Multiuser Coding Techniques
for Mismatched Decoding

Jonathan Scarlett, Alfonso Martinez and Albert Guillén i Fàbregas

Abstract

This paper studies multiuser coding techniques for channel coding with a given (possibly suboptimal) decoding rule. For the mismatched discrete memoryless multiple-access channel, error exponents are obtained which are tight with respect to the ensemble average, and positive within the interior of Lapidoth's achievable rate region. In the special case of maximum-likelihood decoding, the ensemble tightness of the exponent of Liu-Hughes is proved. Alternative expressions for the error exponents and rate regions are given, including expressions obtained using Lagrange duality which extend immediately to general alphabets.

In the setting of single-user mismatched decoding, similar analysis techniques are applied to two types of superposition coding. The standard version is shown to yield an achievable rate which is at least as high as that of Lapidoth's expurgated parallel coding rate after the optimization of the parameters. A refined version of superposition coding is shown to achieve rates at least as good as the standard version for any set of random-coding parameters, and it is shown that the gap between the two can be significant when the input distribution is fixed.

Index Terms

Mismatched decoding, multiple-access channels, superposition coding, random coding, error exponents, ensemble tightness, duality, maximum-likelihood decoding.

I. Introduction

The mismatched decoding problem [1]–[9] seeks to characterize the performance of point-to-point communication when the decoding rule is fixed and possibly suboptimal. This problem is of interest, for example, when the optimal decoding rule is infeasible due to channel uncertainty or implementation constraints. Finding a single-letter expression for the mismatched capacity (i.e. the highest achievable rate with mismatched decoding) remains an open problem even for single-user discrete memoryless channels.

Network information theory problems with mismatched decoding are not only of independent interest, but can also provide valuable insight into the single-user mismatched decoding problem. In particular, it was shown by Lapidoth that higher random-coding rates can be obtained by treating the single-user channel as a multiple-access channel (MAC) and generating multiple codebooks in parallel [6].

In this paper, we build on the work of [6] and study multiuser coding techniques for channels with mismatched decoding. We revisit the mismatched MAC, and provide a derivation of the achievable rate region given in [6] which yields non-asymptotic bounds and ensemble-tight error exponents as intermediate steps. Furthermore, an alternative

J. Scarlett is with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: jmscarlett@gmail.com). A. Martinez is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: alfonso.martinez@ieee.org). A. Guillén i Fàbregas is with the Institució Catalana de Recerca i Estudis Avançats (ICREA), the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

expression for the region is given which extends immediately to general alphabets. Building on the analysis techniques developed for the MAC, we study two types of superposition coding for single-user mismatched channels, and provide the achievable rates and error exponents corresponding to each.

A. System Setup

Throughout the paper, we consider both the mismatched single-user channel and the mismatched multiple-access channel. Here we provide a detailed description of each.

1) Mismatched Single-User Channel: The input and output alphabets are denoted by $\mathcal{X}$ and $\mathcal{Y}$ respectively, and the channel transition law is denoted by $W(y|x)$, thus yielding an n-letter transition law given by

$$W^n(\boldsymbol{y}|\boldsymbol{x}) \triangleq \prod_{i=1}^n W(y_i|x_i). \tag{1}$$

If $\mathcal{X}$ and $\mathcal{Y}$ are finite, the channel is referred to as a discrete memoryless channel (DMC). We consider length-n block coding, in which a codebook $\mathcal{C} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}\}$ is known at both the encoder and decoder. The encoder takes as input a message m uniformly distributed on the set $\{1, \ldots, M\}$, and transmits the corresponding codeword $\boldsymbol{x}^{(m)}$. The decoder receives the vector $\boldsymbol{y}$ at the output of the channel, and forms the estimate

$$\hat{m} = \underset{j \in \{1,\ldots,M\}}{\arg\max} \ q^n(\boldsymbol{x}, \boldsymbol{y}), \tag{2}$$

where n is the length of each codeword, $x_i^{(j)}$ is the i-th entry of $\boldsymbol{x}^{(j)}$ (similarly for y_i), and $q^n(\boldsymbol{x}, \boldsymbol{y}) \triangleq \prod_{i=1}^n q(x_i, y_i)$. The function $q(x, y)$ is called the *decoding metric*, and is assumed to be non-negative. In the case of a tie, a codeword achieving the maximum in (2) is selected uniformly at random. In the case that $q(x, y) = W(y|x)$, the decoding rule in (2) is that of optimal maximum-likelihood (ML) decoding.

A rate R is said to be achievable if, for all $\delta > 0$, there exists a sequence of codebooks with at least $\exp(n(R - \delta))$ codewords and vanishing error probability under the decoding metric q. The mismatched capacity of a given channel and metric is defined to be the supremum of all achievable rates.

An error exponent $E(R)$ is said to be achievable if there exists a sequence of codebooks $\mathcal{C}_n$ with $M \geq \exp(nR)$ codewords of length n such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log p_e(\mathcal{C}_n) \geq E(R). \tag{3}$$

We let $\bar{p}_e(n, M)$ denote the average error probability with respect to a given random-coding ensemble which will be clear from the context. The random-coding error exponent $E_r(R)$ is said to exhibit ensemble tightness if

$$\lim_{n\to\infty} -\frac{1}{n}\log \bar{p}_e(n, e^{nR}) = E_r(R). \tag{4}$$

This definition requires that the limit exists.

2) Mismatched Multiple-Access Channel: We also consider a 2-user memoryless MAC $W(y|x_1, x_2)$ with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. In the case that each alphabet is finite, the MAC is referred to as a discrete memoryless MAC (DM-MAC). The decoding metric is denoted by $q(x_1, x_2, y)$, and we write $W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \triangleq \prod_{i=1}^n W(y_i|x_{1,i}, x_{2,i})$ and $q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \triangleq \prod_{i=1}^n q(x_{1,i}, x_{2,i}, y_i)$.

Each encoder $\nu = 1, 2$ takes as input a message m_ν uniformly distributed on the set $\{1, \ldots, M_\nu\}$, and transmits the corresponding codeword $\boldsymbol{x}_\nu^{(m_\nu)}$ from the codebook $\mathcal{C}_\nu = \{\boldsymbol{x}_\nu^{(1)}, \ldots, \boldsymbol{x}_\nu^{(M_\nu)}\}$. Upon receiving the signal $\boldsymbol{y}$ at the output of the channel, the decoder forms an estimate $(\hat{m}_1, \hat{m}_2)$ of the messages, given by

$$(\hat{m}_1, \hat{m}_2) = \underset{i \in \{1,\ldots,M_1\}, j \in \{1,\ldots,M_2\}}{\arg\max} q^n(\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(j)}, \boldsymbol{y}). \tag{5}$$

We assume that ties are broken at random. An error is said to have occurred if the estimate $(\hat{m}_1, \hat{m}_2)$ differs from (m_1, m_2). The error probability for a given pair of codebooks $(\mathcal{C}_1, \mathcal{C}_2)$ is denoted by $p_e(\mathcal{C}_1, \mathcal{C}_2)$, and the error probability for a given random-coding ensemble is denoted by $\overline{p}_e(n, M_1, M_2)$. We define achievable rates, error exponents and ensemble tightness similarly to the single-user setting. Similarly to the single-user case, optimal ML decoding is recovered by setting $q(x_1, x_2, y) = W(y|x_1, x_2)$.

B. Notation

We use bold symbols for vectors (e.g. $\boldsymbol{x}$), and denote the corresponding i-th entry using a subscript (e.g. x_i). All logarithms have base e, and all rates are in units of nats except in the examples, where bits are used. We define $[c]^+ = \max\{0, c\}$, and denote the indicator function by $\mathbb{1}\{\cdot\}$.

The set of all probability distributions on an alphabet, say $\mathcal{X}$, is denoted by $\mathcal{P}(\mathcal{X})$, and the set of all empirical distributions on a vector in $\mathcal{X}^n$ (i.e. types [10, Ch. 2], [11]) is denoted by $\mathcal{P}_n(\mathcal{X})$. Similar notations $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $\mathcal{P}_n(\mathcal{Y}|\mathcal{X})$ are used for conditional distributions. For a given $Q \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(Q)$ is defined to be the set of all sequences in $\mathcal{X}^n$ with type Q. For a given joint type $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and sequence $\boldsymbol{x} \in T^n(P_X)$, the conditional type class $T^n_{\boldsymbol{x}}(P_{XY})$ is defined to be the set of all sequences $\boldsymbol{y}$ such that $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_{XY})$. Further definitions and results related to the method of types are presented in Appendix A.

The probability of an event is denoted by $\mathbb{P}[\cdot]$. The marginals of a joint distribution $P_{XY}(x, y)$ are denoted by $P_X(x)$ and $P_Y(y)$. We write $P_X = \widetilde{P}_X$ to denote element-wise equality between two probability distributions on the same alphabet. Expectation with respect to a joint distribution $P_{XY}(x, y)$ is denoted by $\mathbb{E}_P[\cdot]$, or simply $\mathbb{E}[\cdot]$ when the associated probability distribution is understood from the context. Similarly, mutual information with respect to P_{XY} is written as $I_P(X; Y)$, or simply $I(X; Y)$ when the distribution is understood from the context. Given a distribution $Q(x)$ and conditional distribution $W(y|x)$, we write $Q \times W$ to denote the joint distribution defined by $Q(x)W(y|x)$.

For two positive sequences f_n and g_n, we write $f_n \doteq g_n$ if $\lim_{n \to \infty} \frac{1}{n} \log \frac{f_n}{g_n} = 0$, $f_n \dot{\leq} g_n$ if $\limsup_{n \to \infty} \frac{1}{n} \log \frac{f_n}{g_n} \leq 0$, and similarly for $\dot{\geq}$. When these symbols are used, we assume that the implicit subexponential factors can be upper and lower bounded by universal subexponential quantities; when we consider the discrete memoryless case these quantities may depend on the input and output alphabet sizes. In particular, this assumption implies that $\sup_i f_i(n) \doteq \sup_i g_i(n)$ for any (possibly infinite) set of sequences such that $f_i(n) \doteq g_i(n)$ for all i. We make use of the standard asymptotic notations $O(\cdot)$, $o(\cdot)$ and $\Omega(\cdot)$.

C. Previous Work

The majority of the literature on mismatched decoding has focused on achievable rates for the single-user setting. The most notable early works are by Hui [1] and Csiszár and Körner [2], who independently derived the achievable rate known as the LM rate, given by

$$I_{\mathrm{LM}}(Q) \triangleq \min_{\substack{\widetilde{P}_{XY} : \widetilde{P}_X = Q, \widetilde{P}_Y = P_Y \\ \mathbb{E}_{\widetilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]}} I_{\widetilde{P}}(X; Y), \tag{6}$$

where $P_{XY} = Q \times W$. This rate can equivalently be expressed as [4]

$$I_{\mathrm{LM}}(Q) = \sup_{s \geq 0, a(\cdot)} \mathbb{E} \left[\log \frac{q(X, Y)^s e^{a(X)}}{\mathbb{E}[q(\overline{X}, Y)^s e^{a(\overline{X})} \,|\, Y]} \right], \tag{7}$$

where $(X, Y) \sim Q \times W$.

In the terminology of Ganti *et al.* [7], (6) is the primal expression and (7) is the dual expression. This is consistent with the fact that the equivalence between the two can be proved using Lagrange duality [4], [12]. The primal expression can be proved using constant-composition random coding and the method of types. On the other hand,

the dual expression can be proved using cost-constrained random-coding without relying on the method of types, thus generalizing immediately to general alphabets [7].

Another rate which as received attention in the literature is the generalized mutual information (GMI), defined similarly to (6)–(7) but without the constraint $\widetilde{P}_X = Q$ in (6), and with the supremum over $a(\cdot)$ replaced by $a(\cdot) = 0$ in (7). The GMI is obtained using i.i.d. random coding, and is in general strictly smaller than the LM rate.

A matching converse to the LM rate for binary-input DMCs was given in [5]. In general, however, the LM rate is not tight [3], [6], and the problem of finding the mismatched capacity is open in general. Motivated by the lack of converse results, the concept of ensemble tightness has been addressed in [4], [7], [8]. It has been shown that, for any DMC, the LM rate is the best rate possible for the constant-composition and cost-constrained random-coding ensembles, and the GMI is the best rate possible for the i.i.d. ensemble. Such results are often referred to as random-coding converses.

In [3], Csiszár and Narayan showed that better achievable rates can be obtained by applying the LM rate to the channel $W^{(2)}\big((y_1, y_2)|(x_1, x_2)\big) = W(y_1|x_1)W(y_2|x_2)$ with the metric $q^{(2)}\big((x_1, x_2), (y_1, x_2)\big) = q(x_1, y_1)q(x_2, y_2)$, and similarly for the k-th order products of W and q. It was conjectured that the resulting rate approaches the mismatched capacity as $k \to \infty$, and this conjecture has recently been reported true by Somekh-Baruch [13]

Random-coding error exponents for mismatched decoding were given in [8], [14], [15], and ensemble tightness was addressed in [8]. In particular, it was shown in [8] that for any DMC and metric, the ensemble-tight exponent for the constant-composition ensemble can be obtained using the cost-constrained ensemble with at most two cost constraints.

The mismatched MAC was considered by Lapidoth [6], who obtained an achievable rate region and showed the surprising fact that the single-user LM rate can be improved by treating the single-user channel as a MAC. Lapidoth also addressed the issue of ensemble tightness, showing that random-coding error probability tends to one for rate pairs outside the given achievable rate region.

As an example, Lapidoth considered the channel in Figure 1 consisting of two parallel binary symmetric channels (BSCs) with crossover probabilities $\delta_1 < 0.5$ and $\delta_2 < 0.5$. The mismatched decoder assumes that both crossover probabilities are equal to $\delta < 0.5$. By treating the channel as a mismatched single-user channel from (x_1, x_2) to (y_1, y_2) and using random coding with a uniform distribution on the quaternary input alphabet, one can only achieve rates R satisfying

$$R \le 2\left(1 - H_2\left(\frac{\delta_1 + \delta_2}{2}\right)\right) \qquad \text{bits/use}, \tag{8}$$

where $H_2(\cdot)$ is the binary entropy function in bits. On the other hand, by treating the channel as a mismatched MAC from x_1 and x_2 to (y_1, y_2) and using random coding with equiprobable input distributions on each binary input alphabet, one can achieve any sum-rate R satisfying

$$R \le \big(1 - H_2(\delta_1)\big) + \big(1 - H_2(\delta_2)\big) \qquad \text{bits/use}. \tag{9}$$

This is the best rate possible even under maximum-likelihood (ML) decoding.

In a work which developed independently of ours, Somekh-Baruch [9] gave error exponents and rate regions for the cognitive MAC (i.e. the MAC where one user knows both messages and the other only knows its own) using two multiuser coding schemes: superposition coding and random binning. When applied to single-user mismatched channels, these yield achievable rates which can improve on those of Lapidoth when certain auxiliary variables are fixed. There is some overlap in our work and the superposition coding results given in [9], but throughout the paper we focus primarily on results which are being presented for the first time.

Error exponents for the matched MAC were presented in [16]–[20]. In particular, Liu and Hughes [19] derived an achievable exponent using constant-composition codes, and showed that in contrast to the single-user setting, the exponent can improve over that of i.i.d. random codes [17] even after the optimization of the input distributions. To our knowledge, ensemble tightness has not been proved previously for these ensembles, even in the matched case.
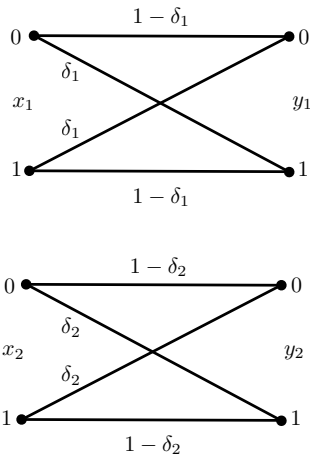
Figure 1.   Parallel BSC example [6].

While error exponents for the mismatched MAC are presented in [20], they are not tight enough to prove the achievability of Lapidoth's rate region. For example, for the parallel BSC example in Figure 1 with uniform input distributions, the exponents of [20] are only positive when the sum-rate R satisfies (8).

D. Contributions

The main contributions of this paper are as follows:

- In Section II, we present a number of general upper and lower bounds on the probability of a multiply-indexed union. We give sufficient conditions under which the upper and lower bounds coincide to within a constant factor, thus providing a tool to perform a tight analysis of the random-coding error probability of multiuser random-coding ensembles. Furthermore, we present generic equivalences of optimization problems using Lagrange duality [12], thus providing the tools to obtain alternative expressions for various rates and exponents.

- In Section III, we study the mismatched MAC. Using constant-composition random coding and the above-mentioned union bounds, we provide an alternative derivation of Lapidoth's achievable rate region [6] which yields non-asymptotic bounds and ensemble-tight error exponents as intermediate steps. By specializing our exponents to ML decoding, we prove the ensemble tightness of the exponent given in [19] for constant-composition random coding.

- Using the method of Lagrange duality [12], we provide alternative expressions for the achievable rate region and error exponent for the DM-MAC. We show that the dual expressions can be obtained directly using cost-constrained random coding, thus proving their validity in the case of general alphabets. Furthermore, we study the number of costs needed for the cost-constrained ensemble to match the performance of constant-composition codes in the discrete memoryless setting.

- Using similar techniques to those used in our MAC analysis, we study two types of superposition coding for single-user mismatched channels. We show that the standard version of superposition coding (Section IV) yields a rate at least as high as that of Lapidoth's expurgated MAC ensemble [6, Thm. 4] after the optimization of the parameters. Furthermore, we show that a refined version of superposition coding (Section V) yields rates at least as high as the standard version for any choice of parameters, with significant improvements possible when the input distribution is fixed. The rates for each ensemble are expressed in both primal and dual forms, and the latter are shown to be valid in the case of general alphabets.

For the mismatched DMC, the results of this paper can be summarized by the following list of random-coding constructions, in decreasing order of rate:

1) Refined superposition coding (Theorems 17 and 19),

2) Standard superposition coding (Theorems 14 and 16),

3) Expurgated parallel coding ([6, Thm. 4] and Theorem 11),

4) Constant-composition or cost-constrained coding with independent codewords (LM Rate [1], [2], [7]),

5) i.i.d. coding with independent codewords (GMI [14]).

Numerical examples are provided in Section V-C.

II. PRELIMINARY RESULTS

In this section, we present a number of general tools and results which will be used throughout the paper.

A. Bounds on the Probability of a Multiply-Indexed Union

In single-user and multiuser channel coding settings, bounds on the random-coding error probability are typically obtained using the union bound, which states that for any set of events $\{A_i\}_{i=1}^M$,

$$\mathbb{P}\Big[\bigcup_i A_i\Big] \le \sum_i \mathbb{P}[A_i]. \tag{10}$$

In many cases the right-hand side of (10) is greater than one, and hence a tighter bound is given by the truncated union bound, namely

$$\mathbb{P}\Big[\bigcup_i A_i\Big] \le \min\Big\{1, \sum_i \mathbb{P}[A_i]\Big\}. \tag{11}$$

In this paper, we will also be interested in lower bounds on the probability of a union, which will be used to prove ensemble tightness results. In particular, we will make use of de Caen's bound [21], which states that

$$\mathbb{P}\Big[\bigcup_i A_i\Big] \ge \sum_i \frac{\mathbb{P}[A_i]^2}{\sum_{i'} \mathbb{P}[A_i \cap A_{i'}]}. \tag{12}$$

In the case that the events are pairwise independent and identically distributed, it is easily verified that (12) proves the tightness of (11) to within a factor of $\frac{1}{2}$ [22]. This tightness result in fact holds even when the events are not identically distributed [23, Lemma A.2].

