Mismatched Multi-letter Successive Decoding for the Multiple-Access Channel

Jonathan Scarlett
University of Cambridge
jms265@cam.ac.uk

Alfonso Martinez
Universitat Pompeu Fabra
alfonso.martinez@ieee.org

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—This paper studies channel coding for the discrete memoryless multiple-access channel with a given decoding rule. A multi-letter successive decoding rule depending on an arbitrary non-negative function $q(x_1, x_2, y)$ is considered, and an achievable rate region and error exponent are derived. The rate region is compared with that of the maximum-metric decoder which uses the function $q(x_1, x_2, y)$, and a numerical example is given for which successive decoding yields a strictly higher sum rate for a given pair of input distributions.**

I. Introduction

The mismatched decoding problem [1]–[3] seeks to characterize the performance of channel coding when the decoding rule is fixed and possibly suboptimal (e.g. due to channel uncertainty or implementation constraints). Extensions of this problem to multiuser settings are not only of interest in their own right, but can also provide valuable insight into the single-user setting [3]–[5]. In particular, significant attention has been paid to the mismatched multiple-access channel (MAC), for which, given the length-n output vector $\boldsymbol{y}$ and codebooks $\mathcal{C}_\nu = \{\boldsymbol{x}_\nu^{(1)}, \cdots, \boldsymbol{x}_\nu^{(M_\nu)}\}$ ($\nu = 1, 2$), the decoder estimates the message pair as

$$(\hat{m}_1, \hat{m}_2) = \arg\max_{(i,j)} q^n(\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(j)}, \boldsymbol{y}), \tag{1}$$

where $q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \triangleq \prod_{i=1}^n q(x_{1,i}, x_{2,i}, y_i)$ for some non-negative decoding metric $q(x_1, x_2, y)$.

Given that the decoder only knows the metric $q^n(\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(j)}, \boldsymbol{y})$ corresponding to each codeword pair, one may question whether there exists a decoding rule with better performance than the maximum-metric rule in (1). In general, this question is only interesting if "reasonable" decoding rules are considered. For example, if the values $\{\log q(x_1, x_2, y)\}$ are rationally independent (i.e. no value can be written as linear combinations of the others with rational coefficients), then there is a one-to-one correspondence between the joint empirical distribution of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ and the possible values of $q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$, and hence the decoder can implement the maximum-likelihood (ML) rule (assuming the channel is memoryless).

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

In this paper, we consider successive decoding of the form

$$\hat{m}_1 = \arg\max_i \sum_j q^n(\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(j)}, \boldsymbol{y}) \tag{2}$$

$$\hat{m}_2 = \arg\max_j q^n(\boldsymbol{x}_1^{(\hat{m}_1)}, \boldsymbol{x}_2^{(j)}, \boldsymbol{y}). \tag{3}$$

Stated formally, we have the following. Let $W(y|x_1, x_2)$ be the transition law of a memoryless MAC, and let $q(x_1, x_2, y)$ be an arbitrary non-negative function. The alphabets are denoted by $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{Y}$, and each is assumed to be finite. Encoder $\nu = 1, 2$ takes as input m_ν equiprobable on $\{1, \cdots, M_\nu\}$, and transmits the corresponding codeword $\boldsymbol{x}_\nu^{(m_\nu)}$ from a codebook $\mathcal{C}_\nu$. We say that a rate pair (R_1, R_2) is achievable if, for all $\delta > 0$, there exist sequences of codebooks $\mathcal{C}_{1,n}$ and $\mathcal{C}_{2,n}$ with $M_1 \ge e^{n(R_1-\delta)}$ and $M_2 \ge e^{n(R_2-\delta)}$ respectively, such that $\mathbb{P}[(\hat{m}_1, \hat{m}_2) \ne (m_1, m_2)] \to 0$ under the decoding rule described by (2)–(3).

Letting $\mathcal{E}_\nu \triangleq \{\hat{m}_\nu \ne m_\nu\}$ for $\nu = 1, 2$, we observe that if $q(x_1, x_2, y) = W(y|x_1, x_2)$, then (2) is the decision rule which minimizes $\mathbb{P}[\mathcal{E}_1]$. That is, (2) is a mismatched version of the optimal decoding rule for (one user of) the interference channel (IC). Thus, as well as giving an achievable rate region for the MAC with mismatched successive decoding, our results will quantify the loss due to mismatch for the IC. In particular, we obtain an achievable error exponent using different techniques to those of [6].

