left[I_{\hat{Q}}(X_2; Y_1) - R_2 \right]_+.$$

(C.8)

Accordingly, from (C.5)-(C.7) and (C.8), we see that (C.3) and (C.4) can be expressed in terms of convex optimization problems, namely, for a given $Q_{Y_1|X_1 X_2}$, the terms $\hat{E}_1(Q, R_2)$ and $\hat{E}_2(Q, R_2)$ (i.e., the inner terms of the minimization problem in (C.1)) can be calculated efficiently, as desired.

Finally, we discuss the computation of (C.1). Generally speaking, the minimization over $Q_{Y_1|X_1 X_2}$ might not be a

$$\hat{E}^{(1)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{2,3,4\}}(Q)} \left[I_{\tilde{Q}}(Z_1; Z_2^4, Y_1) + E_7(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.7}$$

$$\hat{E}^{(2)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{1,3,4\}}(Q)} \left[I_{\tilde{Q}}(Z_2; Z_1, Z_3^4, Y_1) + E_7(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.8}$$

$$\hat{E}_8^{(1)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{2,3,4\}}(Q)} E_6(\tilde{Q}_{Z_1^4 Y}, Q_{Z_1^4 Y_1}), \tag{A.9}$$

$$\hat{E}_8^{(2)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{1,3,4\}}(Q)} E_5(\tilde{Q}_{Z_1^4 Y}, Q_{Z_1^4 Y_1}), \tag{A.10}$$

$$\hat{E}_1^{(3)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{3,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(1)}; Z_3^4, Y_1 | \boldsymbol{Z}_{12 \backslash \mathcal{U}(1)}) + E_5(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y}) \right], \tag{A.11}$$

$$\hat{E}_2^{(3)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{3,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(2)}; Z_3^4, Y_1 | \boldsymbol{Z}_{12 \backslash \mathcal{U}(2)}) + E_6(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y}) \right], \tag{A.12}$$

$$\hat{E}_4^{(3)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{3,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(4)}; Z_3^4, Y_1 | \boldsymbol{Z}_{12 \backslash \mathcal{U}(4)}) + E_7(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y}) \right], \tag{A.13}$$

$$\hat{E}_8^{(3)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{3,4\}}(Q)} E_3(\tilde{Q}_{Z_1^4 Y}, Q_{Z_1^4 Y_1}), \tag{A.14}$$

$$\hat{E}_1^{(4)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{2,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(1)}; Z_2, Z_4, Y_1 | \boldsymbol{Z}_{13 \backslash \mathcal{U}(1)}) + E_4(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.15}$$

$$\hat{E}_3^{(4)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{2,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(3)}; Z_2, Z_4, Y_1 | \boldsymbol{Z}_{13 \backslash \mathcal{U}(3)}) + E_6(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.16}$$

$$\hat{E}_5^{(4)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{2,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(5)}; Z_2, Z_4, Y_1 | \boldsymbol{Z}_{13 \backslash \mathcal{U}(5)}) + E_7(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.17}$$

$$\hat{E}_8^{(4)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{2,4\}}(Q)} E_2(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}), \tag{A.18}$$

$$\hat{E}_2^{(5)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{1,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(2)}; Z_1, Z_4, Y_1 | \boldsymbol{Z}_{23 \backslash \mathcal{U}(2)}) + E_4(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.19}$$

$$\hat{E}_3^{(5)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{1,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(3)}; Z_1, Z_4, Y_1 | \boldsymbol{Z}_{23 \backslash \mathcal{U}(3)}) + E_5(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.20}$$

$$\hat{E}_6^{(5)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{1,4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(6)}; Z_1, Z_4, Y_1 | \boldsymbol{Z}_{23 \backslash \mathcal{U}(6)}) + E_7(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \tag{A.21}$$

$$\hat{E}_8^{(5)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{1,4\}}(Q)} E_1(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}), \tag{A.22}$$

$$\hat{E}_u^{(6)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{4\}}(Q)} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(u)}; Z_4, Y_1 | \boldsymbol{Z}_{123 \backslash \mathcal{U}(u)}) + E_u(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \right], \; u \in \{1, \dots, 7\}, \tag{A.23}$$