In this section, we provide a number of upper and lower bounds on the probability of a multiply-indexed union. In several cases of interest, the upper and lower bounds coincide to within a constant factor, and generalize the above-mentioned tightness result of [22] to certain settings where pairwise independence need not hold.

Lemma 1. *Let $\{Z_1(i)\}_{i=1}^{M_1}$ and $\{Z_2(j)\}_{j=1}^{M_2}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. For any set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, we have*

$$\mathbb{P}\Big[\bigcup_{i,j}\big\{(Z_1(i), Z_2(j)) \in \mathcal{A}\big\}\Big] \le \min\bigg\{1, M_1\mathbb{E}\Big[\min\big\{1, M_2\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \,|\, Z_1\big]\big\}\Big],$$
$$M_2\mathbb{E}\Big[\min\big\{1, M_1\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \,|\, Z_2\big]\big\}\Big]\bigg\} \tag{13}$$

where $(Z_1, Z_2) \sim P_{Z_1} \times P_{Z_2}$.

Proof: Applying the union bound to the union over i, we obtain

$$\mathbb{P}\Big[\bigcup_{i,j}\big\{(Z_1(i), Z_2(j)) \in \mathcal{A}\big\}\Big] \le M_1\mathbb{P}\Big[\bigcup_j\big\{(Z_1, Z_2(j)) \in \mathcal{A}\big\}\Big] \tag{14}$$

$$= M_1\mathbb{E}\Big[\mathbb{P}\Big[\bigcup_j\big\{(Z_1, Z_2(j)) \in \mathcal{A}\big\} \,\Big|\, Z_1\Big]\Big]. \tag{15}$$

Applying the truncated union bound to the union over j, we obtain

$$\mathbb{P}\left[\bigcup_{i,j}\left\{(Z_1(i), Z_2(j)) \in \mathcal{A}\right\}\right] \le M_1 \mathbb{E}\left[\min\left\{1, M_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \,|\, Z_1]\right\}\right], \tag{16}$$

thus recovering the second term in the outer minimization in (13). The third term is obtained similarly by applying the union bounds in the opposite order, and the upper bound of 1 is trivial. ∎

Lemma 2. *Let $\{Z_1(i)\}_{i=1}^{M_1}$ and $\{Z_2(j)\}_{j=1}^{M_2}$ be independent sequences of pairwise independent and identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. For any set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, we have*

$$\mathbb{P}\left[\bigcup_{i,j}\left\{(Z_1(i), Z_2(j)) \in \mathcal{A}\right\}\right] \ge \frac{1}{4}\min\left\{1, M_1 \frac{\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]^2}{\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z_2') \in \mathcal{A}\big]},\right.$$
$$\left. M_2 \frac{\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]^2}{\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1', Z_2) \in \mathcal{A}\big]}, M_1 M_2 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]\right\}, \tag{17}$$

where $(Z_1, Z_1', Z_2, Z_2') \sim P_{Z_1}(z_1) P_{Z_1}(z_1') P_{Z_2}(z_2) P_{Z_2}(z_2')$.

Proof: We make use of de Caen's bound in (12). Noting by symmetry that each term in the outer summation is equal, and splitting the inner summation according to which of the (i, j) indices coincide with (i', j'), we obtain

$$\mathbb{P}\left[\bigcup_{i,j}\left\{(Z_1(i), Z_2(j)) \in \mathcal{A}\right\}\right] \ge M_1 M_2 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]^2 \Big((M_1 - 1)(M_2 - 1)\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]$$
$$+ (M_2 - 1)\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z_2') \in \mathcal{A}\big] + (M_1 - 1)\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1', Z_2) \in \mathcal{A}\big] + \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]\Big)^{-1} \tag{18}$$

$$\ge M_1 M_2 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]^2 \Big(4 \max\left\{M_1 M_2 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big], M_2 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z_2') \in \mathcal{A}\big],\right.$$
$$\left. M_1 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1', Z_2) \in \mathcal{A}\big], \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]\right\}\Big)^{-1}, \tag{19}$$

from which the lemma follows. ∎

The following lemma gives conditions under which a weakened version of (13) coincides with (17) to within a factor of four.

Lemma 3. *Let $\{Z_1(i)\}_{i=1}^{M_1}$ and $\{Z_2(j)\}_{j=1}^{M_2}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. Fix a set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, and define*

$$\mathcal{A}_1 \triangleq \left\{z_1 \,:\, (z_1, z_2) \in \mathcal{A} \text{ for some } z_2\right\} \tag{20}$$
$$\mathcal{A}_2 \triangleq \left\{z_2 \,:\, (z_1, z_2) \in \mathcal{A} \text{ for some } z_1\right\}. \tag{21}$$

1) A general upper bound is given by

$$\mathbb{P}\left[\bigcup_{i,j}\left\{(Z_1(i), Z_2(j)) \in \mathcal{A}\right\}\right] \le \min\left\{1, M_1 \mathbb{P}[Z_1 \in \mathcal{A}_1], M_2 \mathbb{P}[Z_2 \in \mathcal{A}_2], M_1 M_2 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]\right\}, \tag{22}$$

where $(Z_1, Z_2) \sim P_{Z_1} \times P_{Z_2}$.

2) If (i) $\{Z_1(i)\}_{i=1}^{M_1}$ are pairwise independent, (ii) $\{Z_2(j)\}_{j=1}^{M_2}$ are pairwise independent, (iii) $\mathbb{P}\big[(z_1, Z_2) \in \mathcal{A}\big]$ is the

same for all $z_1 \in \mathcal{A}_1$, and (iv) $\mathbb{P}\big[(Z_1, z_2) \in \mathcal{A}\big]$ is the same for all $z_2 \in \mathcal{A}_2$, then

$$\mathbb{P}\bigg[\bigcup_{i,j}\Big\{\big(Z_1(i), Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] \geq \frac{1}{4}\min\Big\{1, M_1\mathbb{P}\big[Z_1 \in \mathcal{A}_1\big], M_2\mathbb{P}\big[Z_2 \in \mathcal{A}_2\big], M_1 M_2 \mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]\Big\}. \quad (23)$$

Proof: We obtain (22) by weakening (13) in multiple ways. The second term in (22) follows since the inner probability in the second term of (13) is zero whenever $\mathbb{P}[Z_1 \notin \mathcal{A}]$, and since $\min\{1, \alpha\} \leq 1$. The third term in (22) is obtained similarly, and the fourth term follows from the fact that $\min\{1, \alpha\} \leq \alpha$.

The lower bound in (23) follows from (17), and since the additional assumptions in the second part of the lemma statement imply

$$\frac{\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A}\big]^2}{\mathbb{P}\big[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z_2') \in \mathcal{A}\big]} = \frac{\mathbb{P}\big[Z_1 \in \mathcal{A}_1\big]^2 \mathbb{P}\big[(z_1, Z_2) \in \mathcal{A}\big]^2}{\mathbb{P}\big[Z_1 \in \mathcal{A}_1\big] \mathbb{P}\big[(z_1, Z_2) \in \mathcal{A}\big]^2}, \quad (24)$$

$$= \mathbb{P}\big[Z_1 \in \mathcal{A}_1\big] \quad (25)$$

where z_1 is an arbitrary element of $\mathcal{A}_1$. The third term in the minimization in (17) can be handled similarly. ∎

While we will primarily be interested in doubly-indexed unions in this paper, it will prove useful to state the following generalization of Lemma 3 to the probability of a union indexed by K values, which is stated without proof. For a given subset $\mathcal{K} = \{k_1, \cdots, k_{|\mathcal{K}|}\}$ of $\{1, \cdots, K\}$, we write $\boldsymbol{Z}_{\mathcal{K}}$ as a shorthand for $(Z_{k_1}, \cdots, Z_{k_{|\mathcal{K}|}})$, and similarly for $\boldsymbol{Z}_{\mathcal{K}^c}$. The corresponding realizations are written by $\boldsymbol{z}_{\mathcal{K}} = (z_{k_1}, \cdots, z_{k_{|\mathcal{K}|}})$, and similarly for $\boldsymbol{z}_{\mathcal{K}^c}$.

Lemma 4. *Let $\{Z_1(i_1)\}_{i_1=1}^{M_1}, \cdots, \{Z_K(i_K)\}_{i_K=1}^{M_K}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1, \cdots, \mathcal{Z}_K$ respectively, with $Z_k(i) \sim P_{Z_k}$ for $k = 1, \cdots, K$. Fix a set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_K$, and for each $\mathcal{K} = \{k_1, \cdots, k_{|\mathcal{K}|}\} \subseteq \{1, \cdots, K\}$ define*

$$\mathcal{A}_{\mathcal{K}} \triangleq \Big\{\boldsymbol{z}_{\mathcal{K}} : (z_1, \cdots, z_K) \in \mathcal{A} \text{ for some } \boldsymbol{z}_{\mathcal{K}^c}\Big\} \quad (26)$$

1) A general upper bound is given by

$$\mathbb{P}\bigg[\bigcup_{i_1, \cdots, i_K}\Big\{\big(Z_1(i_1), \cdots, Z_K(i_K)\big) \in \mathcal{A}\Big\}\bigg] \leq \min\bigg\{1, \min_{\mathcal{K} \subseteq \{1, \cdots, K\}, \mathcal{K} \neq \emptyset}\Big(\prod_{k \in \mathcal{K}} M_k\Big)\mathbb{P}\big[\boldsymbol{Z}_{\mathcal{K}} \in \mathcal{A}_{\mathcal{K}}\big]\bigg\}, \quad (27)$$

where $(Z_1, \cdots, Z_K) \sim \prod_{k=1}^{K} P_{Z_k}$.

2) If $\{Z_k(i)\}_{i=1}^{M_k}$ are pairwise independent for any given k, and $\mathbb{P}\big[(Z_1, \cdots, Z_K) \in \mathcal{A} | \boldsymbol{Z}_{\mathcal{K}} = \boldsymbol{z}_{\mathcal{K}}\big]$ is the same for all $\boldsymbol{z}_{\mathcal{K}} \in \mathcal{A}_{\mathcal{K}}$ for any given $\mathcal{K} \subseteq \{1, \cdots, K\}$, then

$$\mathbb{P}\bigg[\bigcup_{i_1, \cdots, i_K}\Big\{\big(Z_1(i_1), \cdots, Z_K(i_K)\big) \in \mathcal{A}\Big\}\bigg] \geq 2^{-K}\min\bigg\{1, \min_{\mathcal{K} \subseteq \{1, \cdots, K\}, \mathcal{K} \neq \emptyset}\Big(\prod_{k \in \mathcal{K}} M_k\Big)\mathbb{P}\big[\boldsymbol{Z}_{\mathcal{K}} \in \mathcal{A}_{\mathcal{K}}\big]\bigg\}, \quad (28)$$

where $(Z_1, \cdots, Z_K) \sim \prod_{k=1}^{K} P_{Z_k}$.

B. Equivalent Forms of Convex Optimization Problems

The exponents and rates derived in this paper will be presented in both primal and dual forms, analogously to the LM rate in (6)–(7). The corresponding proofs of equivalence are more involved than that of the LM rate (see [4]). Here we provide two lemmas which will be useful in proving the equivalences. The following lemma is a generalization of the fact that (6) and (7) are equivalent, which is proved using Lagrange duality [12].

Lemma 5. *Fix the finite alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$, the non-negative functions $f(z_1, z_2)$ and $g(z_1, z_2)$, the distributions*

P_{Z_1} and P_{Z_2}, and a constant β. The optimization

$$\min_{\substack{\widetilde{P}_{Z_1 Z_2} : \widetilde{P}_{Z_1} = P_{Z_1}, \widetilde{P}_{Z_2} = P_{Z_2}, \\ \mathbb{E}_{\widetilde{P}}[\log f(Z_1, Z_2)] \geq \beta}} I_{\widetilde{P}}(Z_1; Z_2) - \mathbb{E}_{\widetilde{P}}[\log g(Z_1, Z_2)] \tag{29}$$

has the same value as

$$\sup_{\lambda \geq 0, \mu_1(\cdot)} - \sum_{z_2} P_{Z_2}(z_2) \log \sum_{\overline{z}_1} P_{Z_1}(\overline{z}_1) f(\overline{z}_1, z_2)^\lambda g(\overline{z}_1, z_2) e^{\mu_1(\overline{z}_1)} + \sum_{z_1} P_{Z_1}(z_1) \mu_1(z_1) + \lambda\beta, \tag{30}$$

where the supremum over $\mu_1(\cdot)$ is taken over all real-valued functions on $\mathcal{Z}_1$.

Proof: See Appendix B. ∎

When using Lemma 5, we will typically be interested the case that either $g(\cdot, \cdot)$ is absent from (29), or the constraint $\mathbb{E}_{\widetilde{P}}[\log f(Z_1, Z_2)] \geq \beta$ is absent. In the former case, the dual expression is given by (30) with $g(\cdot, \cdot) = 1$, and in the latter case, the dual expression is given by (30) with $\lambda = 0$. Observing the symmetry of (29) in Z_1 and Z_2, it can be seen that the roles of the two can also be swapped in (30).

The following lemma will allow certain convex optimization problems to be expressed in a form where, after some simple manipulations, Lemma 5 can be applied.

Lemma 6. *Fix a positive integer d and let $\mathcal{D}$ be a convex subset of $\mathbb{R}^d$. Let $f(\boldsymbol{z})$, $g(\boldsymbol{z})$, $g_1(\boldsymbol{z})$ and $g_2(\boldsymbol{z})$ be convex functions mapping $\mathbb{R}^d$ to $\mathbb{R}$ such that*

$$g_1(\boldsymbol{z}) + g_2(\boldsymbol{z}) \leq g(\boldsymbol{z}) \tag{31}$$

for all $\boldsymbol{z} \in \mathcal{D}$. Then the optimization

$$\min_{\boldsymbol{z} \in \mathcal{D}} f(\boldsymbol{z}) + \left[\max\left\{ g_1(\boldsymbol{z}), g_2(\boldsymbol{z}), g(\boldsymbol{z}) \right\} \right]^+ \tag{32}$$

has the same value as

$$\max\left\{ \min_{\boldsymbol{z} \in \mathcal{D}} f(\boldsymbol{z}) + \left[\max\left\{ g_1(\boldsymbol{z}), g(\boldsymbol{z}) \right\} \right]^+, \min_{\boldsymbol{z} \in \mathcal{D}} f(\boldsymbol{z}) + \left[\max\left\{ g_2(\boldsymbol{z}), g(\boldsymbol{z}) \right\} \right]^+ \right\}. \tag{33}$$

Proof: See Appendix B. ∎

III. MULTIPLE-ACCESS CHANNEL

In this section, we study the mismatched multiple-access channel introduced in Section I-A2. We consider random coding, in which each codeword of user $\nu = 1, 2$ is generated independently according to some distribution $P_{\boldsymbol{X}_\nu}$. We let $\boldsymbol{X}_\nu^{(i)}$ be the random variable corresponding to the i-th codeword of user ν, yielding

$$\left(\{\boldsymbol{X}_1^{(i)}\}_{i=1}^{M_1}, \{\boldsymbol{X}_2^{(j)}\}_{i=1}^{M_2} \right) \sim \prod_{i=1}^{M_1} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1^{(i)}) \prod_{j=1}^{M_2} P_{\boldsymbol{X}_2}(\boldsymbol{x}_2^{(j)}). \tag{34}$$

We assume without loss of generality that message $(1, 1)$ is transmitted, and write $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in place of $\boldsymbol{X}_1^{(1)}$ and $\boldsymbol{X}_2^{(1)}$. We write $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$ to denote arbitrary codewords which are generated independently of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The random sequence at the output of the channel is denoted by $\boldsymbol{Y}$. It follows that

$$(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}, \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2) \sim P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2). \tag{35}$$

All of our results can be extended to the setting in which the codewords are generated conditionally on a time-sharing sequence $\boldsymbol{u}$ (e.g. see [20]). However, this leads to more complex notation, and also introduces a number of subtle issues in the mismatched setting which are not present in the matched setting. Thus, we focus primarily on the case

that there is no time-sharing. In Section III-E, we state the corresponding results with time-sharing, and discuss the differences between the matched and mismatched settings.

We distinguish between the following three types of error:

(Type 1) $\hat{m}_1 \neq m_1$ and $\hat{m}_2 = m_2$

(Type 2) $\hat{m}_1 = m_1$ and $\hat{m}_2 \neq m_2$

(Type 12) $\hat{m}_1 \neq m_1$ and $\hat{m}_2 \neq m_2$

The corresponding random-coding error probabilities for these events are denoted by $\overline{p}_{e,1}(n, M_1)$, $\overline{p}_{e,2}(n, M_2)$ and $\overline{p}_{e,12}(n, M_1, M_2)$ respectively, and the overall random-coding error probability is denoted by $\overline{p}_e(n, M_1, M_2)$. Clearly we have

$$\max\{\overline{p}_{e,1}, \overline{p}_{e,2}, \overline{p}_{e,12}\} \leq \overline{p}_e \leq \overline{p}_{e,1} + \overline{p}_{e,2} + \overline{p}_{e,12}. \tag{36}$$

A. Random-Coding Bounds

The following theorem extends the random-coding union (RCU) bound for mismatched decoders [8], [24] to the MAC.

Theorem 1. *Under the random-coding distributions $P_{\boldsymbol{X}_1}$ and $P_{\boldsymbol{X}_2}$, the ensemble-average error probabilities for the maximum-metric decoder satisfy*

$$\overline{p}_{e,1}(n, M_1) \leq \mathrm{rcu}_1(n, M_1) \tag{37}$$

$$\overline{p}_{e,2}(n, M_2) \leq \mathrm{rcu}_2(n, M_2) \tag{38}$$

$$\overline{p}_{e,12}(n, M_1, M_2) \leq \min\left\{\mathrm{rcu}_{12,1}(n, M_1, M_2), \mathrm{rcu}_{12,2}(n, M_1, M_2)\right\} \tag{39}$$

for errors of type 1, 2 and 12 respectively, where

$$\mathrm{rcu}_1(n, M_1) \triangleq \mathbb{E}\left[\min\left\{1, (M_1 - 1)\mathbb{P}\left[\frac{q^n(\overline{\boldsymbol{X}}_1, \boldsymbol{X}_2, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1 \,\middle|\, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right\}\right] \tag{40}$$

$$\mathrm{rcu}_2(n, M_2) \triangleq \mathbb{E}\left[\min\left\{1, (M_2 - 1)\mathbb{P}\left[\frac{q^n(\boldsymbol{X}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1 \,\middle|\, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right\}\right] \tag{41}$$

$$\mathrm{rcu}_{12,1}(n, M_1, M_2) \triangleq$$
$$\mathbb{E}\left[\min\left\{1, (M_1 - 1)\mathbb{E}\left[\min\left\{1, (M_2 - 1)\mathbb{P}\left[\frac{q^n(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1 \,\middle|\, \overline{\boldsymbol{X}}_1\right]\right\} \,\middle|\, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right\}\right] \tag{42}$$

$$\mathrm{rcu}_{12,2}(n, M_1, M_2) \triangleq$$
$$\mathbb{E}\left[\min\left\{1, (M_2 - 1)\mathbb{E}\left[\min\left\{1, (M_1 - 1)\mathbb{P}\left[\frac{q^n(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1 \,\middle|\, \overline{\boldsymbol{X}}_2\right]\right\} \,\middle|\, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right\}\right]. \tag{43}$$

Proof: For the type-1 and type-2 error probabilities, the RCU bounds follow using the same steps as the single-user setting [24]. For the type-12 error, we have

$$\overline{p}_{e,12} \leq \mathbb{P}\left[\bigcup_{i \neq 1, j \neq 1}\left\{\frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1\right\}\right] \tag{44}$$

$$= \mathbb{E}\left[\mathbb{P}\left[\bigcup_{i \neq 1, j \neq 1}\left\{\frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1\right\} \,\middle|\, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right]. \tag{45}$$

We obtain (42)–(43) by applying Lemma 1 to the union in (45), with $Z_1(i) = \boldsymbol{X}_1^{(i)}$ and $Z_2(j) = \boldsymbol{X}_2^{(j)}$. In each case we upper bound the minimum of three values in (13) by a minimum of two values. ∎

The type-1 and type-2 RCU bounds in (40)–(41) have appeared previously [25], along with the type-12 RCU bound

$$\mathrm{rcu}'_{12}(n, M_1, M_2) \triangleq \mathbb{E}\left[\min\left\{1, (M_1 - 1)(M_2 - 1)\mathbb{P}\left[\frac{q^n(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1 \;\middle|\; \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right\}\right]. \tag{46}$$

Both $\mathrm{rcu}_{12,1}$ and $\mathrm{rcu}_{12,2}$ are always less than or equal to rcu'_{12}, and hence give better bounds on the error probability. We will see that in the mismatched setting, rcu'_{12} yields not only a worse non-asymptotic bound, but also a worse achievable rate region and error exponent. While no such losses in the rates and exponents are observed under ML decoding (see Section III-D), the refined RCU bounds $\mathrm{rcu}_{12,1}$ and $\mathrm{rcu}_{12,2}$ are still of independent interest for characterizing the non-asymptotic performance.