It can be shown that the exponents and rates with $q = W$ coincide with those of ML decoding (i.e. (1) with $q = W$); this is done by noting that (2) minimizes $\mathbb{P}[\mathcal{E}_1]$, (3) minimizes the probability of favoring some (m_1, j) ($j \ne m_2$) over (m_1, m_2), and (1) minimizes $\mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2]$. In contrast, we will see that when $q \ne W$, the successive decoder can lead to significantly different rate regions to those of maximum-metric decoding.

Notation: Bold symbols are used for vectors (e.g. $\boldsymbol{x}$), and the corresponding i-th entry is written using a subscript (e.g. x_i). Subscripts are used to denote the distributions corresponding to expectations and mutual informations (e.g. $\mathbb{E}_P[\cdot]$, $I_P(X;Y)$). The marginals of a joint distribution P_{XY} are denoted by P_X and P_Y. We write $P_X = \widetilde{P}_X$ to denote element-wise equality between two probability distributions on the same alphabet. The set of all sequences of length n with a given empirical distribution P_X (i.e. type [7, Ch. 2]) is denoted by $T^n(P_X)$. We write $f(n) \doteq g(n)$ if $\lim_{n\to\infty} \frac{1}{n} \log \frac{f(n)}{g(n)} =$

0, and similarly for $\dot{\leq}$ and $\dot{\geq}$. We write $[\alpha]^+ = \max(0, \alpha)$, and denote the indicator function by $\mathbb{1}\{\cdot\}$

II. Main Result

We fix the input distributions Q_1 and Q_2, let $P_{X_1X_2Y} \triangleq Q_1 \times Q_2 \times W$, and define the functions

$$\overline{F}(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}, R_2) \triangleq \max\Big\{\mathbb{E}_{\widetilde{P}}[\log q(X_1,X_2,Y)], \mathbb{E}_{\widetilde{P}'}[\log q(X_1,X_2,Y)] + \big[R_2 - I_{\widetilde{P}'}(X_2;X_1,Y)\big]^+\Big\}, \quad (4)$$

$$\underline{F}(P_{X_1X_2Y}, R_2) \triangleq \max\Big\{\mathbb{E}_P[\log q(X_1,X_2,Y)], \max_{P'_{X_1X_2Y} \in \mathcal{T}'_1(P_{X_1X_2Y}, R_2)} \mathbb{E}_{P'}[\log q(X_1,X_2,Y)] + R_2 - I_{P'}(X_2;X_1,Y)\Big\}, \quad (5)$$

and the sets

$$\mathcal{T}_1(P_{X_1X_2Y}, R_2) \triangleq \Big\{(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}) : \widetilde{P}_{X_2Y} = P_{X_2Y}, \widetilde{P}_{X_1} = P_{X_1}, \widetilde{P}'_{X_1Y} = \widetilde{P}_{X_1Y}, P'_{X_2} = P_{X_2}, \overline{F}(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}, R_2) \geq \underline{F}(P_{X_1X_2Y}, R_2)\Big\} \quad (6)$$

$$\mathcal{T}'_1(P_{X_1X_2Y}, R_2) \triangleq \Big\{P'_{X_1X_2Y} : P'_{X_1Y} = P_{X_1Y}, P'_{X_2} = P_{X_2}, I_{P'}(X_2;X_1,Y) \leq R_2\Big\} \quad (7)$$

$$\mathcal{T}_2(P_{X_1X_2Y}) \triangleq \Big\{\widetilde{P}_{X_1X_2Y} : \widetilde{P}_{X_2} = P_{X_2}, \widetilde{P}_{X_1Y} = P_{X_1Y}, \mathbb{E}_{\widetilde{P}}[\log q(X_1,X_2,Y)] \geq \mathbb{E}_P[\log q(X_1,X_2,Y)]\Big\}. \quad (8)$$

Theorem 1. *For any input distributions Q_1 and Q_2, the pair (R_1, R_2) is achievable for the decoder in* (2)–(3) *provided that*

$$R_1 \leq \min_{(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1(P_{X_1X_2Y}, R_2)} I_{\widetilde{P}}(X_1;X_2,Y) + \big[I_{\widetilde{P}'}(X_2;X_1,Y) - R_2\big]^+ \quad (9)$$

$$R_2 \leq \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_2(P_{X_1X_2Y})} I_{\widetilde{P}}(X_2;X_1,Y). \quad (10)$$