$$\hat{E}_8^{(6)}(Q_{Z_1^4 Y_1}, R_{22}) \triangleq \min_{\tilde{Q}: \, \tilde{Q} \in \mathcal{S}_{\{4\}}(Q)} E_0(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}), \tag{A.24}$$

$$E_{\text{HK}}^{(u)}(Q_{Z_1^4 Y_1}) \triangleq \max \left\{ \left[\hat{E}^{(u)}(Q_{Z_1^4 Y_1}, R_{22}) - \mathcal{R}_u \right]_+, \hat{E}_8^{(u)}(Q_{Z_1^4 Y_1}, R_{22}) \right\}, \; u \in \{1, 2\}, \tag{A.25}$$

$$E_{\text{HK}}^{(3)}(Q_{Z_1^4 Y_1}) \triangleq \max \left\{ \max_{u \in \{1,2,4\}} \left[\hat{E}_u^{(3)}(Q_{Z_1^4 Y_1}, R_{22}) - \mathcal{R}_u \right]_+, \hat{E}_8^{(3)}(Q_{Z_1^4 Y_1}, R_{22}) \right\}, \tag{A.26}$$

$$E_{\text{HK}}^{(4)}(Q_{Z_1^4 Y_1}) \triangleq \max \left\{ \max_{u \in \{1,3,5\}} \left[\hat{E}_u^{(4)}(Q_{Z_1^4 Y_1}, R_{22}) - \mathcal{R}_u \right], \hat{E}_8^{(4)}(Q_{Z_1^4 Y_1}, R_{22}) \right\}, \tag{A.27}$$

$$E_{\text{HK}}^{(5)}(Q_{Z_1^4 Y_1}) \triangleq \max \left\{ \max_{u \in \{2,3,6\}} \left[\hat{E}_u^{(5)}(Q_{Z_1^4 Y_1}, R_{22}) - \mathcal{R}_u \right]_+, \hat{E}_8^{(5)}(Q_{Z_1^4 Y_1}, R_{22}) \right\}, \tag{A.28}$$

$$E_{\text{HK}}^{(6)}(Q_{Z_1^4 Y_1}) \triangleq \max \left\{ \max_{u \in \{1:7\}} \left[\hat{E}_u^{(6)}(Q_{Z_1^4 Y_1}, R_{22}) - \mathcal{R}_u \right]_+, \hat{E}_8^{(6)}(Q_{Z_1^4 Y_1}, R_{22}) \right\}, \tag{A.29}$$

$$\tilde{E}_{\text{HK}}(R_{11}, R_{12}, R_{21}, R_{22}) \triangleq \min_{\substack{Q_{Y_1 | Z_1^4}: \\ Q_{Z_i} = P_{Z_i}, \, 1 \leq i \leq 4}} \left[D(Q_{Y_1 | Z_1^4} || W_{Y_1 | Z_1^4} | P_{Z_1^4}) + \min_{u \in \{1:6\}} E_{\text{HK}}^{(u)}(Q_{Z_1^4 Y_1}) \right]. \tag{A.30}$$

convex problem, and thus one should resort to global opti- mization methods (e.g., a simple algorithm is an exhaustive

$$\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{V_1(i)\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1(i),V_{k+1}(j))\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\right\}\right\}$$

$$\leq\min\left\{1,L_1\mathbb{E}\left[\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|V_1\right\}\right\}\right]\right.,$$

$$\left.L_2\mathbb{E}\left[\min\left\{1,L_1\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|\{V_k\}_{k=2}^{K}\right\}\right\}\right]\right\}\tag{B.1}$$

$$\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{V_1(i)\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1(i),V_{k+1}(j))\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\right\}\right\}$$

$$\leq L_1\Pr\left\{\bigcup_{j}\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1}(j))\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\right\}\right\}$$

$$\leq L_1\mathbb{E}\left\{\Pr\left\{\bigcup_{j}\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1}(j))\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\right\}\Bigg|V_1\right\}\right\}\tag{B.2}$$

$$\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{V_1(i)\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1(i),V_{k+1}(j))\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\right\}\right\}$$