B. Exponents and Rates for the DM-MAC

In this section, we study the DM-MAC using the constant-composition ensemble, previously studied by Liu and Hughes [19], among others. For $\nu = 1, 2$, we fix $Q_\nu \in \mathcal{P}(\mathcal{X}_\nu)$ and take $P_{\boldsymbol{X}_\nu}$ to be the uniform distribution on the type class $T^n(Q_{\nu,n})$, where $Q_{\nu,n} \in \mathcal{P}_n(\mathcal{X}_\nu)$ is the most probable type under Q_ν, i.e.

$$P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu) = \frac{1}{|T^n(Q_{\nu,n})|}\mathbb{1}\{\boldsymbol{x}_\nu \in T^n(Q_{\nu,n})\}. \tag{47}$$

Our analysis is based on the method of types; the main properties used are given in Appendix A. Throughout the section, we write $f(\boldsymbol{Q})$ to denote a function f that depends on Q_1 and Q_2. Similarly, we write $f(\boldsymbol{Q}_n)$ to denote a function that depends on $Q_{1,n}$ and $Q_{2,n}$.

1) Error Exponent: The error exponents and achievable rates will be expressed in terms of the sets

$$\mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q}) \triangleq \left\{P_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : P_{X_1} = Q_1, P_{X_2} = Q_2\right\} \tag{48}$$

$$\mathcal{T}_1^{\mathrm{cc}}(P_{X_1 X_2 Y}) \triangleq \left\{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \right.$$
$$\left. \widetilde{P}_{X_1} = P_{X_1}, \widetilde{P}_{X_2 Y} = P_{X_2 Y}, \mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]\right\} \tag{49}$$

$$\mathcal{T}_2^{\mathrm{cc}}(P_{X_1 X_2 Y}) \triangleq \left\{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \right.$$
$$\left. \widetilde{P}_{X_2} = P_{X_2}, \widetilde{P}_{X_1 Y} = P_{X_1 Y}, \mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]\right\} \tag{50}$$

$$\mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1 X_2 Y}) \triangleq \left\{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \right.$$
$$\left. \widetilde{P}_{X_1} = P_{X_1}, \widetilde{P}_{X_2} = P_{X_2}, \widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]\right\}. \tag{51}$$

The following theorem gives the random-coding error exponent for each error type.

Theorem 2. *The random-coding error probabilities for the constant-composition ensemble in* (47) *satisfy*

$$\overline{p}_{e,1}(n, e^{nR_1}) \doteq \exp\left(-nE_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1)\right) \tag{52}$$

$$\overline{p}_{e,2}(n, e^{nR_2}) \doteq \exp\left(-nE_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2)\right) \tag{53}$$

$$\overline{p}_{e,12}(n, e^{nR_1}, e^{nR_2}) \doteq \exp\left(-nE_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2)\right), \tag{54}$$

where

$$E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1) \triangleq \min_{P_{X_1X_2Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_1^{\mathrm{cc}}(P_{X_1X_2Y})} D(P_{X_1X_2Y}\|Q_1 \times Q_2 \times W) + \left[I_{\widetilde{P}}(X_1; X_2, Y) - R_1\right]^+$$

(55)

$$E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2) \triangleq \min_{P_{X_1X_2Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_2^{\mathrm{cc}}(P_{X_1X_2Y})} D(P_{X_1X_2Y}\|Q_1 \times Q_2 \times W) + \left[I_{\widetilde{P}}(X_2; X_1, Y) - R_2\right]^+$$

(56)

$$E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \triangleq \min_{P_{X_1X_2Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1X_2Y})} D(P_{X_1X_2Y}\|Q_1 \times Q_2 \times W) \qquad (57)$$

$$+ \left[\max\left\{I_{\widetilde{P}}(X_1; Y) - R_1, I_{\widetilde{P}}(X_2; Y) - R_2, D(\widetilde{P}_{X_1X_2Y}\|Q_1 \times Q_2 \times \widetilde{P}_Y) - R_1 - R_2\right\}\right]^+.$$

(58)

Proof: The random-coding error probabilities $\overline{p}_{e,1}$ and $\overline{p}_{e,2}$ can be handled similarly to the single-user setting [8]. Furthermore, equivalent error exponents to (55)–(56) are given in [20]. We therefore focus on $\overline{p}_{e,12}$, which requires a more refined analysis. We first write

$$\overline{p}_{e,12} = c_{12}\mathbb{E}\left[\mathbb{P}\left[\bigcup_{i\neq 1, j\neq 1}\left\{\frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1\right\}\Bigg| \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right]$$

(59)

for some $c_{12} \in [\frac{1}{2}, 1]$. Setting $c_{12} = 1$ yields the average probability of error when ties are decoded as errors, and we have $c_{12} \geq \frac{1}{2}$ since decoding ties at random reduces the error probability by at most a factor of two [26].

We will rewrite (59) in terms of the possible joint types of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$ and $(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})$. To this end, we define

$$\mathcal{S}_n^{\mathrm{cc}}(\boldsymbol{Q}_n) \triangleq \left\{P_{X_1X_2Y} \in \mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : P_{X_1} = Q_{1,n}, P_{X_2} = Q_{2,n}\right\}$$

(60)

$$\mathcal{T}_{12,n}^{\mathrm{cc}}(P_{X_1X_2Y}) \triangleq \mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1X_2Y}) \cap \mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}).$$

(61)

Roughly speaking, $\mathcal{S}_n^{\mathrm{cc}}$ is the set of possible joint types of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$, and $\mathcal{T}_{12,n}^{\mathrm{cc}}(P_{X_1X_2Y})$ is the set of types of $(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})$ which lead to decoding errors when $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \in T^n(P_{X_1X_2Y})$. The constraints on P_{X_ν} and $\widetilde{P}_{X_\nu}$ arise from the fact that we are using constant-composition random coding, and the constraint $\mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]$ holds if and only if $q^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ for $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in T^n(P_{X_1X_2Y})$ and $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1X_2Y})$.

Fixing $P_{X_1X_2Y} \in \mathcal{S}_n^{\mathrm{cc}}(\boldsymbol{Q}_n)$ and letting $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ be an arbitrary triplet of sequences such that $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in T^n(P_{X_1X_2Y})$, it follows that the event

$$\bigcup_{i\neq 1, j\neq 1}\left\{\frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{y})}{q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})} \geq 1\right\}$$

(62)

can be written as

$$\bigcup_{i\neq 1, j\neq 1} \bigcup_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12,n}^{\mathrm{cc}}(P_{X_1X_2Y})}\left\{(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1X_2Y})\right\}.$$

(63)

Expanding the probability and expectation in (59) in terms of types, substituting (63), and interchanging the order of

the unions, we obtain

$$\bar{p}_{e,12} = c_{12} \times \sum_{P_{X_1 X_2 Y} \in \mathcal{S}_n^{\mathrm{cc}}(\boldsymbol{Q}_n)} \mathbb{P}\big[(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \in T^n(P_{X_1 X_2 Y})\big]$$

$$\times \mathbb{P}\left[\bigcup_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12,n}^{\mathrm{cc}}(P_{X_1 X_2 Y})} \bigcup_{i \neq 1, j \neq 1} \left\{ (\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1 X_2 Y}) \right\} \right] \quad (64)$$

$$\doteq \max_{P_{X_1 X_2 Y} \in \mathcal{S}_n^{\mathrm{cc}}(\boldsymbol{Q}_n)} \mathbb{P}\big[(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \in T^n(P_{X_1 X_2 Y})\big]$$

$$\times \max_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12,n}^{\mathrm{cc}}(P_{X_1 X_2 Y})} \mathbb{P}\left[\bigcup_{i \neq 1, j \neq 1} \left\{ (\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1 X_2 Y}) \right\} \right], \quad (65)$$

where $\boldsymbol{y}$ is an arbitrary element of $T^n(P_Y)$, and (65) follows from the union bound and since the number of joint types is polynomial in n.

The exponential behavior of the first probability in (65) is given by the property of types in (A.9), so it only remains to determine the exponential behavior of the second probability. To this end, we make use of Lemma 3 with $Z_1(i) = \boldsymbol{X}_1^{(i)}$, $Z_2(j) = \boldsymbol{X}_2^{(j)}$, $\mathcal{A} = T_{\boldsymbol{y}}^n(\widetilde{P}_{X_1 X_2 Y})$, $\mathcal{A}_1 = T_{\boldsymbol{y}}^n(\widetilde{P}_{X_1 Y})$ and $\mathcal{A}_2 = T_{\boldsymbol{y}}^n(\widetilde{P}_{X_2 Y})$. Using (22)–(23) and the properties of types in (A.10) and (A.11), it follows that the second probability in (65) has an exponent of

$$\left[\max \left\{ I_{\widetilde{P}}(X_1; Y) - R_1, I_{\widetilde{P}}(X_2; Y) - R_2, D\big(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y\big) - R_1 - R_2 \right\} \right]^+. \quad (66)$$

Substituting (66) and the property of types in (A.9) into (65), and using a continuity argument (e.g. see [27, Eq. (30)]) to replace the sets $\mathcal{S}_n^{\mathrm{cc}}$ and $\mathcal{T}_{12,n}^{\mathrm{cc}}$ by $\mathcal{S}^{\mathrm{cc}}$ and $\mathcal{T}_{12}^{\mathrm{cc}}$ respectively, we recover the exponent in (58). ∎

Due to the lack of converse results in mismatched decoding, it is important to determine whether the weakness in the achievability results is due to the ensemble itself, or the bounding techniques used in the analysis. Theorem 2 shows that the overall error exponent

$$E_r^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \triangleq \min \left\{ E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \right\} \quad (67)$$

is not only achievable, but it is also tight with respect to the ensemble average.

The proof of Theorem 2 made use of the refined union bound given in Lemma 3. If we had instead used the standard truncated union bound in (11), we would have obtained exponent

$$E_{r,12}^{\mathrm{cc}'}(\boldsymbol{Q}, R_1, R_2) \triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1 X_2 Y})}$$
$$D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \big[D(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y) - (R_1 + R_2) \big]^+, \quad (68)$$

which coincides with an achievable exponent given in [20]. Following the analysis of [8], one can easily show that this is the exponent of the quantity rcu_{12}' defined in (46).

In the case that there are $K > 2$ users, one can use Lemma 4 in place of Lemma 3. For example, for the joint error event between all K users, the minimum of four terms in (58) is replaced by a minimum of 2^K terms, corresponding to the subsets of $\{1, \cdots, K\}$

2) Achievable Rate Region: The following achievable rate region follows from Theorem 2 in a straightforward fashion, and coincides with the ensemble-tight achievable rate region of Lapidoth [6].

Theorem 3. *The overall error exponent $E_r^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2)$ is positive for all rate pairs (R_1, R_2) in the interior of*

$\mathcal{R}^{\mathrm{LM}}(\boldsymbol{Q})$, where $\mathcal{R}^{\mathrm{LM}}(\boldsymbol{Q})$ is the set of all rate pairs (R_1, R_2) satisfying

$$R_1 \le \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_1^{\mathrm{cc}}(Q_1 \times Q_2 \times W)} I_{\widetilde{P}}(X_1; X_2, Y) \tag{69}$$

$$R_2 \le \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_2^{\mathrm{cc}}(Q_1 \times Q_2 \times W)} I_{\widetilde{P}}(X_2; X_1, Y) \tag{70}$$

$$R_1 + R_2 \le \min_{\substack{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\mathrm{cc}}(Q_1 \times Q_2 \times W) \\ I_{\widetilde{P}}(X_1; Y) \le R_1, \, I_{\widetilde{P}}(X_2; Y) \le R_2}} D(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y). \tag{71}$$

Proof: The conditions in (69)–(71) are obtained from the error exponents in (55)–(58) respectively. Focusing on (71), we see that the objective in (58) is always positive when $D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) > 0$, $I_{\widetilde{P}}(X_1; Y) > R_1$ or $I_{\widetilde{P}}(X_2; Y) > R_2$. Thus, by taking $D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \to 0$ and introducing the constraints $I_{\widetilde{P}}(X_1; Y) \le R_1$ and $I_{\widetilde{P}}(X_2; Y) \le R_2$, we obtain the condition in (71). ∎

Using the usual time-sharing argument [6], [28] (see also Section III-E), it follows from Theorem 3 that we can achieve any rate pair in the convex hull of $\bigcup_{\boldsymbol{Q}} \mathcal{R}^{\mathrm{LM}}(\boldsymbol{Q})$, where the union is over all distributions Q_1 and Q_2 on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively.

The ensemble tightness of the rate region given in Theorem 3 does not follow directly from the fact that the exponents in Theorem 2 are ensemble tight, since in principle the random-coding error probability could vanish subexponentially at higher rates. However, the lower bounding techniques used in Theorem 2 can be adapted to prove the tightness of the rate region (e.g. see Section V-A). We omit the analysis here, since the ensemble tightness of the rate region is well-known [6].

Using a similar argument to the proof of Theorem 3, we see that (68) yields the rate condition

$$R_1 + R_2 \le \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\mathrm{cc}}(Q_1 \times Q_2 \times W)} D(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y). \tag{72}$$

In the following subsection, we compare the ensemble-tight expressions in (58) and (71) with the weaker expressions in (68) and (72).

3) Numerical Examples: We now return to the parallel BSC example given in Figure 1, where the output is given by $y = (y_1, y_2)$. As mentioned in Section I-C, the decoder assumes that both crossover probabilities are equal. It is straightforward to show that the corresponding decoding rule is equivalent to minimizing sum of t_1 and t_2, where t_ν is the number of bit flips from the input sequence $\boldsymbol{x}_\nu$ to the output sequence $\boldsymbol{y}_\nu$. As noted in [6], this decision rule is in fact equivalent to ML. This channel could easily be analyzed by treating the two subchannels separately, but we treat it as a MAC because it serves as a good example for comparing the ensemble-tight achievability results with (68) and (72).

We let both Q_1 and Q_2 be the equiprobable distribution on $\{0, 1\}$. With this choice, it was shown in [6] that the right-hand side of (72) is no greater than

$$2 \left(1 - H_2 \left(\frac{\delta_1 + \delta_2}{2} \right) \right) \text{ bits/use.} \tag{73}$$

On the other hand, the refined condition in (71) can be used to prove the achievability of *any* (R_1, R_2) within the rectangle with corners $(0, 0)$ and (C_1, C_2), where $C_\nu \triangleq 1 - H_2(\delta_\nu)$ [6]. This observation is analogous to the comparison between (8) and (9) in the introduction. The main difference is that the weakness in (8) is in the random-coding ensemble itself, whereas the weakness in (73) is in the bounding techniques used in the analysis.

We evaluate the error exponents using the optimization software YALMIP [29]. Figure 2 plots each of the exponents as a function of α, where the rate pairs are given by $(R_1, R_2) = (\alpha C_1, \alpha C_2)$. While the overall error exponent $E_r(\boldsymbol{Q}, R_1, R_2)$ in (67) is unchanged at low to moderate values of α when $E_{r,12}^{\mathrm{cc}'}$ is used in place of $E_{r,12}$, this is not true for high values of α. Furthermore, consistent with the preceding discussion, $E_{r,12}^{\mathrm{cc}'}$ is non-zero only for $\alpha < 0.865$,
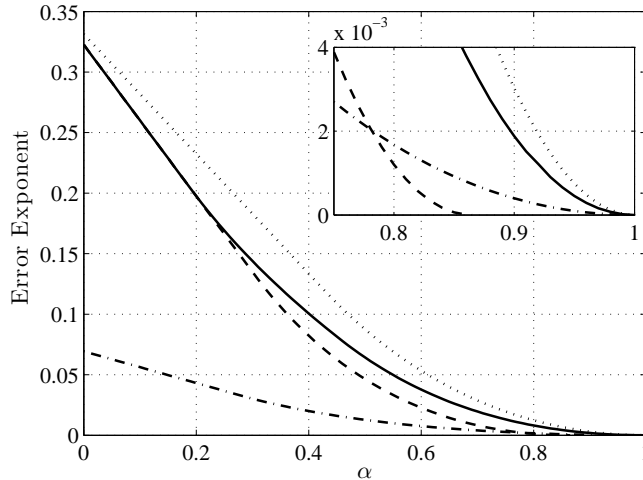
Figure 2. Error exponents $E_{r,1}^{\mathrm{cc}}$ (dotted), $E_{r,2}^{\mathrm{cc}}$ (dash-dot), $E_{r,12}^{\mathrm{cc}}$ (solid) and $E_{r,12}^{\mathrm{cc}'}$ (dashed) for the parallel channel shown in Figure 1 using $\delta_1 = 0.05$, $\delta_2 = 0.25$ and equiprobable input distributions on $\{0, 1\}$.
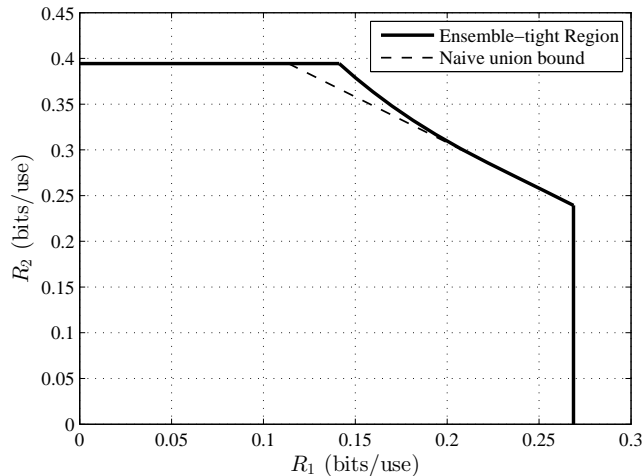

Figure 3. Achievable rate region for the MAC associated with (74) using the corresponding decoding metric in which all values of δ are equal. The crossover probabilities are $\delta_0 = 0.02$, $\delta_1 = 0.2$, $\delta_2 = 0.025$ and $\delta_3 = 0.25$, and both input distributions are equiprobable on $\{0, 1\}$. The ensemble-tight region is given in Theorem 3, whereas the naive union bound replaces (71) with (72).

whereas $E_{r,12}^{\mathrm{cc}}$ is positive for all $\alpha < 1$. It is interesting to note that the curves $E_{r,12}^{\mathrm{cc}}$ and $E_{r,12}^{\mathrm{cc}'}$ coincide at low values of α. This is consistent with [20, Cor. 5], which states that $E_{r,12}^{\mathrm{cc}'}$ is ensemble-tight at low rates.