Proof: See Section III. ■

The minimization in (9) is a non-convex optimization problem, but it can be cast in terms of convex optimization problems; see the Appendix for details. While our focus is on achievable rates, the proof of Theorem 1 reveals that the error exponent corresponding to (10) coincides with one of the three error events for maximum-metric decoding [5], and the error exponent corresponding to (9) is given by

$$\min_{P_{X_1X_2Y} : P_{X_1}=Q_1, P_{X_2}=Q_2} D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times W) + \big[I_0(P_{X_1X_2Y}, R_2) - R_1\big]^+, \quad (11)$$

where $I_0(P_{X_1X_2Y}, R_2)$ denotes the right-hand side of (9) with an arbitrary distribution $P_{X_1X_2Y}$ used in (5)–(8) (rather than $P_{X_1X_2Y} = Q_1 \times Q_2 \times W$).

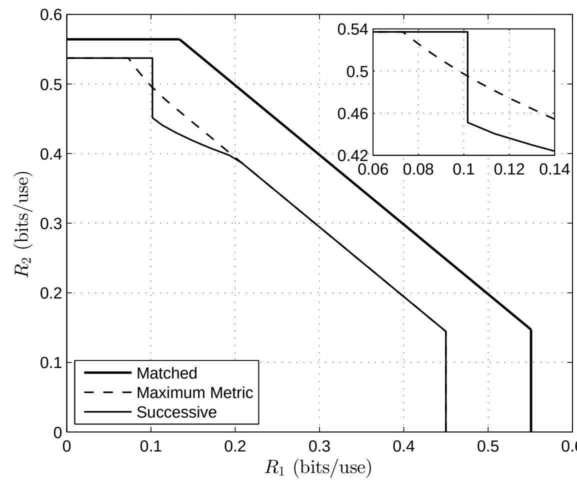

Figure 1. Achievable rate regions for the channel given in (12).

A. Numerical Example

We consider the MAC with $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1\}$, $\mathcal{Y} = \{0,1,2\}$, and

$$W(y|x_1,x_2) = \begin{cases} 1 - 2\eta_{x_1x_2} & y = x_1 + x_2 \\ \eta_{x_1x_2} & \text{otherwise}, \end{cases} \quad (12)$$

where $\{\eta_{x_1x_2}\}$ are constants. The mismatched decoder uses $q(x_1,x_2,y)$ of a similar form, but a fixed value η in place of $\{\eta_{x_1x_2}\}$. We set $\eta_{00} = 0.01$, $\eta_{01} = 0.1$, $\eta_{10} = 0.01$, $\eta_{11} = 0.3$, $\eta = 0.15$, and $Q_1 = Q_2 = (0.5, 0.5)$. Figure 1 plots the achievable rates regions of successive decoding (Theorem 1), maximum-metric decoding (see [3], [5]), and matched decoding (yielding the same region whether successive or maximum-metric).

Interestingly, neither of the mismatched rate regions dominates the other, thus suggesting that the two decoding rules are fundamentally different. For the given input distribution, the sum rate for successive decoding exceeds that of maximum-metric decoding. Furthermore, upon taking the convex hull (which is justified by a time sharing argument [3], [8]), the region for successive decoding is strictly larger. While we observed similar behaviors for other choices of Q_1 and Q_2, it remains unclear as to whether this always holds. Furthermore, while the rate region for maximum-metric decoding is tight with respect to the ensemble average [3], it is unclear whether the same is true for that of successive decoding.

The vertical line at $R_1 \approx 0.1$ is analogous to the interference channel, where for R_1 below a certain threshold, R_2 can take any value while still ensuring user 1's message is estimated correctly [6]. Due to the mismatch, this induces a non-pentagonal shape in the present example.

III. Proof of Theorem 1

Our analysis is based on the method of type class enumeration (e.g. see [6], [9], [10]), and is perhaps most similar to that of Somekh-Baruch and Merhav [10]. We consider constant-composition random coding, where for $\nu = 1, 2$ we have

$$P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu) = \frac{1}{|T^n(Q_\nu)|} \mathbb{1}\big\{\boldsymbol{x}_\nu \in T^n(Q_\nu)\big\}. \quad (13)$$

Here we assume that Q_1 and Q_2 are types for notational convenience; more generally, we can approximate these by types and the analysis is unchanged. The (independent) random codewords are denoted by $(\boldsymbol{X}_\nu^{(1)}, \cdots, \boldsymbol{X}_\nu^{(M_\nu)})$. We assume without loss of generality that $m_1 = m_2 = 1$, and we write $\boldsymbol{X}_\nu = \boldsymbol{X}_\nu^{(1)}$ and let $\overline{\boldsymbol{X}}_\nu$ denote an arbitrary $\boldsymbol{X}_\nu^{(j)}$ with $j \neq 1$. The output sequence is denoted by $\boldsymbol{Y}$, and we write $R_\nu \triangleq \frac{1}{n}\log M_\nu$ ($\nu = 1, 2$).