$$\leq L_1\mathbb{E}\left[\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|V_1\right\}\right\}\right]\tag{B.3}$$

$$\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|V_1\right\}\right\}$$

$$=\mathcal{I}\{\mathcal{F}\}\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|V_1\right\}\right\}$$

$$+\mathcal{I}\{\mathcal{F}^c\}\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|V_1\right\}\right\}$$

$$=\mathcal{I}\{\mathcal{F}\}\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|V_1\right\}\right\}$$

$$\leq\mathcal{I}\{\mathcal{F}\}\tag{B.4}$$

$$L_1\mathbb{E}\left[\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\Bigg|V_1\right\}\right\}\right]$$

$$\leq L_1L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1,V_{k+1})\in\mathcal{A}_{l,k},\ \bigcap_{k=2}^{K}V_k\in\mathcal{G}_{l,k}\right\}\right\}\tag{B.5}$$

---

search over a fine grid of probability simplex). Nonetheless, in the following we somewhat simplify these optimizations.

We first see that (C.1) can be rewritten as

$$\tilde{E}_1(R_1,R_2)\triangleq\min\left\{\tilde{E}_1'(R_1,R_2),\tilde{E}_1''(R_1,R_2)\right\}\tag{C.9}$$

where

$$\tilde{E}_1'(R_1, R_2) = \inf_{Q:\, R_1 < \hat{E}_1(Q, R_2) - \hat{E}_2(Q, R_2)} \left\{ D(Q\|W) + \hat{E}_1(Q, R_2) - R_1 \right\}, \quad \text{(C.10)}$$

and

$$\tilde{E}_1''(R_1, R_2) = \inf_{Q:\, R_1 \geq \hat{E}_1(Q, R_2) - \hat{E}_2(Q, R_2)} \left\{ D(Q\|W) + \hat{E}_2(Q, R_2) \right\}. \quad \text{(C.11)}$$

Let us analyze $\tilde{E}_1'(R_1, R_2)$ as a function of R_1. For $R_1 = 0$, we have

$$\tilde{E}_1'(0, R_2) = \inf_Q \left\{ D(Q\|W) + \hat{E}_1(Q, R_2) \right\}. \quad \text{(C.12)}$$

Now, letting the minimizer be $Q^*_{Y_1|X_1 X_2}\big|_{R_1=0}$, and defining the critical rate $R_{1,\text{crit}} = \hat{E}_1\big(Q^*_{Y_1|X_1 X_2}\big|_{R_1=0}, R_2 \big) - \hat{E}_2\big(Q^*_{Y_1|X_1 X_2}\big|_{R_1=0}, R_2 \big)$, it is easily noticed that for $R_1 \leq R_{1,\text{crit}}$, the exponent is an affine function

$$\tilde{E}_1'(R_1, R_2) = \tilde{E}_1'(0, R_2) - R_1. \quad \text{(C.13)}$$

Furthermore, for $R_1 \leq R_{1,\text{crit}}$, it is readily seen that

$$\tilde{E}_1(R_1, R_2) = \tilde{E}_1'(R_1, R_2). \quad \text{(C.14)}$$

For $R_1 > R_{1,\text{crit}}$, however, since the optimization of Q is not convex, the term cannot be simplified anymore.

APPENDIX D
PROOF OF (10)

First, note that the following region:

$$R_1 < I_W(X_1; Y_1|X_2), \quad \text{(D.1)}$$
$$R_1 < I_W(X_1; Y_1) + [I_W(X_2; Y_1|X_1) - R_2]_+, \quad \text{(D.2)}$$

evaluated with $P_{X_1 X_2 Y_1} = P_{X_1} \times P_{X_2} \times W_{Y_1|X_1 X_2}$, is equivalent to (10). Thus, to show that (10) is achievable, it suffices to show that the above region is achievable. Now, recall that the ordinary random coding exponent, in the single-user setting, is given by

$$E_r(R) = \min_Q \left\{ D(Q_{Y|X}\|W|P_X) + [I_Q(X;Y) - R]_+ \right\}.$$

From [25, Lemma 9] it can be shown that $E_r(R)$ can be rewritten as

$$E_r(R) = \min_{(Q, \tilde{Q}) \in \mathcal{D}:\, f(\tilde{Q}) \leq f(Q)} \left\{ D(\tilde{Q}_{Y|X}\|W|P_X) + [I_Q(X;Y) - R]_+ \right\}, \quad \text{(D.3)}$$