Next, we show that the achievable rate region $\mathcal{R}^{\mathrm{LM}}(\boldsymbol{Q})$ can be non-convex even for a given pair (Q_1, Q_2). We consider the following example, with $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, 2, 3\}$. The channel is described by $W(y|x_1, x_2) = W'(y|2x_1 + x_2)$, where $W'(y|x)$ is defined by the entries of the matrix

$$\boldsymbol{W}' = \begin{bmatrix} 1 - 3\delta_0 & \delta_0 & \delta_0 & \delta_0 \\ \delta_1 & 1 - 3\delta_1 & \delta_1 & \delta_1 \\ \delta_2 & \delta_2 & 1 - 3\delta_2 & \delta_2 \\ \delta_3 & \delta_3 & \delta_3 & 1 - 3\delta_3 \end{bmatrix}. \tag{74}$$

The decoding metric is defined similarly, but with each δ_i replaced by a single value $\delta \in (0, 0.25)$ (all such choices of δ are equivalent).

Figure 3 plots the achievable rate region $\mathcal{R}^{\mathrm{LM}}(\boldsymbol{Q})$ with $\delta_0 = 0.02$, $\delta_1 = 0.2$, $\delta_2 = 0.025$, $\delta_3 = 0.25$, and

$Q_1 = Q_2 = (0.5, 0.5)$. The boundary between the horizontal and vertical limits is not a straight line, but instead a convex curve. To gain some insight as to how this behavior arises, we discuss the condition (71). In this example, the constraint $I_{\widetilde{P}}(X_2; Y) \leq R_2$ is inactive, whereas the constraint $I_{\widetilde{P}}(X_1; Y) \leq R_1$ affects the region. As R_1 increases, the constraint is satisfied by a wider range of distributions $\widetilde{P}_{X_1 X_2 Y}$, and the sum rate $R_1 + R_2$ decreases. The corner point (0.2687,0.2392) is the same whether the ensemble-tight bound or the weakened bound is used, indicating that the constraint $I_{\widetilde{P}}(X_1; Y) \leq R_1$ is also inactive at this point. However, at lower values of R_1, the constraint is active, and the ensemble-tight bound is better. In particular, the upper corner point (0.1412,0.3944) improves on that of (0.1136,0.3944), which corresponds to the weaker rate condition in (72).

4) Alternative Expressions for the Error Exponents and Rates: In this subsection, we derive equivalent dual expressions for the exponents and rates given in Theorems 2 and 3. The proofs are a non-trivial extension of the single-user setting [4], and are made possible by Lemmas 5 and 6. In Section III-C, we will show that these dual expressions generalize to continuous alphabets, analogously to the single-user setting [7], [8].

Lemma 7. *The exponent $E_{r,12}^{\mathrm{cc}}$ in (58) can be expressed as*

$$E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) = \max\left\{E_{r,12,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2), E_{r,12,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2)\right\}, \tag{75}$$

where

$$E_{r,12,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1 X_2 Y})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W)$$
$$+ \left[\max\left\{I_{\widetilde{P}}(X_1; Y) - R_1, D\big(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y\big) - R_1 - R_2\right\}\right]^+ \tag{76}$$

$$E_{r,12,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1 X_2 Y})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W)$$
$$+ \left[\max\left\{I_{\widetilde{P}}(X_2; Y) - R_2, D\big(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y\big) - R_1 - R_2\right\}\right]^+. \tag{77}$$

Proof: This is a special case of Lemma 6 with the following identifications:

$$f(\boldsymbol{z}) = D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \tag{78}$$
$$g(\boldsymbol{z}) = D\big(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y\big) - R_1 - R_2 \tag{79}$$
$$g_1(\boldsymbol{z}) = I_{\widetilde{P}}(X_1; Y) - R_1 \tag{80}$$
$$g_2(\boldsymbol{z}) = I_{\widetilde{P}}(X_2; Y) - R_2. \tag{81}$$

From the latter three choices and the identity

$$D\big(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y\big) = I_{\widetilde{P}}(X_1; Y) + I_{\widetilde{P}}(X_2; Y) + I_{\widetilde{P}}(X_1; X_2 | Y), \tag{82}$$

we see that the condition in (31) is satisfied. ∎

The exponents in (76)–(77) are precisely those that would have been obtained by applying properties of types directly to the RCU bounds in (42)–(43), rather than directly analyzing the random-coding error probability as in the proof of Theorem 2. Thus, our RCU bounds yield the true error exponent for the constant-composition ensemble.

Lemma 8. *The achievable rate region in Theorem 3 is unchanged if the condition that (71) holds is replaced by the*

condition that at least one of the following hold:

$$R_1 + R_2 \leq \min_{\substack{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{cc}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_1;Y) \leq R_1}} D(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y) \tag{83}$$

$$R_1 + R_2 \leq \min_{\substack{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{cc}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_2;Y) \leq R_2}} D(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y). \tag{84}$$

Proof: These conditions can be derived from (76)–(77) in an identical fashion to (71). Specifically, $E_{r,12}^{cc}$ is positive whenever one of (76) and (77) hold, and zero otherwise. A similar statement holds for (71), and thus the two formulations must be equivalent. ∎

The conditions in (83)–(84) yield a notable advantage over the equivalent condition in (71) in terms of computation. Based on (71), one may expect that the achievable rate region must be generated by evaluating the minimization over a grid of (R_1, R_2) values. However, from (83)–(84), we can avoid such a computationally complex search by instead evaluating the minimization in (83) over a range of R_1 values, and the minimization in (84) over a range of R_2 values. Thus, the number of minimizations performed is reduced from $O(N_1 N_2)$ to $O(N_1 + N_2)$, where for $\nu = 1, 2$, N_ν is the number of points used in the quantization of the R_ν axis.

The dual expressions for the error exponents in (55)–(56) and (76)–(77) are given in the following theorem.

Theorem 4. *The exponents $E_{r,1}^{cc}$, $E_{r,2}^{cc}$, $E_{r,12,1}^{cc}$ and $E_{r,12,2}^{cc}$ can be expressed as*

$$E_{r,1}^{cc}(\boldsymbol{Q}, R_1) = \sup_{\rho_1 \in [0,1]} E_{0,1}^{cc}(\boldsymbol{Q}, \rho_1) - \rho_1 R_1 \tag{85}$$

$$E_{r,2}^{cc}(\boldsymbol{Q}, R_2) = \sup_{\rho_2 \in [0,1]} E_{0,2}^{cc}(\boldsymbol{Q}, \rho_1) - \rho_2 R_2 \tag{86}$$

$$E_{r,12,1}^{cc}(\boldsymbol{Q}, R_1, R_2) = \sup_{\rho_1 \in [0,1], \rho_2 \in [0,1]} E_{0,12,1}^{cc}(\boldsymbol{Q}, \rho_1, \rho_2) - \rho_1(R_1 + \rho_2 R_2) \tag{87}$$

$$E_{r,12,2}^{cc}(\boldsymbol{Q}, R_1, R_2) = \sup_{\rho_1 \in [0,1], \rho_2 \in [0,1]} E_{0,12,2}^{cc}(\boldsymbol{Q}, \rho_1, \rho_2) - \rho_2(R_2 + \rho_1 R_1), \tag{88}$$

where

$$E_{0,1}^{cc}(\boldsymbol{Q}, \rho_1) \triangleq \sup_{s \geq 0, a_1(\cdot), a_1'(\cdot), a_2'(\cdot)} - \log \mathbb{E}\left[\left(\frac{\mathbb{E}\left[q(\overline{X}_1, X_2, Y)^s e^{a_1(\overline{X}_1)} \mid X_2, Y \right]}{q(X_1, X_2, Y)^s e^{a_1(X_1)}} \right)^{\rho_1} e^{\sum_{\nu=1}^{2} a_\nu'(X_\nu) - \phi_\nu'} \right] \tag{89}$$

$$E_{0,2}^{cc}(\boldsymbol{Q}, \rho_2) \triangleq \sup_{s \geq 0, a_2(\cdot), a_1'(\cdot), a_2'(\cdot)} - \log \mathbb{E}\left[\left(\frac{\mathbb{E}\left[q(X_1, \overline{X}_2, Y)^s e^{a_2(\overline{X}_2)} \mid X_1, Y \right]}{q(X_1, X_2, Y)^s e^{a_2(X_2)}} \right)^{\rho_1} e^{\sum_{\nu=1}^{2} a_\nu'(X_\nu) - \phi_\nu'} \right] \tag{90}$$

$$E_{0,12,1}^{cc}(\boldsymbol{Q}, \rho_1, \rho_2) \triangleq \sup_{s \geq 0, a_1(\cdot), a_2(\cdot), a_1'(\cdot), a_2'(\cdot)}$$
$$- \log \mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{\mathbb{E}\left[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a_2(\overline{X}_2)} \mid \overline{X}_1 \right]}{q(X_1, X_2, Y)^s e^{a_2(X_2)}} \right)^{\rho_2} \frac{e^{a_1(\overline{X}_1)}}{e^{a_1(X_1)}} \;\middle|\; X_1, X_2, Y \right] \right)^{\rho_1} e^{\sum_{\nu=1}^{2} a_\nu'(X_\nu) - \phi_\nu'} \right] \tag{91}$$

$$E_{0,12,2}^{cc}(\boldsymbol{Q}, \rho_1, \rho_2) \triangleq \sup_{s \geq 0, a_1(\cdot), a_2(\cdot), a_1'(\cdot), a_2'(\cdot)}$$
$$- \log \mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{\mathbb{E}\left[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a_1(\overline{X}_1)} \mid \overline{X}_2 \right]}{q(X_1, X_2, Y)^s e^{a_1(X_1)}} \right)^{\rho_1} \frac{e^{a_2(\overline{X}_2)}}{e^{a_2(X_2)}} \;\middle|\; X_1, X_2, Y \right] \right)^{\rho_2} e^{\sum_{\nu=1}^{2} a_\nu'(X_\nu) - \phi_\nu'} \right], \tag{92}$$

and where $(X_1, X_2, Y, \overline{X}_1, \overline{X}_2) \sim Q_1(x_1) Q_2(x_2) W(y|x_1, x_2) Q_1(\overline{x}_1) Q_2(\overline{x}_2)$ and $\phi_\nu' \triangleq \mathbb{E}_{Q_\nu}[a_\nu'(X_\nu)]$ $(\nu = 1, 2)$. Each supremum over $a_\nu(\cdot)$ and $a_\nu'(\cdot)$ is taken over all real-valued functions on $\mathcal{X}_\nu$.

Proof: Each expression follows from its counterpart in (55)–(56) and (76)–(77) using Lemma 5 and Fan's minimax theorem [30]; details are given in Appendix C. ∎

The expressions in (87)–(88) contain two parameters ρ_1 and ρ_2, thus differing from the usual error exponents containing a single value of ρ (e.g. [31, Ch. 5]). These two parameters correspond to the two $\min\{1,\cdot\}$ terms in the RCU bounds (42)–(43). To our knowledge, the first exponents of this form were by Poltyrev [32] in the setting of broadcast channels with degraded message sets.

The following dual expression for the achievable rate region follows from (85)–(92).

Theorem 5. *The region $\mathcal{R}^{\mathrm{LM}}(\boldsymbol{Q})$ in (69)–(71) can be expressed as the set of rate pairs (R_1, R_2) satisfying*

$$R_1 \leq \sup_{s \geq 0, a_1(\cdot)} \mathbb{E}\left[\log \frac{q(X_1, X_2, Y)^s e^{a_1(X_1)}}{\mathbb{E}\big[q(\overline{X}_1, X_2, Y)^s e^{a_1(\overline{X}_1)} \,\big|\, X_2, Y\big]}\right] \tag{93}$$

$$R_2 \leq \sup_{s \geq 0, a_2(\cdot)} \mathbb{E}\left[\log \frac{q(X_1, X_2, Y)^s e^{a_2(X_2)}}{\mathbb{E}\big[q(X_1, \overline{X}_2, Y)^s e^{a_2(\overline{X}_2)} \,\big|\, X_1, Y\big]}\right] \tag{94}$$

and at least one of

$$R_1 \leq \sup_{\rho_2 \in [0,1], s \geq 0, a_1(\cdot), a_2(\cdot)} \mathbb{E}\left[\log \frac{\big(q(X_1, X_2, Y)^s e^{a_2(X_2)}\big)^{\rho_2} e^{a_1(X_1)}}{\mathbb{E}\Big[\big(\mathbb{E}\big[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a_2(\overline{X}_2)} \,\big|\, \overline{X}_1\big]\big)^{\rho_2} e^{a_1(\overline{X}_1)} \,\big|\, Y\Big]}\right] - \rho_2 R_2 \tag{95}$$

$$R_2 \leq \sup_{\rho_1 \in [0,1], s \geq 0, a_1(\cdot), a_2(\cdot)} \mathbb{E}\left[\log \frac{\big(q(X_1, X_2, Y)^s e^{a_1(X_1)}\big)^{\rho_1} e^{a_2(X_2)}}{\mathbb{E}\Big[\big(\mathbb{E}\big[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a_1(\overline{X}_1)} \,\big|\, \overline{X}_2\big]\big)^{\rho_1} e^{a_2(\overline{X}_2)} \,\big|\, Y\Big]}\right] - \rho_1 R_1 \tag{96}$$

where $(X_1, X_2, Y, \overline{X}_1, \overline{X}_2) \sim Q_1(x_1)Q_2(x_2)W(y|x_1,x_2)Q_1(\overline{x}_1)Q_2(\overline{x}_2)$, and for $\nu = 1, 2$, each supremum over $a_\nu(\cdot)$ is taken over all real-valued functions on $\mathcal{X}_\nu$.

Proof: The conditions in (93)–(96) follow from (85)–(88) respectively using a similar argument to Gallager's single-user analysis [33]. For example, (95) follows from (87) by setting $a'_1(x_1) = a'_2(x_2) = 0$, fixing s, $a_1(\cdot)$ and $a_2(\cdot)$, and taking the limit as $\rho_1 \to 0$. The equivalence to (69)–(71) follows by making use of the alternative conditions in (83)–(84) and applying Lagrange duality techniques analogously to Theorem 4. For example, for showing the equivalence of (83) and (95), the desired result follows from Steps 1–2 of the proof of Theorem 4 in Appendix C-A, using $Q_1 \times Q_2 \times W$ in place of $P_{X_1 X_2 Y}$. ∎

C. Exponents and Rates for General Alphabets

In this section, we show that the expressions given in Theorems 4 and 5 remain valid for the memoryless MAC with general alphabets. We make use of the cost-constrained ensemble [8], [15], [31], defined as follows. We fix $Q_1 \in \mathcal{P}(\mathcal{X}_1)$ and $Q_2 \in \mathcal{P}(\mathcal{X}_1)$ and choose

$$P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu) = \frac{1}{\mu_{\nu,n}} \prod_{i=1}^{n} Q_\nu(x_{\nu,i}) \mathbb{1}\big\{\boldsymbol{x}_\nu \in \mathcal{D}_{\nu,n}\big\}, \tag{97}$$

where for $\nu = 1, 2$, $\mu_{\nu,n}$ is a normalizing constant, and

$$\mathcal{D}_{\nu,n} \triangleq \left\{\boldsymbol{x}_\nu \,:\, \left|\frac{1}{n}\sum_{i=1}^{n} a_{\nu,l}(x_{\nu,i}) - \phi_{\nu,l}\right| \leq \frac{\delta}{n}, \, l = 1, \ldots, L_\nu\right\} \tag{98}$$

$$\phi_{\nu,l} \triangleq \mathbb{E}_{Q_\nu}\left[a_{\nu,l}(X_\nu)\right], \tag{99}$$

where $\{a_{\nu,l}\}_{l=1}^{L_\nu}$ are cost functions, and δ is a positive constant. Thus, the codewords for user ν are constrained to satisfy L_ν cost constraints in which the empirical average of $a_{\nu,l}(\cdot)$ is close to the true mean. We allow each of the

parameters to be optimized, including the cost functions. In principle, δ could vary with l, ν and n, but a fixed value will suffice for our purposes. By choosing $L_1 = L_2 = 0$, we recover the i.i.d. ensemble studied in [17].

The cost-constrained ensemble has primarily been used with $L_\nu = 1$ [15], [31], but the inclusion of multiple cost functions has proven beneficial in the mismatched single-user setting [8], [34]. In this section, we will see that the use of multiple costs is beneficial for both the matched and mismatched MAC. It should be noted that the costs are not being introduced to meet system constraints (e.g. power limitations), but instead to improve the performance of the random-coding ensemble itself. Thus, the costs can be thought of as *auxiliary costs* rather than *system costs*. The latter can easily be handled similarly to the former (e.g. see [8, Sec. VII], [34]), but in this paper we assume that the channel is unconstrained.

The following proposition from [8] will be useful.

Proposition 1. [8, Prop. 1] *For $\nu = 1, 2$, fix L_ν and the cost functions $\{a_{\nu,l}\}_{l=1}^{L_\nu}$. Then $\mu_{\nu,n} = \Omega(n^{-L_\nu/2})$ provided that the second moment of $a_{\nu,l}(X_\nu)$ is finite under $X_\nu \sim Q_\nu$ for $l = 1, \ldots, L_\nu$.*

We begin with the following generalization of Theorem 4.

Theorem 6. *For any memoryless MAC, the error exponents in (85)–(92) are achievable, where each supremum is subject to the constraints $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)^2] < \infty$ and $\mathbb{E}_{Q_\nu}[a'_\nu(X_\nu)^2] < \infty$ ($\nu = 1, 2$). Furthermore, these exponents can be achieved using the cost-constrained ensemble with $L_1 = L_2 = 5$.*

Proof: The proof is similar for each exponent, so we focus on $E_{r,12,1}$. We define $Q_\nu^n(\boldsymbol{x}_\nu) \triangleq \prod_{i=1}^n Q_\nu(x_{\nu,i})$ for $\nu = 1, 2$. Expanding (42) and applying Markov's inequality and $\min\{1, \alpha\} \le \alpha^\rho$ ($0 \le \rho \le 1$), we obtain[1]

$$\mathrm{rcu}_{12,1}(n, M_1)$$
$$\le \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \left(M_1 \sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) \left(M_2 \frac{\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) q^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y})^s}{q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})^s} \right)^{\rho_2} \right)^{\rho_1} \quad (100)$$

for any $\rho_1 \in [0, 1]$, $\rho_2 \in [0, 1]$ and $s \ge 0$. For $\nu = 1, 2$, we let $a_\nu(x)$ and $a'_\nu(x)$ be two of the five cost functions in the ensemble, and we define $a_\nu^n(\boldsymbol{x}_\nu) \triangleq \sum_{i=1}^n a_\nu(x_{\nu,i})$, $a'^n_\nu(\boldsymbol{x}_\nu) \triangleq \sum_{i=1}^n a'_\nu(x_{\nu,i})$, $\phi_\nu \triangleq \mathbb{E}_{Q_\nu}[a_\nu(X_\nu)]$ and $\phi'_\nu \triangleq \mathbb{E}_{Q_\nu}[a'_\nu(X_\nu)]$. Using the bounds on the cost functions in (98), we obtain

$$\mathrm{rcu}_{12,1}(n, M_1) \le e^{2\delta(\rho_1 + \rho_1\rho_2 + 1)} \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \left(M_1 \sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) \right.$$
$$\left. \times \left(M_2 \frac{\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) q^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y})^s e^{a_2^n(\overline{\boldsymbol{x}}_2)}}{q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})^s e^{a_2^n(\boldsymbol{x}_2)}} \right)^{\rho_2} \frac{e^{a_1^n(\overline{\boldsymbol{x}}_1)}}{e^{a_1^n(\boldsymbol{x}_1)}} \right)^{\rho_1} e^{\sum_{\nu=1}^2 a'^n_\nu(\boldsymbol{x}_\nu) - n\phi_\nu}. \quad (101)$$

We further upper bound (101) by substituting (97) and replacing the summations over $\mathcal{D}_{\nu,n}$ by summations over all sequences on $\mathcal{X}_\nu^n$. Expanding the resulting terms (e.g. $W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)$) as a product from 1 to n and taking the supremum over (s, ρ_1, ρ_2) and the cost functions, we obtain a bound whose exponent is given by (87), with a prefactor of

$$\frac{e^{2\delta(\rho_1 + \rho_1\rho_2 + 1)}}{\mu_{1,n}^{1+\rho_1} \mu_{2,n}^{1+\rho_1\rho_2}}. \quad (102)$$

From Proposition 1 and the assumptions on the second moments of the cost functions, this prefactor is subexponential in n.