As noted by Grant *et al.* [11], we can analyze the error probability of the second decoding step (see (3)) assuming that no error occurred on the first step (see (2)), while still using the unconditional statistics of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$. The subsequent analysis has been done in the study of maximum-metric decoding [3], [5], and the corresponding rate condition is precisely (10). In the remainder of this section, we focus on the first decoding step.

Let $\overline{p}_{e,1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ denote the random-coding error probability for the first decoding step conditioned on $(\boldsymbol{X}_1^{(1)}, \boldsymbol{X}_2^{(1)}, \boldsymbol{Y}) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$. The joint type of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ is denoted by $P_{X_1X_2Y}$.[1] We write the objective in (2) as

$$\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{x}}_1) \triangleq q^n(\overline{\boldsymbol{x}}_1, \boldsymbol{x}_2, \boldsymbol{y}) + \sum_{j\neq 1} q^n(\overline{\boldsymbol{x}}_1, \boldsymbol{X}_2^{(j)}, \boldsymbol{y}), \tag{14}$$

which is random due to the randomness of $\{\boldsymbol{X}_2^{(j)}\}$. Using the union bound, we have

$$\overline{p}_{e,1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \leq (M_1 - 1)\mathbb{P}\big[\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{X}}_1) \geq \Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\boldsymbol{x}_1)\big]. \tag{15}$$

We proceed by analyzing the statistics of $\Xi_{\boldsymbol{x}_2\boldsymbol{y}}$. From (14),

$$\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{x}}_1) = q^n(\widetilde{P}_{X_1X_2Y}) + \sum_{\widetilde{P}'_{X_1X_2Y}} N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) q^n(\widetilde{P}'_{X_1X_2Y}), \tag{16}$$

where $\widetilde{P}_{X_1X_2Y}$ is the joint type of $(\overline{\boldsymbol{x}}_1, \boldsymbol{x}_2, \boldsymbol{y})$, $N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y})$ is the random number of $\boldsymbol{X}_2^{(j)}$ ($j \neq 1$) such that $(\overline{\boldsymbol{x}}_1, \boldsymbol{X}_2^{(j)}, \boldsymbol{y}) \in T^n(\widetilde{P}'_{X_1X_2Y})$, and we write $q^n(\widetilde{P}'_{X_1X_2Y}) \triangleq q^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y})$ for an arbitrary triplet $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}) \in T^n(\widetilde{P}'_{X_1X_2Y})$. Since the codewords are generated independently, $N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y})$ is binomially distributed with $M_2 - 1$ trials and success probability $\mathbb{P}\big[(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{y}) \in T^n(\widetilde{P}'_{X_1X_2Y})\big]$. By construction, we have $N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) = 0$ unless $\widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y})$, where

$$\mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y}) \triangleq \Big\{\widetilde{P}'_{X_1X_2Y} \,:\, \widetilde{P}'_{X_1Y} = \widetilde{P}_{X_1Y}, \widetilde{P}'_{X_2} = Q_2\Big\}. \tag{17}$$

The following lemma characterizes the behavior of $N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y})$ for fixed R_2 and $\widetilde{P}'_{X_1X_2Y}$. The proof can be found in [6], [10], and is based on the fact that

$$\mathbb{P}\big[(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{y}) \in T^n(\widetilde{P}'_{X_1X_2Y})\big] \doteq e^{-nI_{\widetilde{P}'}(X_2;X_1,Y)}. \tag{18}$$

Roughly speaking, the lemma states that if $R_2 > I_{\widetilde{P}'}(X_2; X_1, Y)$ then the corresponding type enumerator is highly concentrated about its mean, whereas if $R_2 < I_{\widetilde{P}'}(X_2; X_1, Y)$ then the type enumerator takes a subexponential value (possibly zero) with overwhelming probability.

[1]This is a slight abuse of notation in light of the previous definition $P_{X_1X_2Y} = Q_1 \times Q_2 \times W$, but this substitution will be made later.