where $f(Q) = \mathbb{E}_Q \{ \log W(Y|X) \}$, and $\mathcal{D}$ is the set of $(Q, \tilde{Q})$ distributions such that $Q_Y = \tilde{Q}_Y$, and $Q_X = \tilde{Q}_X = P_X$. The last representation is very similar, in some sense, to the error exponent formula in Theorem 1. It can be seen that $E_r(R)$ is positive as long as

$$R < \min_{Q:\, Q_Y = W_Y, Q_X = P_X, f(W) \leq f(Q)} I_Q(X;Y). \quad \text{(D.4)}$$

Obviously, we should get that $R < I_W(X;Y)$, namely, the minimum in (D.4) should be equal to $I_W(X;Y)$. To see that this is indeed the case, note that since the above optimization problem is convex, the linear constraint is met with equality, and we note that $f(W) = -H_W(Y|X)$, $f(Q) = -D(Q\|W|P_X) - H_Q(Y|X)$, and $I_Q(X;Y) = H_W(Y) - H_Q(Y|X)$. Using the last facts, we get

$$R < \min_{Q:\, Q_Y = W_Y, Q_X = P_X, f(W) = f(Q)} I_Q(X;Y)$$
$$= \min_{\substack{Q:\, Q_Y = W_Y, Q_X = P_X, \\ H_W(Y|X) = D(Q\|W|P_X) + H_Q(Y|X)}} H_W(Y) - H_Q(Y|X)$$
$$= I_W(X;Y)$$
$$\quad + \min_{\substack{Q:\, Q_Y = W_Y, Q_X = P_X, \\ H_W(Y|X) = D(Q\|W|P_X) + H_Q(Y|X)}} D(Q\|W|P_X)$$
$$= I_W(X;Y),$$

as required.

In our case, using the equivalent representation of our error exponent in (C.5)-(C.7), we readily get that the error exponent in Theorem 1 is positive if[6]:

$$R_1 < \min_{\tilde{Q} \in \mathcal{S}(W), \max[t_0(W), f(W)] \leq f(\tilde{Q})} \tilde{I}(\tilde{Q}), \quad \text{(D.5)}$$

$$R_1 < \min_{\substack{\tilde{Q} \in \mathcal{S}(W), \hat{Q} \in \hat{S}(\tilde{Q}), \hat{I}(\hat{Q}) \leq R_2 \\ \max[t_0(W), f(W)] \leq f(\hat{Q}) + R_2 - \hat{I}(\hat{Q})}} \tilde{I}(\tilde{Q}), \quad \text{(D.6)}$$

$$R_1 < \min_{\substack{\tilde{Q} \in \mathcal{S}(W), \hat{Q} \in \hat{S}(\tilde{Q}) \\ \max[t_0(W), f(W)] \leq f(\hat{Q})}} \left\{ \tilde{I}(\tilde{Q}) + \left[\hat{I}(\hat{Q}) - R_2 \right]_+ \right\} \quad \text{(D.7)}$$

where we recall that $\tilde{I}(\tilde{Q}) \equiv I_{\tilde{Q}}(X_1; X_2, Y_1)$ and $\hat{I}(\hat{Q}) \equiv I_{\hat{Q}}(X_2; X_1, Y_1)$. In the following, we show that (D.5) and (D.7) correspond to (D.1) and (D.2), respectively. Finally, we show that (D.5) is dominated by (D.5) and (D.7), and thus superfluous. Indeed, for (D.5), we have

$$\min_{\tilde{Q} \in \mathcal{S}(W), \max[t_0(W), f(W)] \leq f(\tilde{Q})} \tilde{I}(\tilde{Q})$$
$$= \min_{\tilde{Q}_{Y|X_2} = W_{Y|X_2}, \max[t_0(W), f(W)] \leq f(\tilde{Q})} I_{\tilde{Q}}(X_1; X_2, Y)$$
$$\geq \min_{\tilde{Q}_{Y|X_2} = W_{Y|X_2}, f(W) \leq f(\tilde{Q})} I_{\tilde{Q}}(X_1; X_2, Y). \quad \text{(D.8)}$$