Dealing with the other three exponents in (85)–(92) similarly, we conclude that it suffices to choose $L_1 = L_2 = 5$. To see this, one can simply count the number of occurrences of each $a_\nu(\cdot)$ and $a'_\nu(\cdot)$ in (89)–(92). It suffices to let the cost functions for $E_{0,12,1}$ and $E_{0,12,2}$ coincide, since the error exponent for the type-12 event is the maximum of the

---

[1]In the case of continuous alphabets, the summations should be replaced by integrals.

two. ■

From Theorem 6, we immediately obtain the following generalization of Theorem 5.

Theorem 7. *An achievable rate region for the memoryless MAC is given by $\mathcal{R}^{\mathrm{LM}}(\boldsymbol{Q})$ in (93)–(96), where each supremum is subject to $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)^2] < \infty$ ($\nu = 1, 2$). Furthermore, any point in the region can be achieved using cost-constrained random coding with $L_1 = L_2 = 3$.*

Proof: The proof of the first part is identical to that of Theorem 5. The sufficiency of $L_1 = L_2 = 3$ follows from the fact that the functions $a_1'(\cdot)$ and $a_2'(\cdot)$ are set to zero in the proof of Theorem 5. ■

D. Application to the Matched MAC

In this section, we apply our results to the setting of ML decoding, where $q(x_1, x_2, y) = W(y|x_1, x_2)$. As mentioned in Section I-C, an exponent for the constant-composition ensemble was derived by Liu and Hughes [19], and was shown to yield a strict improvement over Gallager's exponent for the i.i.d. ensemble [17] even after the optimization of the input distributions. This is in contrast to the single-user setting, where the two coincide after the optimization of the input distribution [11].

1) Simplified Primal Expressions for the Exponents: We will show that, in contrast to the mismatched setting, the overall error exponent E_r^{cc} given in (67) is unchanged when $E_{r,12}^{\mathrm{cc}'}$ in (68) is used in place of $E_{r,12}^{\mathrm{cc}}$. We begin with the following proposition, which shows that the exponents $(E_{r,1}^{\mathrm{cc}}, E_{r,2}^{\mathrm{cc}}, E_{r,12}^{\mathrm{cc}'})$ (see (55)–(56) and (68)) under ML decoding coincide with those of Liu and Hughes in the absence of time-sharing [19].

Proposition 2. *Under ML decoding (i.e. $q(x_1, x_2, y) = W(y|x_1, x_2)$), $E_{r,1}^{\mathrm{cc}}$, $E_{r,2}^{\mathrm{cc}}$ and $E_{r,12}^{\mathrm{cc}'}$ can be expressed as*

$$E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1) = \min_{P_{X_1 X_2 Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \big[I_P(X_1; X_2, Y) - R_1\big]^+ \tag{103}$$

$$E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2) = \min_{P_{X_1 X_2 Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \big[I_P(X_2; X_1, Y) - R_2\big]^+ \tag{104}$$

$$E_{r,12}^{\mathrm{cc}'}(\boldsymbol{Q}, R_1, R_2) = \min_{P_{X_1 X_2 Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \big[D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times P_Y) - (R_1 + R_2)\big]^+. \tag{105}$$

Proof: This proposition states that in the matched case, the minimizations over $\widetilde{P}_{X_1 X_2 Y}$ in (55), (56) and (68) are achieved by $\widetilde{P}_{X_1 X_2 Y} = P_{X_1 X_2 Y}$. The proof is nearly identical to [11, Lemma 9] and is omitted. ■

The statement that $E_{r,12}^{\mathrm{cc}}$ can be replaced by $E_{r,12}^{\mathrm{cc}'}$ is formally given in the following theorem.

Theorem 8. *Under ML decoding (i.e. $q(x_1, x_2, y) = W(y|x_1, x_2)$), we have for any input distributions (Q_1, Q_2) and rates (R_1, R_2) that*

$$\min\big\{E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2)\big\} = \min\big\{E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), E_{r,12}^{\mathrm{cc}'}(\boldsymbol{Q}, R_1, R_2)\big\}. \tag{106}$$

Proof: We know that $E_{r,12}^{\mathrm{cc}} \geq E_{r,12}^{\mathrm{cc}'}$ always holds, and hence the left-hand side of (106) is greater than or equal to the right-hand side. It remains to show that the converse is also true. To this end, we will show that the converse statement holds when $E_{r,12}^{\mathrm{cc}}$ is replaced by a quantity which is greater than or equal to $E_{r,12}^{\mathrm{cc}}$. From the definition of $\mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1 X_2 Y})$, $\widetilde{P}_{X_1 X_2 Y} = P_{X_1 X_2 Y}$ always satisfies the constraints of (58), and hence

$$E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \leq F_{12}(\boldsymbol{Q}, R_1, R_2), \tag{107}$$

where

$$F_{12}(\boldsymbol{Q}, R_1, R_2) \triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W)$$
$$+ \left[\max \left\{ I_P(X_1; Y) - R_1, I_P(X_2; Y) - R_2, D\big(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times P_Y\big) - R_1 - R_2 \right\} \right]^+. \quad (108)$$

We wish to show that

$$\min \left\{ E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), F_{12}(\boldsymbol{Q}, R_1, R_2) \right\} \leq \min \left\{ E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), E_{r,12}^{\mathrm{cc}'}(\boldsymbol{Q}, R_1, R_2) \right\}. \quad (109)$$

It suffices to show that whenever F_{12} exceeds $E_{r,12}^{\mathrm{cc}'}$, F_{12} also greater than or equal to either $E_{r,1}^{\mathrm{cc}}$ or $E_{r,2}^{\mathrm{cc}}$. Comparing (105) and (108), the objective in (108) only exceeds that of (105) when the maximum in (108) is achieved by $I_P(X_1; Y) - R_1$ or $I_P(X_2; Y) - R_2$. We show that the former implies $F_{12} \geq E_{r,1}^{\mathrm{cc}}$; it can similarly be shown that the latter implies $F_{12} \geq E_{r,2}^{\mathrm{cc}}$. If $I_P(X_1; Y) - R_1$ achieves the maximum, we have

$$I_P(X_1; Y) - R_1 \geq D\big(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times P_Y\big) - R_1 - R_2. \quad (110)$$

Using the identity

$$D\big(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times P_Y\big) = I_P(X_1; Y) + I_P(X_2; X_1, Y), \quad (111)$$

we can equivalently write (110) as $I_P(X_2; X_1, Y) \leq R_2$. However, substituting any such $P_{X_1 X_2 Y}$ into (104) yields an objective of $D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W)$, which cannot exceed the objective in (108). It follows that $F_{12} \geq E_{r,1}^{\mathrm{cc}}$. ∎

While it is possible that $E_{r,12}^{\mathrm{cc}} > E_{r,12}^{\mathrm{cc}'}$ under ML decoding, Theorem 8 shows that this never occurs in the region where $E_{r,12}$ achieves the minimum in (67). Thus, combining Theorem 8 and Theorem 2, we conclude that the exponent given in [19] is the true exponent for the constant-composition ensemble under ML decoding. While we have only proved this in the absence of time-sharing, the same holds true when time-sharing is used; see Section III-E for further discussion.

2) Dual Expression for the Exponent: The following theorem gives dual expressions for the exponents of Liu and Hughes [19] in a form closely resembling Gallager's exponents for the i.i.d. ensemble [17].

Theorem 9. *For any DM-MAC with ML decoding (i.e. $q(x_1, x_2, y) = W(y|x_1, x_2)$), the error exponents for constant-composition random coding* (103)–(105) *can be expressed as*

$$E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1) = \sup_{\rho \in [0,1]} E_{0,1}^{\mathrm{cc}}(\boldsymbol{Q}, \rho) - \rho R_1 \quad (112)$$

$$E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2) = \sup_{\rho \in [0,1]} E_{0,2}^{\mathrm{cc}}(\boldsymbol{Q}, \rho) - \rho R_2 \quad (113)$$

$$E_{r,12}^{\mathrm{cc}'}(\boldsymbol{Q}, R_1, R_2) = \sup_{\rho \in [0,1]} E_{0,12}^{\mathrm{cc}'}(\boldsymbol{Q}, \rho) - \rho(R_1 + R_2), \quad (114)$$

where

$$E_{0,1}^{\mathrm{cc}}(\boldsymbol{Q}, \rho) = \sup_{a_1(\cdot), a_2(\cdot)} - \log \sum_{x_2, y} Q_2(x_2) \left(\sum_{x_1} Q_1(x_1) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu} \right)^{1+\rho} \quad (115)$$

$$E_{0,2}^{\mathrm{cc}}(\boldsymbol{Q}, \rho) = \sup_{a_1(\cdot), a_2(\cdot)} - \log \sum_{x_1, y} Q_1(x_1) \left(\sum_{x_2} Q_2(x_2) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu} \right)^{1+\rho} \quad (116)$$

$$E_{0,12}^{\mathrm{cc}'}(\boldsymbol{Q}, \rho) = \sup_{a_1(\cdot), a_2(\cdot)} - \log \sum_y \left(\sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu} \right)^{1+\rho}. \quad (117)$$

Furthermore, these exponents are achievable in the case of general alphabets using the cost-constrained ensemble with $L_1 = L_2 = 3$, provided that each supremum is restricted to cost functions $a_\nu(X_\nu)$ whose second moment is finite

under $X_\nu \sim Q_\nu$.

Proof: By definition, $E_{r,12}^{\text{cc}'}$ can be obtained from either of (76)–(77) by removing the first term in the $\max\{\cdot,\cdot\}$. From the proof of Theorem 4, we conclude that the analogous statement in the dual expressions (85)–(92) is that the supremum over ρ_2 in (91) (respectively, (92)) is replaced by $\rho_2 = 1$ (respectively, $\rho_1 = 1$). With these substitutions, (91) and (92) become equal. Thus, it suffices to show that (89)–(91) coincide with (115)–(117) under ML decoding. We focus on (91) and (117), since the others are handled similarly.

Setting $\rho_2 = 1$ in (91) and renaming ρ_1 as ρ, we obtain the E_0 function

$$E_{0,12}^{\text{cc}''}(\boldsymbol{Q}, \rho) \triangleq -\log \sum_{x_1, x_2, y} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2)$$
$$\times \left(\frac{\sum_{\overline{x}_1, \overline{x}_2} Q_1(\overline{x}_1) Q_2(\overline{x}_2) W(y|\overline{x}_1, \overline{x}_2)^s e^{\sum_{\nu=1}^2 a_\nu(\overline{x}_\nu) - \phi_\nu}}{W(y|x_1, x_2)^s e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu}} \right)^\rho e^{\sum_{\nu=1}^2 a'_\nu(x_\nu) - \phi'_\nu} \tag{118}$$

$$= -\log \sum_y \left(\sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2)^{1-\rho s} e^{\sum_{\nu=1}^2 a'_\nu(x_\nu) - \phi'_\nu - \rho(a_\nu(x_\nu) - \phi_\nu)} \right)$$
$$\times \left(\sum_{\overline{x}_1, \overline{x}_2} Q_1(\overline{x}_1) Q_2(\overline{x}_2) W(y|\overline{x}_1, \overline{x}_2)^s e^{\sum_{\nu=1}^2 a_\nu(\overline{x}_\nu) - \phi_\nu} \right)^\rho, \tag{119}$$

where in (118) we multiplied the numerator and denominator of the fraction by $e^{-(\phi_1 + \phi_2)}$, where $\phi_\nu \triangleq \mathbb{E}_{Q_\nu}[a_\nu(X_\nu)]$. Using Hölder's inequality similarly to [31, Ex. 5.6], we can lower bound the argument to the logarithm in (119) by

$$\sum_y \left(\sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu} \right)^{1+\rho}. \tag{120}$$

Furthermore, using the sufficient condition for equality in [31, Ex. 5.6], we conclude that this lower bound can be achieved by choosing $s = \frac{1}{1+\rho}$ and $a'_\nu(x_\nu) = (1+\rho) a_\nu(x_\nu)$. By handling the other E_0 functions similarly, we obtain the first part of the theorem. For the second part, it only remains to show that the choice $L_1 = L_2 = 3$ suffices. This is seen by noting that whenever two cost functions are scalar multiples of each other (in this case, $a_\nu(\cdot)$ and $a'_\nu(\cdot)$) one of them can be removed without affecting the exponent obtained. ∎

It is possible to derive Theorem 9 more directly, without using the mismatched decoding results of the previous sections; see [35, Sec. V] for details. Gallager's exponents [17] are obtained by setting $a_1(\cdot) = a_2(\cdot) = 0$ in (115)–(117).

E. Time-Sharing

In this subsection, we discuss the concept of time-sharing for the MAC. We begin with a brief overview of time-sharing for the matched MAC, which is included in order to elucidate the subtle issues which arise in the mismatched setting.

1) Matched Setting: For any pair of input distributions (Q_1, Q_2), the achievable rate region for ML decoding is pentagonal and therefore convex. However, the region obtained by taking the union over all (Q_1, Q_2) is non-convex in general. To obtain the remainder of the capacity region, one can use the standard time-sharing argument in which one pair of codebooks is used for λn symbols of the block, and another pair of codebooks is used for the remaining $(1 - \lambda)n$ symbols. We refer to this scheme as explicit time-sharing, or simply time-sharing.

A drawback of explicit time-sharing is that the corresponding error exponents can be low. If $E_r^{(1)}$ is the exponent for some rate pair $(R_1^{(1)}, R_2^{(1)})$ and $E_r^{(2)}$ is the exponent for another rate pair $(R_1^{(2)}, R_2^{(2)})$, then the resulting exponent for the weighted rate pair

$$R_\nu = \lambda R_\nu^{(1)} + (1 - \lambda) R_\nu^{(2)}, \quad \nu = 1, 2 \tag{121}$$

is given by

$$E_r(R_1, R_2) = \min\left\{\lambda E_r^{(1)}(R_1^{(1)}, R_2^{(1)}), (1-\lambda)E_r^{(2)}(R_1^{(2)}, R_2^{(2)})\right\}, \tag{122}$$

which may be particularly low if λ is close to zero or one. Nevertheless, we have $E_r(R_1, R_2) > 0$ provided that $E_r^{(1)} > 0$, $E_r^{(2)} > 0$ and $\lambda \in (0,1)$.

Higher error exponents can be obtained by considering the coded time-sharing ensemble [28], which directly achieves any point in the capacity region. Each codeword is generated conditionally on a (possibly random) time-sharing sequence $\boldsymbol{U}$ in $\mathcal{U}^n$, where $\mathcal{U}$ is a finite time-sharing alphabet. For example, one may consider conditional i.i.d. random coding of the form

$$P_{\boldsymbol{X}_\nu|\boldsymbol{U}}(\boldsymbol{x}_\nu|\boldsymbol{u}) = \prod_{i=1}^{n} Q_\nu(x_{\nu,i}|u_i), \tag{123}$$

where Q_1 and Q_2 are (conditional) input distributions. Let $E_0^{(u)}$ denote the E_0 function (e.g. see (115)–(117)) for some error type under i.i.d. random coding on $Q_1(\cdot|u)$. Using coded time-sharing with $\mathcal{U} = \{1, 2\}$, $Q_U = (\lambda, 1-\lambda)$, and $\boldsymbol{U}$ equal to a deterministic sequence with type close to Q_U, we obtain the E_0 function [17]

$$E_0 = \lambda E_0^{(1)} + (1-\lambda)E_0^{(2)}. \tag{124}$$

In contrast to explicit time-sharing, we see that the exponent does not tend to zero as λ approaches zero or one.

Similar improvements apply for the constant-composition ensemble and cost-constrained ensemble. For the former, we let $Q_{\nu,n}$ be the closest conditional type to Q_ν for $\nu = 1, 2$, and write

$$P_{\boldsymbol{X}_\nu|\boldsymbol{U}}(\boldsymbol{x}_\nu|\boldsymbol{u}) = \frac{1}{|T_{\boldsymbol{u}}(Q_{\nu,n})|}\mathbb{1}\{\boldsymbol{x}_\nu \in T_{\boldsymbol{u}}(Q_{\nu,n})\}, \tag{125}$$

The cost-constrained ensemble is described by

$$P_{\boldsymbol{X}_\nu|\boldsymbol{U}}(\boldsymbol{x}_\nu|\boldsymbol{u}) = \frac{1}{\mu_{\nu,n}}\prod_{i=1}^{n} Q_\nu(x_{\nu,i}|u_i)\mathbb{1}\{\boldsymbol{x}_\nu \in \mathcal{D}_{\nu,n}\}, \tag{126}$$

where

$$\mathcal{D}_{\nu,n} \triangleq \left\{\boldsymbol{x}_\nu : \left|\sum_{i=1}^{n} a_{\nu,l}(u_i, x_{\nu,i}) - \phi_{\nu,l}\right| \leq \frac{\delta}{n}, l = 1, \ldots, L_\nu\right\} \tag{127}$$

and where each quantity is defined similarly to (98)–(99) (e.g. $\phi_{\nu,l} \triangleq \mathbb{E}_{Q_U \times Q_\nu}[a_{\nu,l}(U, X_\nu)]$).

2) Mismatched Setting: Under mismatched decoding, explicit time-sharing behaves similarly to the matched setting. For time-sharing between two achievable pairs $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$ with exponents $E_r^{(1)}$ and $E_r^{(2)}$, one can achieve the rate pair described by (121), with the exponent given by (122).

It may seem reasonable to conjecture that one can use coded time-sharing with $|\mathcal{U}| = 2$ to prove the achievability of any point on the line segment between two achievable points $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$. However, in the mismatched setting, this is not true in general. Before showing this, we outline some of the results when coded time-sharing is used. Since the changes compared to the case $\mathcal{U} = \emptyset$ are similar for each error type, we focus on the type-12 quantities.