Lemma 1. [6], [10] *Fix the pair* $(\overline{\boldsymbol{x}}_1, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1Y})$*, a constant* $\delta > 0$*, and a type* $\widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y})$*.*

(i) If $R_2 \geq I_{\widetilde{P}'}(X_2; X_1, Y) + \delta$*, then*

$$M_2 e^{-n(I_{\widetilde{P}'}(X_2;X_1,Y)+\delta)} \leq N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) \tag{19}$$

$$\leq M_2 e^{-n(I_{\widetilde{P}'}(X_2;X_1,Y)-\delta)} \tag{20}$$

with probability approaching one super-exponentially fast.

(ii) If $R_2 < I_{\widetilde{P}'}(X_2; X_1, Y) + \delta$*, then*

$$N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) \leq e^{-n\,2\delta} \tag{21}$$

with probability approaching one super-exponentially fast.

Given a joint type $\widetilde{P}_{X_1X_2Y}$, let $\mathcal{A}_\delta(\widetilde{P}_{X_1X_2Y})$ denote the event that the high-probability events in Lemma 1 occur for all $\widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y})$. Since $\mathbb{P}[\mathcal{A}_\delta(\widetilde{P}_{X_1X_2Y})] \to 1$ super-exponentially fast, we can safely condition any event on $\mathcal{A}_\delta(\widetilde{P}_{X_1X_2Y})$ without changing the exponential behavior of the corresponding probability.

Conditioned on $\mathcal{A}_\delta(P_{X_1X_2Y})$, we have the following:

$$\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\boldsymbol{x}_1) = q^n(P_{X_1X_2Y}) + \sum_{P'_{X_1X_2Y}} N_{\boldsymbol{x}_1\boldsymbol{y}}(P'_{X_1X_2Y}) q^n(P'_{X_1X_2Y}) \tag{22}$$

$$\geq q^n(P_{X_1X_2Y}) + \max_{\substack{P'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, P_{X_1Y}) \\ R_2 \geq I_{P'}(X_2;X_1,Y)+\delta}} N_{\boldsymbol{x}_1\boldsymbol{y}}(P'_{X_1X_2Y}) q^n(P'_{X_1X_2Y}) \tag{23}$$

$$\geq q^n(P_{X_1X_2Y}) + \max_{\substack{P'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, P_{X_1Y}) \\ R_2 \geq I_{P'}(X_2;X_1,Y)+\delta}} M_2 e^{-n(I_{P'}(X_2;X_1,Y)+\delta)} q^n(P'_{X_1X_2Y}) \tag{24}$$

$$\triangleq \underline{F}_\delta(P_{X_1X_2Y}), \tag{25}$$

where (24) follows from part (i) of Lemma 1. Unlike $\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\boldsymbol{x}_1)$, the quantity $\underline{F}_\delta(P_{X_1X_2Y})$ is deterministic. Substituting (25) into (15), we obtain

$$\overline{p}_{e,1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \,\dot{\leq}\, M_1 \mathbb{P}\big[\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{X}}_1) \geq \underline{F}_\delta(P_{X_1X_2Y})\big]. \tag{26}$$

Since the statistics of $\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{x}}_1)$ depend on $\overline{\boldsymbol{x}}_1$ only through the joint type of $(\overline{\boldsymbol{x}}_1, \boldsymbol{x}_2, \boldsymbol{y})$, we can write (26) as follows:

$$\overline{p}_{e,1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \,\dot{\leq}\, M_1 \sum_{\widetilde{P}_{X_1X_2Y}} \mathbb{P}\big[(\overline{\boldsymbol{X}}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1X_2Y})\big] \times \mathbb{P}\big[\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{x}}_1) \geq \underline{F}_\delta(P_{X_1X_2Y})\big] \tag{27}$$

$$\doteq M_1 \max_{\widetilde{P}_{X_1X_2Y} \in \mathcal{S}_1(Q_1, P_{X_2Y})} e^{-nI_{\widetilde{P}}(X_1;X_2,Y)} \times \mathbb{P}\big[\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{x}}_1) \geq \underline{F}_\delta(P_{X_1X_2Y})\big], \tag{28}$$

where $\overline{\boldsymbol{x}}_1$ denotes an arbitrary sequence such that $(\overline{\boldsymbol{x}}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1X_2Y})$, and

$$\mathcal{S}_1(Q_1, P_{X_2Y}) \triangleq \Big\{\widetilde{P}_{X_1X_2Y} \,:\, \widetilde{P}_{X_1} = Q_1, \widetilde{P}_{X_2Y} = P_{X_2Y}\Big\}. \tag{29}$$

In (28), we have used an analogous property to (18).