Now, as before, we note that (using the fact that the minimization over $\tilde{Q}$ in (D.8) is such that $\tilde{Q}_{Y|X_2} = W_{Y|X_2}$)

$$I_{\tilde{Q}}(X_1; X_2, Y) = I_{\tilde{Q}}(X_1; X_2) + I_{\tilde{Q}}(X_1; Y|X_2) \quad \text{(D.9)}$$
$$= I_{\tilde{Q}}(X_1; X_2) + H_W(Y|X_2)$$
$$\quad - H_{\tilde{Q}}(Y|X_1, X_2), \quad \text{(D.10)}$$
$$f(W) = -H_W(Y|X_1, X_2), \quad \text{(D.11)}$$

and

$$f(\tilde{Q}) = -D(\tilde{Q}\|W|P_X) - H_{\tilde{Q}}(Y|X_1, X_2). \quad \text{(D.12)}$$

Thus, we have (D.13), shown at the top of the next page, where

---

[6]To show that (10) is achievable, we consider a lower bound on $\tilde{E}(R_1, R_2)$, which ignores the contribution of $\hat{E}_2(Q, R_2)$, namely, $\min_Q \left\{ D(Q\|W|P_X) + \left[\hat{E}_1(Q, R_2) - R_1 \right]_+ \right\}$.

$$\min_{\tilde{Q}_{Y|X_2}=W_{Y|X_2},\, f(W)\leq f(\tilde{Q})} I_{\tilde{Q}}(X_1;X_2,Y)$$

$$= \min_{\substack{\tilde{Q}_{Y|X_2}=W_{Y|X_2},\\ D(\tilde{Q}||W|P_X)+H_{\tilde{Q}}(Y|X_1,X_2)\leq H_W(Y|X_1,X_2)}} I_{\tilde{Q}}(X_1;X_2)+H_W(Y|X_2)-H_{\tilde{Q}}(Y|X_1,X_2)$$

$$\geq I_W(X_1;Y|X_2)+\min_{\substack{\tilde{Q}_{Y|X_2}=W_{Y|X_2},\\ D(\tilde{Q}||W|P_X)+H_{\tilde{Q}}(Y|X_1,X_2)\leq H_W(Y|X_1,X_2)}} I_{\tilde{Q}}(X_1;X_2)+D(\tilde{Q}||W|P_X)$$

$$= I_W(X_1;Y|X_2) \tag{D.13}$$

---

the inequality follows from the fact that $-H_{\tilde{Q}}(Y|X_1,X_2) \geq D(\tilde{Q}||W|P_X) - H_W(Y|X_1,X_2)$ induced by the optimization constraint, and the last equality is achieved by taking $\tilde{Q} = W$ and $\tilde{Q}_{X_1,X_2} = P_{X_1}P_{X_2}$. The constraint in (D.7) is handled in a similar manner. Indeed, using the same manipulations, we get (D.14), shown at the top of the next page, where the inequality is due to the fact that $-H_{\tilde{Q}}(Y|X_1,X_2) \geq D(\tilde{Q}||W|P_X) - H_W(Y|X_1,X_2)$, the second equality follows by taking $\hat{Q} = W$, and the last equality follows by taking $\tilde{Q}$ such that $X_2 - Y - X_1$ is a Markov chain.