For general conditional codeword distributions $P_{\boldsymbol{X}_\nu|\boldsymbol{U}}$ with a (possibly random) time-sharing sequence $\boldsymbol{U} \sim P_U$, the RCU bounds in (40)–(43) should be modified so that outer expectation is over $(\boldsymbol{U}, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$. For each of the ensembles in (125)–(126), we let $\boldsymbol{U}$ be a deterministic sequence $\boldsymbol{u} \in T^n(Q_{U,n})$, where $Q_{U,n}$ is the most likely type under Q_U. Following the analysis of Section III-B1, the analogous exponent to $E_{r,12}^{cc}$ in (58) is given by

$$E_{r,12}^{cc\text{-}ts}(\boldsymbol{Q}, R_1, R_2) \triangleq \min_{P_{UX_1X_2Y} \in \mathcal{S}^{cc\text{-}ts}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}^{cc\text{-}ts}(P_{UX_1X_2Y})} D(P_{UX_1X_2Y} \| Q_U \times Q_1 \times Q_2 \times W)$$

$$+ \left[\max\left\{I_{\widetilde{P}}(X_1; Y|U) - R_1, I_{\widetilde{P}}(X_2; Y|U) - R_2, D(\widetilde{P}_{UX_1X_2Y} \| Q_U \times Q_1 \times Q_2 \times \widetilde{P}_Y) - R_1 - R_2\right\}\right]^+, \tag{128}$$

where

$$\mathcal{S}^{\text{cc-ts}}(\boldsymbol{Q}) \triangleq \left\{ P_{UX_1X_2Y} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : P_U = Q_U, P_{X_1|U} = Q_1, P_{X_2|U} = Q_2 \right\} \tag{129}$$

$$\mathcal{T}^{\text{cc-ts}}_{12}(P_{UX_1X_2Y}) \triangleq \left\{ \widetilde{P}_{UX_1X_2Y} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}) : \right.$$
$$\left. \widetilde{P}_{X_1|U} = P_{X_1|U}, \widetilde{P}_{X_2|U} = P_{X_2|U}, \widetilde{P}_{UY} = P_{UY}, \mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \right\}. \tag{130}$$

In the dual domain, the analogous quantity to $E^{\text{cc}}_{0,12,1}$ in (91) is given by

$$E^{\text{cc-ts}}_{0,12,1}(\boldsymbol{Q}, \rho_1, \rho_2) \triangleq \sup_{s \geq 0, a_1(\cdot,\cdot), a_2(\cdot,\cdot), a'_1(\cdot,\cdot), a'_2(\cdot,\cdot)}$$
$$- \sum_u Q_U(u) \log \mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{\mathbb{E}\left[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a_2(u,\overline{X}_2)} \,\middle|\, \overline{X}_1 \right]}{q(X_1, X_2, Y)^s e^{a_2(u,X_2)}} \right)^{\rho_2} \frac{e^{a_1(u,\overline{X}_1)}}{e^{a_1(u,X_1)}} \,\middle|\, X_1, X_2, Y \right] \right)^{\rho_1} e^{\sum_{\nu=1}^2 a'_\nu(u,X_\nu) - \phi'_\nu} \right], \tag{131}$$

where for each u, the expectations are taken with respect to the conditional distribution $(X_1, X_2, Y, \overline{X}_1, \overline{X}_2 \,|\, U = u) \sim Q_1(x_1|u)Q_2(x_2|u)W(y|x_1,x_2)Q_1(\overline{x}_1|u)Q_2(\overline{x}_2|u)$. Similarly, the rate condition corresponding to (71) is

$$R_1 + R_2 \leq \min_{\substack{\widetilde{P}_{UX_1X_2Y} \in \mathcal{T}^{\text{cc-ts}}_{12}(Q_U \times Q_1 \times Q_2 \times W) \\ I_{\widetilde{P}}(X_1;Y|U) \leq R_1, I_{\widetilde{P}}(X_2;Y|U) \leq R_2}} D(\widetilde{P}_{UX_1X_2Y} \| Q_U \times Q_1 \times Q_2 \times \widetilde{P}_Y), \tag{132}$$

and the rate condition corresponding to (95) is

$$R_1 \leq \sup_{\rho_2 \in [0,1], s \geq 0, a_1(\cdot,\cdot), a_2(\cdot,\cdot)} \mathbb{E}\left[\log \frac{\left(q(X_1, X_2, Y)^s e^{a_2(U,X_2)} \right)^{\rho_2} e^{a_1(U,X_1)}}{\mathbb{E}\left[\left(\mathbb{E}\left[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a_2(U,\overline{X}_2)} \,\middle|\, \overline{X}_1 \right] \right)^{\rho_2} e^{a_1(U,\overline{X}_1)} \,\middle|\, Y \right]} \right] - \rho_2 R_2, \tag{133}$$

where $U \sim Q_U$.

We now compare the above rate conditions with those obtained by explicit time-sharing, focusing on the case that $\mathcal{U} = \{1, 2\}$. We again write $Q_U(1) = \lambda$ and $Q_U(2) = 1 - \lambda$, and we assume that the time-sharing is done using the input distribution vectors $\boldsymbol{Q}^{(1)}$ and $\boldsymbol{Q}^{(2)}$ and the rate pairs $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$. We make use of the quantities R_1 and R_2 defined in (121).

We first consider the primal domain expressions. We have from (122) that under explicit time-sharing, the condition corresponding to (71) is given by

$$R_1 \leq \lambda \min_{\widetilde{P}_{X_1X_2Y}} D(\widetilde{P}_{X_1X_2Y} \| Q_1^{(1)} \times Q_2^{(1)} \times \widetilde{P}_Y) + (1 - \lambda) \min_{\widetilde{P}_{X_1X_2Y}} D(\widetilde{P}_{X_1X_2Y} \| Q_1^{(2)} \times Q_2^{(2)} \times \widetilde{P}_Y), \tag{134}$$

where the minimizations are subject to constraints of the form in (71), the first using $\boldsymbol{Q}^{(1)}$ and the second using $\boldsymbol{Q}^{(2)}$. By writing the first minimization variable in (134) as $\widetilde{P}_{X_1X_2Y|U=1}$ and the second as $\widetilde{P}_{X_1X_2Y|U=2}$, we can combine these into a single minimization over $\widetilde{P}_{UX_1X_2Y}$, where $\widetilde{P}_U = Q_U$. The resulting rate condition has the same form as (132), but with the constraint $\mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]$ replaced by

$$\mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y) \,|\, U = u] \geq \mathbb{E}_P[\log q(X_1, X_2, Y) \,|\, U = u], \; u = 1, 2, \tag{135}$$

and each constraint $I_{\widetilde{P}}(X_\nu; Y|U) \leq R_\nu$ $(\nu = 1, 2)$ replaced by

$$I_{\widetilde{P}}(X_\nu; Y|U = u) \leq R_\nu^{(u)}, \quad u = 1, 2. \tag{136}$$

It follows that the constraints on $\widetilde{P}_{UX_1X_2Y}$ are stricter for explicit time-sharing than for coded time-sharing; the former must hold for each value of u, while the latter only need to hold for the average over u with respect to Q_U. Since the

minimization for the coded time-sharing ensemble is over a larger set, the resulting rate cannot be higher, and can in fact be strictly lower.

A similar observation applies in the dual domain. Let us write $I_{12}(\boldsymbol{Q}, s, \rho_2)$ to denote (95) with fixed values of s and ρ_2 in place of the suprema. Using explicit time-sharing, the condition corresponding to (93) is given by

$$R_1 \leq \lambda \sup_{s \geq 0, \rho_2 \in [0,1]} I_{12}(\boldsymbol{Q}^{(1)}, s, \rho_2) + (1 - \lambda) \sup_{s \geq 0, \rho_2 \in [0,1]} I_{12}(\boldsymbol{Q}^{(2)}, s, \rho_2), \tag{137}$$

whereas from (133), coded time-sharing only permits

$$R_1 \leq \sup_{s \geq 0, \rho_2 \in [0,1]} \left(\lambda I_{12}(\boldsymbol{Q}^{(1)}, s, \rho_2) + (1 - \lambda) I_{12}(\boldsymbol{Q}^{(2)}, s, \rho_2) \right). \tag{138}$$

In other words, the coded time-sharing ensemble only allows a single choice of (s, ρ_2), whereas explicit time-sharing allows (s, ρ_2) to vary with u. An analogy to the above discussion regarding the primal expressions can be seen by interpreting (s, ρ_2) as Lagrange multipliers.

In summary, the achievable rate regions of the coded time-sharing ensembles can be strictly smaller than those of explicit time-sharing in the mismatched setting, in contrast to the matched setting. However, from the above discussion preceding (122), we expect coded time-sharing to yield higher error exponents for many channels and metrics, at least for some rate pairs. It remains an open question as to whether the rate region of explicit time-sharing can be achieved using a different approach which yields better exponents.

F. Application to Single-User Mismatched Channels

The parallel BSC example [6] given in Section I-C shows that the achievable rate region for the MAC can be used to improve on the LM rate in the single-user setting. In [6, Thm. 4], Lapidoth showed that we can in fact do even better by expurgating all codeword pairs except for those have a fixed joint type approximately equal to $Q_1 \times Q_2$. Such expurgation is not possible for the mismatched MAC, since it requires cooperation between the users.

We state the achievable rate region of [6, Thm. 4] in the following theorem, making using of the sets

$$\mathcal{T}_\nu^{\text{cc-ex}}(Q_1 \times Q_2 \times W) \triangleq \mathcal{T}_\nu^{\text{cc}}(Q_1 \times Q_2 \times W) \cap \left\{ \widetilde{P}_{X_1 X_2 Y} : \widetilde{P}_{X_1 X_2} = Q_1 \times Q_2 \right\} \tag{139}$$

for $\nu = 1, 2, 12$, where the sets $\mathcal{T}_\nu^{\text{cc}}$ are defined in (49)–(51).

Theorem 10. [6, Thm. 4] *Let the DMC $W'(y|x)$ and metric $q'(x, y)$ be given, and let $W(y|x_1, x_2) = W'(y|\psi(x_1, x_2))$ and $q(x_1, x_2, y) = q'(\psi(x_1, x_2), y)$ for some finite alphabets $\mathcal{X}_1$, $\mathcal{X}_2$ and function $\psi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}$. An achievable rate for W' with the metric q' is given by $R = R_1 + R_2$ for any (R_1, R_2) satisfying*

$$R_1 \leq \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_1^{\text{cc-ex}}(Q_1 \times Q_2 \times W)} I_{\widetilde{P}}(X_1; Y | X_2) \tag{140}$$

$$R_2 \leq \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_2^{\text{cc-ex}}(Q_1 \times Q_2 \times W)} I_{\widetilde{P}}(X_2; Y | X_1) \tag{141}$$

$$R_1 + R_2 \leq \min_{\substack{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{cc-ex}}(Q_1 \times Q_2 \times W) \\ I_{\widetilde{P}}(X_1; Y) \leq R_1, \, I_{\widetilde{P}}(X_2; Y) \leq R_2}} I_{\widetilde{P}}(X_1, X_2; Y) \tag{142}$$

for some Q_1, Q_2.

Analogously to Lemma 8, we have the following.

Lemma 9. *The achievable rate region in Theorem 10 is unchanged if the condition that* (142) *holds is replaced by*

the condition that at least one of the following hold:

$$R_1 + R_2 \leq \min_{\substack{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{cc-ex}}(Q_1 \times Q_2 \times W) \\ I_{\widetilde{P}}(X_1;Y) \leq R_1}} I_{\widetilde{P}}(X_1, X_2; Y). \tag{143}$$

$$R_1 + R_2 \leq \min_{\substack{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{cc-ex}}(Q_1 \times Q_2 \times W) \\ I_{\widetilde{P}}(X_2;Y) \leq R_2}} I_{\widetilde{P}}(X_1, X_2; Y). \tag{144}$$

Proof: The condition in (142) can equivalently be written as

$$0 \leq \min_{\widetilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{cc-ex}}(Q_1 \times Q_2 \times W)} \max \Big\{ I_{\widetilde{P}}(X_1;Y) - R_1, I_{\widetilde{P}}(X_2;Y) - R_2, I_{\widetilde{P}}(X_1, X_2; Y) - R_1 - R_2 \Big\}. \tag{145}$$

The lemma follows by writing (143)–(144) in a similar form and applying Lemma 6, analogously to the proof of Lemma 7. ∎

The following theorem is analogous to Theorem 4, and can be proved similarly using Lagrange duality techniques. The only significant difference in the derivation is the additional constraint $\widetilde{P}_{X_1 X_2} = Q_1 \times Q_2$ in each primal expression, and the presence of joint functions $a(x_1, x_2)$ in each dual expression rather than separate functions $a_1(x_1)$ and $a_2(x_2)$. The proof is omitted to avoid repetition.

Theorem 11. *Under the setup of Theorem 10, the pair (R_1, R_2) satisfies (140)–(142) if and only if it satisfies*

$$R_1 \leq \sup_{s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{q(X_1, X_2, Y)^s e^{a(X_1, X_2)}}{\mathbb{E}\big[q(\overline{X}_1, X_2, Y)^s e^{a(\overline{X}_1, X_2)} \mid X_2, Y\big]} \right] \tag{146}$$

$$R_2 \leq \sup_{s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{q(X_1, X_2, Y)^s e^{a(X_1, X_2)}}{\mathbb{E}\big[q(X_1, \overline{X}_2, Y)^s e^{a(X_1, \overline{X}_2)} \mid X_1, Y\big]} \right] \tag{147}$$

and at least one of

$$R_1 \leq \sup_{\rho_2 \in [0,1], s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{\big(q(X_1, X_2, Y)^s e^{a(X_1, X_2)}\big)^{\rho_2}}{\mathbb{E}\Big[\big(\mathbb{E}\big[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a(\overline{X}_1, \overline{X}_2)} \mid \overline{X}_1\big]\big)^{\rho_2} \mid Y\Big]} \right] - \rho_2 R_2 \tag{148}$$

$$R_2 \leq \sup_{\rho_1 \in [0,1], s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{\big(q(X_1, X_2, Y)^s e^{a(X_1, X_2)}\big)^{\rho_1}}{\mathbb{E}\Big[\big(\mathbb{E}\big[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a(\overline{X}_1, \overline{X}_2)} \mid \overline{X}_2\big]\big)^{\rho_1} \mid Y\Big]} \right] - \rho_1 R_1 \tag{149}$$

where $(X_1, X_2, Y, \overline{X}_1, \overline{X}_2) \sim Q_1(x_1) Q_2(x_2) W(y|x_1, x_2) Q_1(\overline{x}_1) Q_2(\overline{x}_2)$.

It is possible to prove the achievability of $R = R_1 + R_2$ for (R_1, R_2) satisfying (146)–(149) directly, and to extend the validity of the result to general alphabets. The main difference compared to [6, Thm. 4] is that instead of expurgating all codewords except those of a given type, we expurgate all codewords which fail to meet a set of *joint* cost constraints of the form

$$\left| \frac{1}{n} \sum_{i=1}^{n} a_l(x_{1,i}, x_{2,i}) - \phi_l \right| \leq \delta_n \tag{150}$$

for some functions $\{a_l\}_{l=1}^{L}$ and a sequence δ_n, where $\phi_l \triangleq \mathbb{E}_{Q_1 \times Q_2}[a_l(X_1, X_2)]$.

We do not pursue the above approach further in this paper. Instead, we show in the following section that superposition coding yields an achievable rate at least as high as that of Theorem 10 for any DMC. The superposition coding approach extends more easily to general alphabets, without the need for expurgation. Furthermore, unlike the proof of Theorem 10 (see [6, Thm. 4]) or its variation outlined above, our analysis of superposition coding yields non-asymptotic bounds and error exponents as intermediate steps.

IV. SUPERPOSITION CODING

In this section, we study superposition coding (SC) [36]–[38] for the mismatched single-user channel introduced in Section I-A1. The parameters of the ensemble are an auxiliary alphabet $\mathcal{U}$, an auxiliary codeword distribution $P_{\boldsymbol{U}}$, and a conditional codeword distribution $P_{\boldsymbol{X}|\boldsymbol{U}}$. We fix the rates R_0 and R_1. An auxiliary codebook $\{\boldsymbol{U}^{(i)}\}_{i=1}^{M_0}$ with $M_0 \triangleq \exp(R_0)$ codewords is generated at random, with each auxiliary codeword independently distributed according to $P_{\boldsymbol{U}}$. For each $i = 1, \cdots, M_0$, a codebook $\{\boldsymbol{X}^{(i,j)}\}_{j=1}^{M_1}$ with $M_1 \triangleq \exp(R_1)$ codewords is generated at random, with each codeword conditionally independently distributed according to $P_{\boldsymbol{X}|\boldsymbol{U}}$. The message m at the input to the encoder is indexed as (m_0, m_1), and for any such pair the corresponding transmitted codeword is $\boldsymbol{X}^{(m_0, m_1)}$. Thus, the overall number of messages is $M = M_1 M_2$, yielding a rate of $R = R_1 + R_2$. More compactly, we have

$$\left\{ \left(\boldsymbol{U}^{(i)}, \{\boldsymbol{X}^{(i,j)}\}_{j=1}^{M_1} \right) \right\}_{i=1}^{M_0} \sim \prod_{i=1}^{M_0} \left(P_{\boldsymbol{U}}(\boldsymbol{u}^{(i)}) \prod_{j=1}^{M_1} P_{\boldsymbol{X}|\boldsymbol{U}}(\boldsymbol{x}^{(i,j)}|\boldsymbol{u}^{(i)}) \right). \tag{151}$$

We assume without loss of generality that message $(1, 1)$ is transmitted, and write $\boldsymbol{U}$ and $\boldsymbol{X}$ in place of $\boldsymbol{U}^{(1)}$ and $\boldsymbol{X}^{(1,1)}$ respectively. We write $\widetilde{\boldsymbol{X}}$ to denote an arbitrary codeword $\boldsymbol{X}^{(1,j)}$ with $j \neq 1$, and we write $\overline{\boldsymbol{U}}$ and $\overline{\boldsymbol{X}}$ to denote arbitrary codewords $\boldsymbol{U}^{(i)}$ and $\boldsymbol{X}^{(i,j)}$ with $i \neq 1$. Thus, defining $\boldsymbol{Y}$ to be the channel output, we have

$$(\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y}, \widetilde{\boldsymbol{X}}, \overline{\boldsymbol{U}}, \overline{\boldsymbol{X}}) \sim P_{\boldsymbol{U}}(\boldsymbol{u}) P_{\boldsymbol{X}|\boldsymbol{U}}(\boldsymbol{x}|\boldsymbol{u}) W^n(\boldsymbol{y}|\boldsymbol{x}) P_{\boldsymbol{X}|\boldsymbol{U}}(\widetilde{\boldsymbol{x}}|\boldsymbol{u}) P_{\boldsymbol{U}}(\overline{\boldsymbol{u}}) P_{\boldsymbol{X}|\boldsymbol{U}}(\overline{\boldsymbol{x}}|\overline{\boldsymbol{u}}). \tag{152}$$

The decoder forms an estimate $\hat{m} = (\hat{m}_0, \hat{m}_1)$ according to (2). We distinguish between the following two types of error:

(Type 0) $\hat{m}_0 \neq m_0$

(Type 1) $\hat{m}_0 = m_0$ and $\hat{m}_1 \neq m_1$

The random-coding error probabilities for a given random-coding ensemble are denoted by $\overline{p}_{e,0}(n, M_0, M_1)$ and $\overline{p}_{e,1}(n, M_1)$ respectively. The overall random-coding error probability $\overline{p}_e(n, M_0, M_1)$ thus satisfies

$$\max\{\overline{p}_{e,0}, \overline{p}_{e,1}\} \leq \overline{p}_e \leq \overline{p}_{e,0} + \overline{p}_{e,1}. \tag{153}$$

The primal-domain exponents and rates in this section (Theorems 13 and 14 below) coincide with those obtained in the independent work of Somekh-Baruch [9]. We thus omit their proofs, and focus our attention on results which are being presented here for the first time.