Next, we again use Lemma 1 in order to replace $\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{x}}_1)$ in (28) by a deterministic quantity. We have from (16) that

$$\Xi_{\boldsymbol{x}_2\boldsymbol{y}}(\overline{\boldsymbol{x}}_1) \le q^n(\widetilde{P}_{X_1X_2Y}) + p_0(n) \max_{\widetilde{P}'_{X_1X_2Y}} N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) q^n(\widetilde{P}'_{X_1X_2Y}), \tag{30}$$

where $p_0(n)$ is a polynomial corresponding to the total number of joint types. Substituting (30) into (28), we obtain

$$\overline{p}_{e,1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \,\dot{\le}\, M_1 \max_{\widetilde{P}_{X_1X_2Y} \in \mathcal{S}_1(Q_1, P_{X_2Y})} \max_{\widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y})} e^{-nI_{\widetilde{P}}(X_1;X_2,Y)} \mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y})\big], \tag{31}$$

where

$$\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y}) \triangleq \Big\{q^n(\widetilde{P}_{X_1X_2Y}) + p_0(n) N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) q^n(\widetilde{P}'_{X_1X_2Y}) \ge \underline{F}_\delta(P_{X_1X_2Y})\Big\}, \tag{32}$$

and we have used the union bound to take the maximum over $\widetilde{P}'_{X_1X_2Y}$ outside the probability in (31). Continuing, we have

$$\max_{\widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y})} \mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y})\big] = \max\Bigg\{ \max_{\substack{\widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y}) \\ R_2 \ge I_{\widetilde{P}'}(X_2;X_1,Y)+\delta}} \mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y})\big], \max_{\substack{\widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y}) \\ R_2 < I_{\widetilde{P}'}(X_2;X_1,Y)+\delta}} \mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y})\big]\Bigg\}. \tag{33}$$

For the first maximization in (33), observe that conditioned on $\mathcal{A}_\delta(\widetilde{P}_{X_1X_2Y})$ (defined following Lemma 1), we have for $\widetilde{P}'_{X_1X_2Y}$ satisfying $R_2 \ge I_{\widetilde{P}'}(X_2;X_1,Y) + \delta$ that

$$N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) q^n(\widetilde{P}'_{X_1X_2Y}) \le M_2 e^{-n(I_{\widetilde{P}'}(X_2;X_1,Y)-\delta)} q^n(\widetilde{P}'_{X_1X_2Y}). \tag{34}$$

Hence, and using Lemma 1, we have

$$\mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y})\big] \,\dot{\le}\, \mathbb{1}\Big\{q^n(\widetilde{P}_{X_1X_2Y}) + M_2 p_0(n) e^{-n(I_{\widetilde{P}'}(X_2;X_1,Y)-\delta)} q^n(\widetilde{P}'_{X_1X_2Y}) \ge \underline{F}_\delta(P_{X_1X_2Y})\Big\}. \tag{35}$$

For the second maximization in (33), we define the event $\mathcal{B} \triangleq \big\{N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) > 0\big\}$, yielding

$$\mathbb{P}[\mathcal{B}] \,\dot{\le}\, M_2 e^{-nI_{\widetilde{P}'}(X_2;X_1,Y)}, \tag{36}$$

which follows from the union bound and the identity in (18). Whenever $R_2 < I_{\widetilde{P}'}(X_2; X_1, Y) + \delta$, we have

$$\begin{aligned}
&\mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y})\big] \\
&\le \mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y}) \,\big|\, \mathcal{B}^c\big] + \mathbb{P}[\mathcal{B}]\mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y}) \,\big|\, \mathcal{B}\big] \qquad (37)\\
&\,\dot{\le}\, \mathbb{1}\big\{q^n(\widetilde{P}_{X_1X_2Y}) \ge \underline{F}_\delta(P_{X_1X_2Y})\big\} \\
&\qquad + M_2 e^{-nI_{\widetilde{P}'}(X_2;X_1,Y)} \mathbb{P}\big[\mathcal{E}_{P,\widetilde{P}}(\widetilde{P}'_{X_1X_2Y}) \,\big|\, \mathcal{B}\big], \qquad (38)\\
&\,\dot{\le}\, \mathbb{1}\big\{q^n(\widetilde{P}_{X_1X_2Y}) \ge \underline{F}_\delta(P_{X_1X_2Y})\big\} + M_2 e^{-nI_{\widetilde{P}'}(X_2;X_1,Y)} \\
&\quad \times \mathbb{1}\big\{q^n(\widetilde{P}_{X_1X_2Y}) + p_0(n)e^{-n2\delta} q^n(\widetilde{P}'_{X_1X_2Y}) \ge \underline{F}_\delta(P_{X_1X_2Y})\big\}, \qquad (39)
\end{aligned}$$

where (38) follows from (36) and since $\mathcal{B}^c$ implies $N_{\overline{\boldsymbol{x}}_1\boldsymbol{y}}(\widetilde{P}'_{X_1X_2Y}) = 0$, and (39) uses part (ii) of Lemma 1.