Finally, we show that the constraint in (D.6) is superfluous. To this end, we will show that for $R_2 < I_W(X_2;Y|X_1)$, the r.h.s. of (D.6) reduces to $R_1 + R_2 < I(X_1,X_2;Y)$, which is dominated by (D.5) and (D.7) (or, equivalently, by (D.13) and (D.14)), and for $R_2 \geq I_W(X_2;Y|X_1)$, (D.6) reduces to $R_1 < I_W(X_1;Y)$, already supported by (D.7) (see (D.14)). Whence, (D.6) is redundant. Indeed, for $R_2 \geq I_W(X_2;Y|X_1)$, the r.h.s. of (D.6) can be lower bounded as in (D.15), presented at the top of the next page, where in (a) we use the definition of $t_0(W)$ in (8b), (b) follows from the assumption that $R_2 \geq I_W(X_2;Y|X_1)$, (c) is due to (D.11)-(D.12), (d) follows from the fact that $-H_{\tilde{Q}}(Y|X_1) \geq D(\hat{Q}||W|P_X) + I_{\hat{Q}}(X_1;X_2) - H_W(Y|X_1)$ induced by the optimization constraint, and (e) is achieved by taking $\hat{Q} = W$, $\hat{Q}_{X_1X_2} = \hat{Q}_{X_1}\hat{Q}_{X_2}$, and $\tilde{Q}$ such that $X_2 - Y - X_1$ is a Markov chain. Thus, for $R_2 \geq I_W(X_2;Y|X_1)$, we obtained that $R_1 < I_W(X_1;Y)$, as required. On the other hand, for $R_2 < I_W(X_2;Y|X_1)$, the r.h.s. of (D.6) can be lower bounded as shown in (D.16), given in the next page, where (a) is due to (D.11)-(D.12), (b) is because $\hat{Q} \in \mathcal{S}(\tilde{Q})$ and thus $H_{\tilde{Q}}(Y|X_1) = H_{\hat{Q}}(Y|X_1)$, and (c) follows from the fact that $-H_{\hat{Q}}(Y|X_1) \geq D(\hat{Q}||W|P_X) + I_{\hat{Q}}(X_1;X_2) - H_W(Y|X_1,X_2) - R_2$ induced by the optimization constraint. Whence, we obtained that $R_1 + R_2 \leq I_W(X_1,X_2;Y)$, as required.

ACKNOWLEDGMENT

The authors would like to thank the associate editor, Jun Chen, and the anonymous referees for their suggestions and comments which helped improving the content of this paper. The authors are also grateful to Jonathan Scarlett for very helpful comments on a previous version of this paper, and for drawing our attention to [21] (and its extended version which

is not published yet), which was unknown to the authors at the time of writing the original version of this paper.

REFERENCES

[1] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[2] E. C. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1. Berkeley, CA: Univ. California Press, 1961, pp. 611–64.
[3] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Annals Probabil.*, vol. 2, no. 5, pp. 805–814, 1974.
[4] A. B. Carleial, "Interference channels," *IEEE Trans. on Inf. Theory*, vol. IT-24, pp. 60–70, Jan. 1978.
[5] H. Sato, "Two-user communication channels," *IEEE Trans. on Inf. Theory*, vol. IT-23, pp. 295–304, May. 1977.
[6] A. B. Carleial, "A case where intereference does not reduce capacity," *IEEE Trans. on Inf. Theory*, vol. IT-21, pp. 569–570, Sep. 1975.
[7] R. Benzel, "The capacity region of a class of discrete additive degraded interference channels," *IEEE Trans. on Inf. Theory*, vol. IT-25, pp. 228–231, Mar. 1979.
[8] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. on Inf. Theory*, vol. IT, no. 27, pp. 49–60, Jan. 1981.
[9] R. Etkin, N. Merhav, and E. Ordentlich, "Error exponents of optimum decoding for the interference channel," *IEEE Trans. on Inf. Theory*, vol. 56, no. 1, pp. 40–56, Jan. 2010.
[10] C. Chang, R. Etkin, and E. Ordentlich, "Interference channel capacity region for randomized fixed-composition codes," *HP Labs Technical Report*, 2009. [Online]. Available: http://www.hpl.hp.com/techreports/2008/HPL-2008-194R1.html
[11] J. Scarlett, A. Martinez, and A. G. i Fàbregas, "Multiuser coding techniques for mismatched decoding," *submitted to IEEE Trans. on Inf. Theory*, Nov. 2013. [Online]. Available: arxiv.org/pdf/1311.6635
[12] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[13] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 2011.
[14] N. Merhav, "Relations between random coding exponents and the statistical physics of random codes," *IEEE Trans. on Inf. Theory*, vol. 55, no. 1, pp. 83–92, Jan. 2009.
[15] ——, "Exact random coding exponents of optimal bin index decoding," *IEEE Trans. on Inf. Theory*, vol. 60, no. 10, pp. 6024–6031, Oct. 2014.
[16] ——, "List decoding-random coding exponents and expurgated exponents," *IEEE Trans. on Inf. Theory*, vol. 60, no. 11, pp. 6749–6759, Nov. 2014.
[17] ——, "Exact correct-decoding exponent for the wiretap channel decoder," *IEEE Trans. on Inf. Theory*, vol. 60, no. 12, pp. 7606–7615, Dec. 2014.
[18] W. Huleihel, N. Weinberger, and N. Merhav, "Erasure/list random coding error exponents are not universally achievable," *submitted to IEEE Trans. on Inf. Theory*, Dec. 2015. [Online]. Available: http://arxiv.org/abs/1410.7005
[19] N. Merhav, "The generalized random energy model and its application to the statistical physics of ensembles of hierarchical codes," *IEEE Trans. on Inf. Theory*, vol. 55, no. 3, pp. 1250–1268, May 2009.
[20] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel-Aviv University, 2003, http://citeseerx.ist.psu.edu/viewdoc.