A. Random-Coding Bounds

In Section III, we used Lemma 3 to obtain ensemble-tight error exponents for the mismatched MAC. Since the proof used the method of types, the corresponding subexponential prefactors differ by a polynomial factor depending on the alphabet sizes. For superposition coding, we can in fact obtain non-asymptotic upper and lower bounds which coincide to within a constant factor.

Theorem 12. *Under the random-coding distributions $P_{\boldsymbol{U}}$ and $P_{\boldsymbol{X}|\boldsymbol{U}}$, the ensemble-average error probabilities for the maximum-metric decoder satisfy*

$$\frac{1}{8}\mathrm{rcu}_0(n, M_0, M_1) \leq \overline{p}_{e,0}(n, M_0, M_1) \leq \mathrm{rcu}_0(n, M_0, M_1), \tag{154}$$

$$\frac{1}{4}\mathrm{rcu}_1(n, M_1) \leq \overline{p}_{e,1}(n, M_1) \leq \mathrm{rcu}_1(n, M_1) \tag{155}$$

where

$$\mathrm{rcu}_0(n, M_0, M_1) \triangleq \mathbb{E}\left[\min\left\{1, (M_0-1)\mathbb{E}\left[\min\left\{1, M_1\mathbb{P}\left[\frac{q^n(\overline{\boldsymbol{X}}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1 \,\Big|\, \overline{\boldsymbol{U}}\right]\right\}\,\Big|\, \boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y}\right]\right\}\right]. \tag{156}$$

$$\mathrm{rcu}_1(n, M_1) \triangleq \mathbb{E}\left[\min\left\{1, (M_1-1)\mathbb{P}\left[\frac{q^n(\widetilde{\boldsymbol{X}}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1 \,\Big|\, \boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y}\right]\right\}\right] \tag{157}$$

Proof: Equation (155) follows using an identical argument to the single-user setting, where the RCU bound is known to be ensemble-tight to within a factor of $\frac{1}{4}$ [8]. We thus focus on (154). By upper bounding the error probability by that of a decoder which decodes ties as errors, we obtain

$$\overline{p}_{e,0} \leq \mathbb{P}\left[\bigcup_{i\neq 1, j\neq 1}\left\{\frac{q^n(\boldsymbol{X}^{(i,j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1\right\}\right]. \tag{158}$$

Writing the probability as an expectation given $(\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y})$ and applying the truncated union bound to the union over i, we obtain

$$\overline{p}_{e,0} \leq \mathbb{E}\left[\min\left\{1, (M_0-1)\mathbb{P}\left[\bigcup_{j\neq 1}\left\{\frac{q^n(\overline{\boldsymbol{X}}^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1\right\}\right]\right\}\right]. \tag{159}$$

Applying the same argument to the union over j, we obtain the upper bound in (154). The matching lower bound follows since breaking ties as errors increases the error probability by at most a factor of two [26], and since by the construction of the ensemble, each truncated union bound was applied to independent events, thus guaranteeing tightness to within a factor of $\frac{1}{2}$ [23, Lemma A.2]. ∎

B. Exponents and Rates for DMCs

In this subsection, we assume that the channel is a DMC, and consider the constant-composition superposition coding ensemble, which depends on a joint input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$. We let $Q_{UX,n}$ be the most probable type under Q_{UX}, and we denote the corresponding U marginal by $Q_{U,n}$ and the corresponding distribution of X given U by $Q_{X|U,n}$. The ensemble is characterized by

$$P_{\boldsymbol{U}}(\boldsymbol{u}) = \frac{1}{|T^n(Q_{U,n})|}\mathbb{1}\left\{\boldsymbol{u} \in T^n(Q_{U,n})\right\} \tag{160}$$

$$P_{\boldsymbol{X}|\boldsymbol{U}}(\boldsymbol{x}|\boldsymbol{u}) = \frac{1}{|T_{\boldsymbol{u}}^n(Q_{X|U,n})|}\mathbb{1}\left\{\boldsymbol{x} \in T_{\boldsymbol{u}}^n(Q_{X|U,n})\right\}. \tag{161}$$

We state the exponents and rates corresponding to the non-asymptotic bounds in Theorem 12 without proof, since the proofs are similar to those given in Section III, and since equivalent exponents and rates in a slightly different form have been given by Somekh-Baruch [9]. We define the sets

$$\mathcal{S}^{\mathrm{cc}}(Q_{UX}) \triangleq \left\{P_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : P_{UX} = Q_{UX}\right\} \tag{162}$$

$$\mathcal{T}_0^{\mathrm{cc}}(P_{UXY}) \triangleq \left\{\widetilde{P}_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : \widetilde{P}_{UX} = P_{UX}, \widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]\right\}. \tag{163}$$

$$\mathcal{T}_1^{\mathrm{cc}}(P_{UXY}) \triangleq \left\{\widetilde{P}_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : \widetilde{P}_{UX} = P_{UX}, \widetilde{P}_{UY} = P_{UY}, \mathbb{E}_{\widetilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]\right\} \tag{164}$$

Theorem 13. *The random-coding error probabilities for the constant-composition superposition coding ensemble satisfy*

$$\overline{p}_{e,0}(n, e^{nR_0}, e^{nR_1}) \doteq \exp\left(-nE_{r,0}^{\mathrm{cc}}(Q_{UX}, R_0, R_1)\right) \tag{165}$$

$$\overline{p}_{e,1}(n, e^{nR_1}) \doteq \exp\left(-nE_{r,1}^{\mathrm{cc}}(Q_{UX}, R_1)\right) \tag{166}$$

where

$$E_{r,0}^{\rm cc}(Q_{UX}, R_0, R_1) \triangleq \min_{P_{UXY} \in \mathcal{S}^{\rm cc}(Q_{UX})} \min_{\widetilde{P}_{UXY} \in \mathcal{T}_0^{\rm cc}(P_{UXY})}$$

$$D(P_{UXY}\|Q_{UX} \times W) + \left[I_{\widetilde{P}}(U;Y) + \left[I_{\widetilde{P}}(X;Y|U) - R_1\right]^+ - R_0\right]^+ \quad (167)$$

$$E_{r,1}^{\rm cc}(Q_{UX}, R_1) \triangleq \min_{P_{UXY} \in \mathcal{S}^{\rm cc}(Q_{UX})} \min_{\widetilde{P}_{UXY} \in \mathcal{T}_1^{\rm cc}(P_{UXY})} D(P_{UXY}\|Q_{UX} \times W) + \left[I_{\widetilde{P}}(X;Y|U) - R_1\right]^+ \quad (168)$$

The overall error exponent is given by

$$E_r^{\rm cc}(Q_{UX}, R_0, R_1) \triangleq \min\left\{E_{r,0}^{\rm cc}(Q_{UX}, R_0, R_1), E_{r,1}^{\rm cc}(Q_{UX}, R_1)\right\} \quad (169)$$

and is tight with respect to the ensemble average. Following the steps of Section III-D, it can be shown that in the case of ML decoding (i.e. $q(x,y) = W(y|x)$) the overall exponent can be expressed as

$$E_r^{\rm cc}(Q_{UX}, R_0, R_1) = \min\left\{E_{r,0}^{\rm cc'}(Q_{UX}, R_0, R_1), E_{r,1}^{\rm cc}(Q_{UX}, R_1)\right\}, \quad (170)$$

where

$$E_{r,0}^{\rm cc'}(Q_{UX}, R_0, R_1) \triangleq \min_{P_{UXY} \in \mathcal{S}^{\rm cc}(Q_{UX})} D(P_{UXY}\|Q_{UX} \times W) + [I_P(U,X;Y) - R_0 - R_1]^+ \quad (171)$$

$$E_{r,1}^{\rm cc}(Q_{UX}, R_1) = \min_{P_{UXY} \in \mathcal{S}^{\rm cc}(Q_{UX})} D(P_{UXY}\|Q_{UX} \times W) + [I_P(X;Y|U) - R_1]^+. \quad (172)$$

This exponent coincides with the error exponent for the primary user of a broadcast channel with degraded message sets given in [38]. Thus, the exponent given in [38] is tight with respect to the ensemble average for ML decoding.

Next, we provide the achievable rates resulting from Theorem 13.

Theorem 14. *The overall error exponent* $E_r^{\rm cc}(Q_{UX}, R_0, R_1)$ *is positive for all rate pairs* (R_1, R_2) *in the interior of* $\mathcal{R}^{\rm LM}(Q_{UX})$, *where* $\mathcal{R}^{\rm LM}(Q_{UX})$ *is the set of all rate pairs* (R_0, R_1) *satisfying*

$$R_1 \leq \min_{\widetilde{P}_{UXY} \in \mathcal{T}_1^{\rm cc}(Q_{UX} \times W)} I_{\widetilde{P}}(X;Y|U) \quad (173)$$

$$R_0 + R_1 \leq \min_{\substack{\widetilde{P}_{UXY} \in \mathcal{T}_0^{\rm cc}(Q_{UX} \times W) \\ I_{\widetilde{P}}(U;Y) \leq R_0}} I_{\widetilde{P}}(U,X;Y). \quad (174)$$

Proof: These expressions follow by taking $P_{UX} \to Q_{UX} \times W$ in (167)–(168), and noting that the objective of (167) is always positive when $I_{\widetilde{P}}(U;Y) > R_0$. $\blacksquare$

As discussed in Section III-B2, ensemble tightness of the exponent does not directly imply ensemble tightness of the rate. In the present setting, the ensemble tightness of the rate follows via a straightforward extension of [6, Thm. 3] (see [9]). From the discussion following (169), we recover the rate conditions for the primary user of the broadcast channel with degraded message sets [37] when $q(x,y) = W(y|x)$.

The following theorem gives dual expressions for the rates. In the following subsection we will use cost-constrained coding to show that these expressions remain valid in the case of continuous alphabets. Similar arguments can be applied to the exponents, but we focus on rates the sake of brevity.

Theorem 15. *The region* $\mathcal{R}^{\rm LM}(Q_{UX})$ *in (173)–(174) can be expressed as the set of all rate pairs* (R_0, R_1) *satisfying*

$$R_1 \leq \sup_{s \geq 0, a(\cdot, \cdot)} \mathbb{E}\left[\log \frac{q(X,Y)^s e^{a(U,X)}}{\mathbb{E}[q(\widetilde{X}, Y)^s e^{a(U, \widetilde{X})} \,|\, U, Y]}\right] \quad (175)$$

$$R_0 \leq \sup_{\rho_1 \in [0,1], s \geq 0, a(\cdot,\cdot)} \mathbb{E}\left[\log \frac{(q(X,Y)^s e^{a(U,X)})^{\rho_1}}{\mathbb{E}\left[\left(\mathbb{E}\left[q(\overline{X},Y)^s e^{a(\overline{U},\overline{X})} \,\middle|\, \overline{U} \right] \right)^{\rho_1} \,\middle|\, Y \right]} \right] - \rho_1 R_1, \tag{176}$$

where $(U, X, Y, \widetilde{X}, \overline{U}, \overline{X}) \sim Q_{UX}(u,x)W(y|x)Q_{X|U}(\widetilde{x}|u)Q_{UX}(\overline{u},\overline{x})$.

Proof: The proof is nearly identical to that of Theorem 5. Equation (175) follows from (173) analogously to the steps in obtaining (93)–(94) from (69)–(70), and (176) follows from (174) analogously to the steps in obtaining (95)–(96) from (83)–(84) (see the proof of Theorem 5 in Appendix C-A). ∎

C. Comparison to the Expurgated MAC

The achievable rate of Theorem 14 (superposition coding) has a similar form to that of Theorem 10 (expurgated parallel coding). The following proposition shows that the former can be weakened to the latter.

Proposition 3. *After the optimization of $\mathcal{U}$ and Q_{UX}, the achievable rate $R^{\mathrm{sc}} = \max_{R_0, R_1} R_0 + R_1$ described by (173)–(174) is at least as high as the achievable rate $R^{\mathrm{mac}} = \max_{R_1, R_2} R_1 + R_2$ described by (140)–(142) with optimized parameters $\mathcal{X}_1$, $\mathcal{X}_2$, Q_1, Q_2 and $\psi(\cdot, \cdot)$.*

Proof: We will make use of the equivalent type-12 conditions in (143)–(144) for the expurgated MAC. To avoid ambiguity between overloaded symbols (e.g. R_1), we use a superscript to specify the relevant ensemble for some quantities (e.g. R_1^{sc}, R_1^{mac}).

We will show that by identifying $U^{\mathrm{sc}} = X_1^{\mathrm{mac}}$, $R_1^{\mathrm{sc}} = R_1^{\mathrm{mac}}$ and $R_0^{\mathrm{sc}} = R_2^{\mathrm{mac}}$, we can weaken (173) to (140) and (174) to (144). A similar argument with $U^{\mathrm{sc}} = X_2^{\mathrm{mac}}$, $R_1^{\mathrm{sc}} = R_2^{\mathrm{mac}}$ and $R_0^{\mathrm{sc}} = R_1^{\mathrm{mac}}$ yields that we can weaken (173) to (141) and (174) to (143). Combining these results, it follows that the achievability of the superposition coding rate implies the achievability of any $R = R_1 + R_2$ such that (R_1, R_2) satisfy (140)–(141) and at least one of (143)–(144), as desired.

The process of weakening (173) to (140) is similar to that of weakening (174) to (143), so we focus on the former. Setting $U^{\mathrm{sc}} = X_1^{\mathrm{mac}}$, the optimization in (173) can be written as

$$\min_{\widetilde{P}_{Y|X_1 X}} \sum_{x_1, x, y} Q_{X_1 X}(x_1, x) \widetilde{P}_{Y|X_1 X}(y|x_1, x) \log \frac{\widetilde{P}_{Y|X_1 X}(y|x_1, x)}{\sum_x Q_{X|X_1}(x|x_1) W(y|x)} \tag{177}$$

subject to the constraints

$$\sum_x Q_{X_1 X}(x_1, x) \widetilde{P}_{Y|X_1 X}(y|x_1, x) = \sum_x Q_{X_1 X}(x_1, x) W(y|x) \tag{178}$$

$$\sum_{x_1, x, y} Q_{X_1 X}(x_1, x) \widetilde{P}_{Y|X_1 X}(y|x_1, x) \log q(x, y) \geq \sum_{x, y} Q_X(x) W(y|x) \log q(x, y). \tag{179}$$

Since $Q_{X_1 X}$ can be chosen arbitrarily, we set

$$Q_{X_1 X}(x_1, x) = \sum_{x_2} Q_1(x_1) Q_2(x_2) \mathbb{1}\{x = \psi(x_1, x_2)\} \tag{180}$$

for some Q_1, Q_2 and $\psi(\cdot, \cdot)$. Under this choice, the objective in (177) is given by

$$\sum_{x_1, x_2, y} Q_1(x_1) Q_2(x_2) \widetilde{P}_{Y|X_1 X}(y|x_1, \psi(x_1, x_2)) \log \frac{\widetilde{P}_{Y|X_1 X}(y|x_1, \psi(x_1, x_2))}{\sum_{x_2} Q_2(x_2) W(y|\psi(x_1, x_2))}, \tag{181}$$

and the constraints in (178)–(179) are written similarly with $\psi(x_1, x_2)$ playing the role of x. Since the objective and constraints depend on $\widetilde{P}_{Y|X_1 X}(y|x_1, x)$ only through $\widetilde{P}_{Y|X_1 X}(y|x_1, \psi(x_1, x_2))$, we can lower bound the minimization

by instead taking it over all conditional distributions $\widetilde{P}_{Y|X_1 X_2}$. Thus, we recover the minimization

$$\min_{\widetilde{P}_{Y|X_1 X_2}} \sum_{x_1, x_2, y} Q_1(x_1) Q_2(x_2) \widetilde{P}_{Y|X_1 X_2}(y|x_1, x_2) \log \frac{\widetilde{P}_{Y|X_1 X_2}(y|x_1, x_2)}{\sum_{x_2} Q_2(x_2) W(y|\psi(x_1, x_2))} \tag{182}$$

subject to the constraints

$$\sum_{x_2} Q_1(x_1) Q_2(x_2) \widetilde{P}_{Y|X_1 X_2}(y|x_1, x_2) = \sum_{x_2} Q_1(x_1) Q_2(x_2) W(y|\psi(x_1, x_2)) \tag{183}$$

$$\sum_{x_1, x_2, y} Q_1(x_1) Q_2(x_2) \widetilde{P}_{Y|X_1 X_2}(y|x_1, x_2) \log q(\psi(x_1, x_2), y)$$

$$\geq \sum_{x_1, x_2, y} Q_1(x_1) Q_2(x_2) W(y|\psi(x_1, x_2)) \log q(\psi(x_1, x_2), y). \tag{184}$$

Rewriting this minimization as a minimization over joint distributions $\widetilde{P}_{X_1 X_2 Y}$ subject to $\widetilde{P}_{X_1 X_2} = Q_1 \times Q_2$, we recover the right-hand side of (140). ∎

The result of Proposition 3 can be understood intuitively as follows. The condition in (173) corresponds to either (69) or (70) (depending on the choice of parameters), and the condition in (174) corresponds to either (143) or (144). The superposition coding rate only has two rate conditions (type-0 and type-1), whereas the expurgated MAC has three (type-1, type-2 and type-12), thus allowing the former to be weakened to the latter. The sum rate condition in (71) may appear to improve on that of (174) via the inclusion of two constraints of the form $I_{\widetilde{P}}(X_\nu; Y) \leq R_\nu$ instead of just one. However, from Lemma 9, at most one of these constraints is active for any given rate pair.

It may appear that the ability to choose Q_{UX} in (173)–(174) provides more freedom than the ability to choose Q_1 and Q_2 in (69)–(71). However, it is known that in the latter setting, any joint distribution of (X_1, X) (or (X_2, X)) can be induced by a suitable choice of $\psi(\cdot, \cdot)$ [28, p. 626]. This suggests (but does not prove) that the converse to Proposition 3 holds true, i.e. that the two rates are equivalent after the full optimization of the parameters.

Regardless of whether the two rates are equivalent after the full optimization of the random-coding parameters, the SC bound appears to be more amenable to local optimization techniques, since the optimization over $(\mathcal{X}_1, \mathcal{X}_2, Q_1, Q_2, \psi)$ is replaced by a seemingly simpler optimization over $(\mathcal{U}, Q_{UX})$. The optimization of the input distribution is a non-convex problem in general even for the single-user LM rate [7], and thus the ability to perform local optimizations is of interest.

D. Rates for General Alphabets

The following theorem shows that the expressions in Theorem 15 remain valid in the setting of general alphabets.