Observe that $\underline{F}(P_{X_1X_2Y}, R_2)$ in (5) is obtained from $\underline{F}_\delta$ in (25) in the limit as $\delta \to 0$. Similarly, the exponents corresponding to the other quantities appearing in the indicator functions in (35) and (39) tend toward the following:

$$\overline{F}_1(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}, R_2) \triangleq \max\Big\{\mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)], \mathbb{E}_{\widetilde{P}'}[\log q(X_1, X_2, Y)] + R_2 - I_{\widetilde{P}'}(X_2; X_1, Y)\Big\} \tag{40}$$

$$\overline{F}_2(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}) \triangleq \max\Big\{\mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)], \mathbb{E}_{\widetilde{P}'}[\log q(X_1, X_2, Y)]\Big\}. \tag{41}$$

Combining (31), (33), (35) and (39) with these expressions, taking $\delta \to 0$, and using the continuity of the underlying terms in the optimizations, we obtain

$$\overline{p}_{e,1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \,\dot{\le}\, \max\Bigg\{ \max_{(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(1)}(P_{X_1X_2Y}, R_2)} M_1 e^{-nI_{\widetilde{P}}(X_1;X_2,Y)}, \max_{(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(2')}(P_{X_1X_2Y}, R_2)} M_1 e^{-nI_{\widetilde{P}}(X_1;X_2,Y)}, \max_{(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(2)}(P_{X_1X_2Y}, R_2)} M_1 e^{-nI_{\widetilde{P}}(X_1;X_2,Y)} \times M_2 e^{-nI_{\widetilde{P}'}(X_2;X_1,Y)} \Bigg\}, \tag{42}$$

where[2]

$$\mathcal{T}_1^{(1)}(P_{X_1X_2Y}, R_2) \triangleq \Big\{(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}) \,:\, \widetilde{P}_{X_1X_2Y} \in \mathcal{S}_1(Q_1, P_{X_2Y}), \widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2, \widetilde{P}_{X_1Y}), I_{\widetilde{P}'}(X_2; X_1, Y) \le R_2, \overline{F}_1(\widetilde{P}_{X_1X_2Y}, \widetilde{P}'_{X_1X_2Y}, R_2) \ge \underline{F}(P_{X_1X_2Y}, R_2)\Big\} \tag{43}$$

[2]Strictly speaking, these sets depend on (Q_1, Q_2), but this dependence need not be explicit, since we have $P_{X_1} = Q_1$ and $P_{X_2} = Q_2$.

$$\mathcal{T}_1^{(2')}(P_{X_1X_2Y},R_2) \triangleq \Big\{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \,:\, \widetilde{P}_{X_1X_2Y} \in \mathcal{S}_1(Q_1,P_{X_2Y}), \widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2,\widetilde{P}_{X_1Y}), \mathbb{E}_{\widetilde{P}}[\log q(X_1,X_2,Y)] \ge \underline{F}(P_{X_1X_2Y},R_2)\Big\} \quad (44)$$

$$\mathcal{T}_1^{(2)}(P_{X_1X_2Y},R_2) \triangleq \Big\{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \,:\, \widetilde{P}_{X_1X_2Y} \in \mathcal{S}_1(Q_1,P_{X_2Y}), \widetilde{P}'_{X_1X_2Y} \in \mathcal{S}'_1(Q_2,\widetilde{P}_{X_1Y}), I_{\widetilde{P}'}(X_2;X_1,Y) \ge R_2, \overline{F}_2(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \ge \underline{F}(P_{X_1X_2Y},R_2)\Big\}. \quad (45)$$

The three terms in the maximization in (42) respectively correspond to (35) and the two terms in (39).