$$\min_{\substack{\tilde{Q}_{Y|X_2}=W_{Y|X_2},\hat{Q}_{Y|X_1}=\tilde{Q}_{Y|X_1} \\ D(\hat{Q}||W|P_X)+H_{\hat{Q}}(Y|X_1,X_2)\leq H_W(Y|X_1,X_2)}} \left\{ \tilde{I}(\tilde{Q}) + \left[I_{\hat{Q}}(X_1;X_2) + H_{\tilde{Q}}(Y|X_1) - H_{\hat{Q}}(Y|X_1,X_2) - R_2 \right]_+ \right\}$$

$$\geq \min_{\substack{\tilde{Q}_{Y|X_2}=W_{Y|X_2},\hat{Q}_{Y|X_1}=\tilde{Q}_{Y|X_1} \\ D(\hat{Q}||W|P_X)+H_{\hat{Q}}(Y|X_1,X_2)\leq H_W(Y|X_1,X_2)}} \left\{ \tilde{I}(\tilde{Q}) + \left[I_{\hat{Q}}(X_1;X_2) + D(\hat{Q}||W|P_X) + H_{\tilde{Q}}(Y|X_1) \right. \right.$$

$$\left. \left. - H_W(Y|X_1,X_2) - R_2 \right]_+ \right\}$$

$$= \min_{\tilde{Q}_{Y|X_2}=W_{Y|X_2},\hat{Q}_{Y|X_1}=W_{Y|X_1}} \left\{ I_{\tilde{Q}}(X_1;Y) + I_{\tilde{Q}}(X_1;X_2|Y) + \left[H_{\tilde{Q}}(Y|X_1) - H_W(Y|X_1,X_2) - R_2 \right]_+ \right\}$$

$$= \min_{\tilde{Q}_{Y|X_2}=W_{Y|X_2},\hat{Q}_{Y|X_1}=W_{Y|X_1}} \left\{ I_W(X_1;Y) + I_{\tilde{Q}}(X_1;X_2|Y) + \left[H_W(Y|X_1) - H_W(Y|X_1,X_2) - R_2 \right]_+ \right\}$$

$$= I_W(X_1;Y) + \left[I_W(X_2;Y|X_2) - R_2 \right]_+ , \tag{D.14}$$

$$\min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ \max[t_0(W),f(W)]\leq f(\hat{Q})+R_2-\hat{I}(\hat{Q})}} \tilde{I}(\tilde{Q}) \geq \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ t_0(W)\leq f(\hat{Q})+R_2-\hat{I}(\hat{Q})}} \tilde{I}(\tilde{Q})$$

$$\overset{(a)}{=} \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ \max_{\tilde{Q}:\ \tilde{Q}\in\mathcal{S}(W),\hat{I}(\hat{Q})\leq R_2}\left[f(\hat{Q})-I_{\hat{Q}}(X_2;X_1,Y)\right]\leq f(\hat{Q})-I_{\hat{Q}}(X_2;X_1,Y)}} I_{\tilde{Q}}(X_1;X_2,Y)$$

$$\overset{(b)}{\geq} \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ f(W)-I_W(X_2;X_1,Y)\leq f(\hat{Q})-I_{\hat{Q}}(X_2;X_1,Y)}} I_{\tilde{Q}}(X_1;X_2,Y)$$

$$\overset{(c)}{=} \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ -H_W(Y|X_1)\leq -D(\hat{Q}||W|P_X)-I_{\hat{Q}}(X_1;X_2)-H_{\tilde{Q}}(Y|X_1)}} H_W(Y) - H_{\tilde{Q}}(Y|X_1) + I_{\tilde{Q}}(X_1;X_2|Y)$$

$$\overset{(d)}{\geq} I_W(X_1;Y) + \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ -H_W(Y|X_1)\leq -D(\hat{Q}||W|P_X)-I_{\hat{Q}}(X_1;X_2)-H_{\tilde{Q}}(Y|X_1)}} D(\hat{Q}||W|P_X) + I_{\hat{Q}}(X_1;X_2) + I_{\tilde{Q}}(X_1;X_2|Y)$$

$$\overset{(e)}{=} I_W(X_1;Y) \tag{D.15}$$

$$\min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ \max[t_0(W),f(W)]\leq f(\hat{Q})+R_2-\hat{I}(\hat{Q})}} \tilde{I}(\tilde{Q}) \geq \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ f(W)\leq f(\hat{Q})+R_2-\hat{I}(\hat{Q})}} \tilde{I}(\tilde{Q})$$

$$\overset{(a)}{=} \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ -H_W(Y|X_1,X_2)\leq -D(\hat{Q}||W|P_X)-H_{\hat{Q}}(Y|X_1)-I_{\hat{Q}}(X_1;X_2)+R_2}} H_W(Y) - H_{\tilde{Q}}(Y|X_1) + I_{\tilde{Q}}(X_1;X_2|Y)$$

$$\overset{(b)}{=} \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ -H_W(Y|X_1,X_2)\leq -D(\hat{Q}||W|P_X)-H_{\hat{Q}}(Y|X_1)-I_{\hat{Q}}(X_1;X_2)+R_2}} H_W(Y) - H_{\hat{Q}}(Y|X_1) + I_{\tilde{Q}}(X_1;X_2|Y)$$

$$\overset{(c)}{\geq} I_W(X_1,X_2;Y) - R_2$$

$$+ \min_{\substack{\tilde{Q}\in\mathcal{S}(W),\hat{Q}\in\hat{S}(\tilde{Q}),\hat{I}(\hat{Q})\leq R_2 \\ -H_W(Y|X_1,X_2)\leq -D(\hat{Q}||W|P_X)-H_{\hat{Q}}(Y|X_1)-I_{\hat{Q}}(X_1;X_2)+R_2}} D(\hat{Q}||W|P_X) + I_{\hat{Q}}(X_1;X_2) + I_{\tilde{Q}}(X_1;X_2|Y)$$

$$\geq I_W(X_1,X_2;Y) - R_2 \tag{D.16}$$

[21] J. Scarlett, A. Martinez, and A. G. i Fábregas, "Mismatched multi-letter successive decoding for the multiple-access channel," in *Proc. ISIT 2014*, Jul. 2014, pp. 2539–2543.

[22] J. Scarlett, "*Reliable Communication Under Mismatched Decoding*," Ph.D. dissertation, University of Cambridge, 2014, http://itc.upf.edu/biblio/1061.

[23] B. Bandemer, A. El Gamal, and Y. K. Kim, "Optimal achievable rates for interference networks with random codes," *IEEE Trans. on Inf. Theory*, vol. 61, no. 12, pp. 6536–6549, Dec. 2015.

[24] L. Weng, S. S. Pradhan, and A. Anastasopoulos, "Error exponent regions for Gaussian broadcast and multiple-access channels," *IEEE Trans. on Inf. Theory*, vol. 54, no. 7, pp. 2919–2942, July 2008.

[25] R. Gallager, "Fixed composition arguments and lower bounds to error probability," July 1994. [Online]. Available: www.rle.mit.edu/rgallager/documents/notes5.pdf