Since $\overline{F}_1 \ge \mathbb{E}_{\widetilde{P}}[\log q]$, we see that $\mathcal{T}_1^{(1)} \subseteq \mathcal{T}_1^{(2')}$, and hence the second term in the outer maximum of (42) can be removed. Furthermore, we can safely substitute $P_{X_1X_2Y} = Q_1 \times Q_2 \times W$, since $P_{X_1X_2Y} \to Q_1 \times Q_2 \times W$ with probability approaching one by the law of large numbers. We thus obtain the following rate conditions for the first decoding step:

$$R_1 \le \min_{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(1)}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y) \quad (46)$$

$$R_1 + R_2 \le \min_{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(2)}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y) + I_{\widetilde{P}'}(X_2;X_1,Y). \quad (47)$$

Finally, using the definitions of $\overline{F}$, $\mathcal{S}_1$, $\mathcal{S}'_1$, $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_1^{(2)}$ (see (4), (17), (29), (43) and (45)) to unite (46)–(47) yields (9).

Appendix

Here we write (9) in terms of convex optimization problems, starting with the alternative expression in (46)–(47). We first note that (47) holds if and only if

$$R_1 \le \min_{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(2)}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y) + \big[I_{\widetilde{P}'}(X_2;X_1,Y) - R_2\big]^+, \quad (48)$$

due to the constraint $I_{\widetilde{P}'}(X_2;X_1,Y) \ge R_2$. Next, we claim that when combining (46) and (48), the rate region is unchanged if the constraint $I_{\widetilde{P}'}(X_2;X_1,Y) \ge R_2$ is omitted from (48). To see this, note that for $I_{\widetilde{P}'}(X_2;X_1,Y) < R_2$, the objective in (48) coincides with that of (46). The desired result follows from the identity $\overline{F}_1 > \overline{F}_2$ (using (40)–(41) and the assumption $I_{\widetilde{P}'}(X_2;X_1,Y) < R_2$), implying that (46) is more restrictive.

We now deal with the non-concavity of $\overline{F}_1$ and $\overline{F}_2$. Using the identity

$$\min_{x \le \max\{a,b\}} f(x) = \min\Big\{\min_{x \le a} f(x), \min_{x \le b} f(x)\Big\}, \quad (49)$$

we obtain the following rate conditions from (46) and (48):

$$R_1 \le \min_{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(1,1)}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y) \quad (50)$$

$$R_1 \le \min_{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(1,2)}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y) \quad (51)$$

$$R_1 \le \min_{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(2,1)}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y) + \big[I_{\widetilde{P}'}(X_2;X_1,Y) - R_2\big]^+ \quad (52)$$

$$R_1 \le \min_{(\widetilde{P}_{X_1X_2Y},\widetilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1^{(2,2)}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y) + \big[I_{\widetilde{P}'}(X_2;X_1,Y) - R_2\big]^+, \quad (53)$$

where for $k = 1,2$ and $l = 1,2$, $\mathcal{T}_1^{(k.l)}$ is defined in the same way as $\mathcal{T}_1^{(k)}$ with the following modifications: (i) The constraint $\overline{F}_k \ge \underline{F}$ is changed so that the left-hand side contains the l-th term in the maximization in $\overline{F}_k$ (see (40)–(41)); (ii) For $k = 2$, the constraint $I_{\widetilde{P}'}(X_2;X_1,Y) \ge R_2$ is removed, in accordance with the discussion following (48).

The variable $\widetilde{P}'_{X_1X_2Y}$ can be removed from both (50) and (52), since in both cases the choice $\widetilde{P}'_{X_1X_2Y}(x_1,x_2,y) = P_{X_2}(x_2)\widetilde{P}_{X_1Y}(x_1,y)$ is feasible and yields an objective of $I_{\widetilde{P}}(X_1;X_2,Y)$. It follows that (50) and (52) yield the same value, and we conclude that (9) can equivalently be expressed in terms of three conditions: (51), (53), and

$$R_1 \le \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_1^{(1,1')}(P_{X_1X_2Y},R_2)} I_{\widetilde{P}}(X_1;X_2,Y), \quad (54)$$

where $\mathcal{T}_1^{(1,1')}$ is defined in the same way as $\mathcal{T}_1^{(1,1)}$ with the variable $\widetilde{P}'_{X_1X_2Y}$ removed. These three conditions are all written as convex optimization problems, as desired.

References

[1] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 35–43, Jan. 1995.

[2] A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[3] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sept. 1996.

[4] A. Somekh-Baruch, "On achievable rates for channels with mismatched decoding," 2013, submitted to *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1305.0547].

[5] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Multiuser coding techniques for mismatched decoding," 2013, submitted to *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1311.6635].

[6] R. Etkin, N. Merhav, and E. Ordentlich, "Error exponents of optimum decoding for the interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 40–56, 2010.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[8] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[9] N. Merhav, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4439–4447, Oct. 2008.

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

[11] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 873–890, 2001.