Theorem 16. *For an arbitrary memoryless channel W and auxiliary alphabet $\mathcal{U}$, the rate $R = R_0 + R_1$ is achievable for any (R_0, R_1) satisfying (175)–(176), where the supremum over $a(\cdot, \cdot)$ is taken over all functions such that the second moment of $a(U, X)$ is finite under $(U, X) \sim Q_{UX}$.*

The main step in proving Theorem 16 is proving the existence of a suitable cost-constrained ensemble parameterized by an input distribution Q_{UX} and L cost functions $\{a_l(u, x)\}$ $(l = 1, \cdots, L)$. We will present an ensemble of the form

$$P_{\boldsymbol{U}}(\boldsymbol{u}) = \frac{1}{\mu_{1,n}} \prod_{i=1}^{n} Q_U(u_i) \mathbb{1}\{\boldsymbol{u} \in \mathcal{D}_1\} \tag{185}$$

$$P_{\boldsymbol{X}|\boldsymbol{U}}(\boldsymbol{x}|\boldsymbol{u}) = \frac{1}{\mu_{2,n}(\boldsymbol{u})} \prod_{i=1}^{n} Q_{X|U}(x_i|u_i) \mathbb{1}\{\boldsymbol{x} \in \mathcal{D}_2(\boldsymbol{u})\} \tag{186}$$

with $\mathcal{D}_1$ and $\mathcal{D}_2(\boldsymbol{u})$ chosen such that (i) $\lim_{n\to\infty} \frac{1}{n} \log \mu_{1,n} = 0$, (ii) $\lim_{n\to\infty} \frac{1}{n} \log \mu_{2,n}(\boldsymbol{u}) = 0$ for all $\boldsymbol{u} \in \mathcal{D}_1$, and

(iii) there exists constants $\{\phi_l\}$ and δ such that for each $l = 1, \cdots, L$, we have $\left|\frac{1}{n}\sum_{i=1}^n a_l(u_i, x_i) - \phi_l\right| \leq \frac{\delta}{n}$. Once this ensemble is established, the proof of Theorem 16 will follow using similar steps to Section III-C, starting with the RCU bounds given in Theorem 12. These remaining details are omitted to avoid repetition.

The ensemble presented here is a straightforward generalization of that used in [39] for a Gaussian setting, but is described in detail for completeness. We fix $\delta > 0$ and a sequence g_n, and choose

$$\mathcal{D}_1 \triangleq \left\{\boldsymbol{u} \,:\, \mathbb{P}\big[\boldsymbol{X}'(\boldsymbol{u}) \in \mathcal{D}_2(\boldsymbol{u})\big] \geq \frac{1}{g_n}\right\} \tag{187}$$

$$\mathcal{D}_2(\boldsymbol{u}) \triangleq \left\{\boldsymbol{x} \,:\, \left|\frac{1}{n}\sum_{i=1}^n a_l(u_i, x_i) - \phi_l\right| \leq \frac{\delta}{n}, \, l = 1, \cdots, L\right\}, \tag{188}$$

where $\boldsymbol{X}'(\boldsymbol{u}) \sim Q_{X|U}^n(\boldsymbol{x}|\boldsymbol{u})$ and $\phi_l \triangleq \mathbb{E}_Q[a_l(U, X)]$. For any $\boldsymbol{u} \in \mathcal{D}_1$, we have from (187) that $\mu_{2,n}(\boldsymbol{u})$ decays subexponentially provided that g_n grows subexponentially. Hence, it only remains to show that there exists a subexponential sequence g_n such that $\mu_{1,n}$ decays subexponentially.

Defining

$$(\boldsymbol{U}', \boldsymbol{X}') \sim Q_U^n(\boldsymbol{u}')Q_{X|U}^n(\boldsymbol{x}'|\boldsymbol{u}'), \tag{189}$$

we have from Proposition 1 that

$$\mathbb{P}\big[\boldsymbol{X}' \in \mathcal{D}_2(\boldsymbol{U}')\big] = \frac{1}{f_n} \tag{190}$$

for some f_n growing at most polynomially in n. Since $\mu_{1,n} = \mathbb{P}[\boldsymbol{U}' \in \mathcal{D}_1]$ by definition, we can expand (190) as

$$\frac{1}{f_n} = \mathbb{P}\big[\boldsymbol{X}' \in \mathcal{D}_2(\boldsymbol{U}')\big|\boldsymbol{U}' \in \mathcal{D}_1\big]\mu_{1,n} + \mathbb{P}\big[\boldsymbol{X}' \in \mathcal{D}_2(\boldsymbol{U}')\big|\boldsymbol{U}' \notin \mathcal{D}_1\big](1 - \mu_{1,n}) \tag{191}$$

$$\leq \mu_{1,n} + \frac{1}{g_n}(1 - \mu_{1,n}) \tag{192}$$

$$\leq \mu_{1,n} + \frac{1}{g_n}, \tag{193}$$

where (192) follows by upper bounding the first probability in (191) by one and upper bounding the second probability using the definition of $\mathcal{D}_1$. Choosing $g_n = 2f_n$, we have that g_n grows polynomially, and $\mu_{1,n} \geq \frac{1}{2f_n}$ decays at most polynomially, as desired.

Since the conditions (175)–(176) each contain a supremum over $a(\cdot, \cdot)$, it suffices to choose $L = 2$ in (188) to prove Theorem 16.

V. REFINED SUPERPOSITION CODING

In this section, we study a refined version of the superposition coding ensemble which yields higher achievable rates for single-user mismatched decoding than the standard version studied in Section IV. While our analysis of the error probability will yield non-asymptotic bounds and error exponents as intermediate steps, we focus on the resulting achievable rates for the sake of clarity.

The ensemble is defined as follows. We fix a finite alphabet $\mathcal{U}$, an input distribution $Q_U \in \mathcal{P}(\mathcal{U})$ and the rates R_0 and $\{R_{1u}\}_{u \in \mathcal{U}}$. We write $M_0 \triangleq e^{nR_0}$ and $M_{1u} \triangleq e^{nR_{1u}}$. We let $P_{\boldsymbol{U}}(\boldsymbol{u})$ be the uniform distribution on the type class $T^n(Q_{U,n})$, where $Q_{U,n} \in \mathcal{P}_n(\mathcal{U})$ is the most probable type under Q_U. Thus, as in (160), we have

$$P_{\boldsymbol{U}}(\boldsymbol{u}) = \frac{1}{|T^n(Q_{U,n})|}\mathbb{1}\Big\{\boldsymbol{u} \in T^n(Q_{U,n})\Big\}. \tag{194}$$

As was done for standard superposition coding, we randomly generate the length-n auxiliary codewords $\{\boldsymbol{U}^{(i)}\}_{i=1}^{M_0}$ independently according to $P_{\boldsymbol{U}}$. The difference here is that the actual codewords are not generated conditionally independently given $\boldsymbol{U}^{(i)}$. Instead, we generate a number of partial codewords, and construct the length-n codeword by placing the entries of a partial codeword in the indices where $\boldsymbol{U}$ takes a particular value.
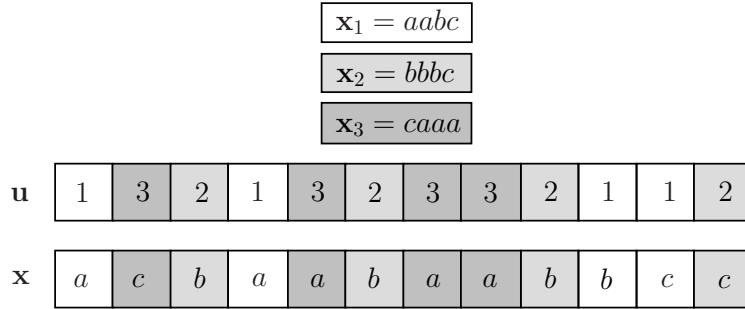
Figure 4. The construction of the final codeword from the auxiliary sequence $\boldsymbol{u}$ and the partial codeword $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ for refined superposition coding.

A more precise description is as follows. For each $u \in \mathcal{U}$, we define

$$n_u \triangleq Q_{U,n}(u)n \tag{195}$$

and fix a number of partial codeword distributions $P_{\boldsymbol{X}_u} \in \mathcal{P}(\mathcal{X}^{n_u})$. For example, $P_{\boldsymbol{X}_u}$ may be i.i.d., constant-composition or cost-constrained. For each $i = 1, \ldots, M_0$ and $u \in \mathcal{U}$, we further generate the length-n_u partial codewords $\{\boldsymbol{X}_u^{(i,j_u)}\}_{j_u=1}^{M_{1u}}$ independently according to $P_{\boldsymbol{X}_u}$. For example, when $\mathcal{U} = \{1, 2\}$ we have

$$\left\{ \left(\boldsymbol{U}^{(i)}, \{\boldsymbol{X}_1^{(i,j_1)}\}_{j_1=1}^{M_{11}}, \{\boldsymbol{X}_2^{(i,j_2)}\}_{j_2=1}^{M_{12}} \right) \right\}_{i=1}^{M_0} \sim \prod_{i=1}^{M_0} \left(P_{\boldsymbol{U}}(\boldsymbol{u}^{(i)}) \prod_{j_1=1}^{M_{11}} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1^{(i,j_1)}) \prod_{j_2=1}^{M_{12}} P_{\boldsymbol{X}_2}(\boldsymbol{x}_2^{(i,j_2)}) \right). \tag{196}$$

The message m at the encoder is indexed as $(m_0, m_{11}, \ldots, m_{1|\mathcal{U}|})$. To transmit a given message, we treat $\boldsymbol{U}^{(m_0)}$ as a time-sharing sequence; at the indices where $\boldsymbol{U}^{(m_0)}$ equals u, we transmit the symbols of $\boldsymbol{X}_u^{(m_0,m_{1u})}$. There are $M_0 \prod_u M_{1u}$ codewords, and hence the rate is $R = R_0 + \sum_u Q_{U,n}(u)R_{1u}$. An example of the construction of the codeword $\boldsymbol{x}$ from an auxiliary sequences $\boldsymbol{u}$ and partial codewords $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ is shown in Figure 4, where we have $\mathcal{U} = \{1, 2, 3\}$ and $\mathcal{X} = \{a, b, c\}$.

For clarity of exposition, we will present the analysis in the case that $\mathcal{U} = \{1, 2\}$. The same arguments apply to the general case. We let $\Xi(\boldsymbol{u}, \boldsymbol{x}_1, \boldsymbol{x}_2)$ denote the function for constructing the length-n codeword from the auxiliary sequence and partial codewords, and write

$$\boldsymbol{X}^{(i,j_1,j_2)} \triangleq \Xi(\boldsymbol{U}^{(i)}, \boldsymbol{X}_1^{(i,j_1)}, \boldsymbol{X}_2^{(i,j_2)}). \tag{197}$$

We let $\boldsymbol{y}_u(\boldsymbol{u})$ denote the subsequence of $\boldsymbol{y}$ corresponding to the indices where $\boldsymbol{u}$ equals u.

Upon receiving $\boldsymbol{y}$, the decoder forms the estimate

$$(\hat{m}_0, \hat{m}_1, \hat{m}_2) = \arg\max_{(i,j_1,j_2)} q^n(\boldsymbol{x}^{(i,j_1,j_2)}, \boldsymbol{y}) \tag{198}$$

$$= \arg\max_{(i,j_1,j_2)} q^{n_1}\left(\boldsymbol{x}_1^{(i,j_1)}, \boldsymbol{y}_1(\boldsymbol{u}^{(i)})\right) q^{n_2}\left(\boldsymbol{x}_2^{(i,j_2)}, \boldsymbol{y}_2(\boldsymbol{u}^{(i)})\right), \tag{199}$$

where the objective in (199) follows by separating the indices where $u = 1$ from those where $u = 2$. By writing the objective in this form, it is easily seen that for any given i, the pair (j_1, j_2) with the highest metric is the one for which j_1 maximizes $q^{n_1}(\boldsymbol{x}_1^{(i,j_1)}, \boldsymbol{y}_1(\boldsymbol{u}^{(i)}))$ and j_2 maximizes $q^{n_2}(\boldsymbol{x}_2^{(i,j_2)}, \boldsymbol{y}_2(\boldsymbol{u}^{(i)}))$. Therefore, we can split the error event into three (not necessarily disjoint) events:

(Type 0) $\quad \hat{m}_0 \neq m_0$

(Type 1) $\quad \hat{m}_0 = m_0$ and $\hat{m}_{11} \neq m_{11}$

(Type 2) $\quad \hat{m}_0 = m_0$ and $\hat{m}_{12} \neq m_{12}$

We denote the corresponding random-coding error probabilities by $\overline{p}_{e,0}(n, M_0, M_{11}, M_{12})$, $\overline{p}_{e,1}(n, M_{11})$ and $\overline{p}_{e,2}(n, M_{12})$ respectively. The overall random-coding error probability $\overline{p}_e(n, M_0, M_{11}, M_{12})$ satisfies

$$\max\{\overline{p}_{e,0}, \overline{p}_{e,1}, \overline{p}_{e,2}\} \leq \overline{p}_e \leq \overline{p}_{e,0} + \overline{p}_{e,1} + \overline{p}_{e,2}. \tag{200}$$

We assume without loss of generality that $(m_0, m_1, m_2) = (1, 1, 1)$. We let $\boldsymbol{U}$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ and $\boldsymbol{X}$ be the codewords corresponding to $(1, 1, 1)$, yielding $\boldsymbol{X} = \Xi(\boldsymbol{U}, \boldsymbol{X}_1, \boldsymbol{X}_2)$. We let $\overline{\boldsymbol{U}}$, $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$ be the codewords corresponding to an arbitrary message with $m_0 \neq 1$. For the index i corresponding to $\overline{\boldsymbol{U}}$, we write $\overline{\boldsymbol{X}}_1^{(j_1)}$, $\overline{\boldsymbol{X}}_2^{(j_2)}$ and $\overline{\boldsymbol{X}}^{(j_1, j_2)}$ in place of $\boldsymbol{X}_1^{(i, j_1)}$, $\boldsymbol{X}_2^{(i, j_2)}$ and $\boldsymbol{X}^{(i, j_1, j_2)}$ respectively. It follows that $\overline{\boldsymbol{X}}^{(j_1, j_2)} = \Xi(\overline{\boldsymbol{U}}, \overline{\boldsymbol{X}}_1^{(j_1)}, \overline{\boldsymbol{X}}_2^{(j_2)})$.

A. Rates for DMCs

In this section, we assume that the channel is a DMC, and consider the refined SC ensemble with each $P_{\boldsymbol{X}_u}$ chosen to be uniform over a conditional type class. We fix a joint distribution Q_{UX} and for each $u \in \mathcal{U}$ we let $Q_{u, n_u} \in \mathcal{P}_{n_u}(\mathcal{X})$ be the most probable type under $Q_{X|U=u}$ for sequences of length n_u. We let $P_{\boldsymbol{X}_u}$ be the uniform distribution on the type class $T^{n_u}(Q_{u, n_u})$, yielding

$$P_{\boldsymbol{X}_u}(\boldsymbol{x}_u) = \frac{1}{|T^{n_u}(Q_{u, n_u})|} \mathbb{1}\left\{\boldsymbol{x}_u \in T^{n_u}(Q_{u, n_u})\right\}. \tag{201}$$

The main result of this section is stated in the following theorem, which makes use of the LM rate defined in (6) and the set $\mathcal{T}_0^{\mathrm{cc}}$ defined in (163).

Theorem 17. *For any finite set $\mathcal{U}$ and input distribution Q_{UX}, the rate*

$$R = R_0 + \sum_u Q_U(u) R_{1u} \tag{202}$$

is achievable provided that R_0 and $\{R_{1u}\}_{u=1}^{|\mathcal{U}|}$ satisfy

$$R_{1u} \leq I_{\mathrm{LM}}(Q_{X|U=u}) \tag{203}$$

$$R_0 \leq \min_{\widetilde{P}_{UXY} \in \mathcal{T}_0(Q_{UX} \times W)} I_{\widetilde{P}}(U; Y) + \left[\max_{\mathcal{K} \subseteq \mathcal{U}, \mathcal{K} \neq \emptyset} \sum_{u \in \mathcal{K}} Q_U(u)\left(I_{\widetilde{P}}(X; Y|U = u) - R_{1u}\right)\right]^+. \tag{204}$$

Proof: From the construction of the random-coding ensemble, the type-1 error probability $\overline{p}_{e,1}$ is precisely that of the single-user constant-composition ensemble with rate R_{11}, length $n_1 = nQ_U(1)$, and input distribution $Q_{X|U=1}$. A similar statement holds for the type-2 error probability $\overline{p}_{e,2}$, and the analysis for these error events is identical to the derivation of the LM rate [1], [2], yielding (203). For the remainder of the proof, we analyze the type-0 event.

The error probability for the type-0 event satisfies

$$\overline{p}_{e,0} = c_0 \mathbb{P}\left[\bigcup_{i \neq 1} \bigcup_{j_1, j_2} \left\{\frac{q^n(\boldsymbol{X}^{(i, j_1, j_2)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1\right\}\right], \tag{205}$$

where $(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}) \sim W^n(\cdot|\boldsymbol{x})$ and $c_0 \in [\frac{1}{2}, 1]$, since breaking ties randomly increases the error probability by at most a factor of two [26]. Writing the probability as an expectation given $(\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y})$ and applying the truncated union bound, we obtain

$$\overline{p}_{e,0} = c_0' \mathbb{E}\left[\min\left\{1, (M_0 - 1)\mathbb{E}\left[\mathbb{P}\left[\bigcup_{j_1, j_2} \left\{\frac{q^n(\overline{\boldsymbol{X}}^{(j_1, j_2)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1\right\}\Big|\overline{\boldsymbol{U}}\right]\Big|\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y}\right]\right\}\right], \tag{206}$$

where $c_0' \in [\frac{1}{4}, 1]$, since for independent events the truncated union bound is tight to within a factor of $\frac{1}{2}$ [23, Lemma A.2]. We have written the probability of the union over j_1 and j_2 as an expectation given $\overline{\boldsymbol{U}}$.

Let the joint types of $(\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y})$ and $(\overline{\boldsymbol{U}}, \overline{\boldsymbol{X}}^{(j_1, j_2)}, \boldsymbol{Y})$ be denoted by P_{UXY} and $\widetilde{P}_{UXY}$ respectively. We claim that

$$\frac{q^n(\overline{\boldsymbol{X}}^{(j_1, j_2)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1 \tag{207}$$

if and only if

$$\widetilde{P}_{UXY} \in \mathcal{T}_{0,n}^{cc}(P_{UXY}) \triangleq \mathcal{T}_0^{cc}(P_{UXY}) \cap \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}), \tag{208}$$

where $\mathcal{T}_0^{cc}$ is defined in (163). The constraint $\widetilde{P}_{UX} = P_{UX}$ follows from the construction of the random coding ensemble, $\widetilde{P}_Y = P_Y$ follows since $(\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y})$ and $(\overline{\boldsymbol{U}}, \overline{\boldsymbol{X}}^{(j_1, j_2)}, \boldsymbol{Y})$ share the same $\boldsymbol{Y}$ sequence, and $\mathbb{E}_{\widetilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]$ coincides with the condition in (207). Thus, expanding (206) in terms of types yields

$$\overline{p}_{e,0} = c_0' \sum_{P_{UXY}} \mathbb{P}\Big[(\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y}) \in T^n(P_{UXY}) \Big]$$

$$\times \min \left\{ 1, (M_0 - 1) \sum_{\widetilde{P}_{UXY} \in \mathcal{T}_{0,n}^{cc}(P_{UXY})} \mathbb{P}\Big[(\overline{\boldsymbol{U}}, \boldsymbol{y}) \in T^n(\widetilde{P}_{UY}) \Big] \mathbb{P}\Big[\bigcup_{j_1, j_2} \Big\{ (\overline{\boldsymbol{u}}, \overline{\boldsymbol{X}}^{(j_1, j_2)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{UXY}) \Big\} \Big] \right\}, \tag{209}$$

where we write $(\overline{\boldsymbol{u}}, \boldsymbol{y})$ to denote an arbitrary pair such that $\boldsymbol{y} \in T^n(P_Y)$ and $(\overline{\boldsymbol{u}}, \boldsymbol{y}) \in T^n(\widetilde{P}_{UY})$. The dependence of these sequences on P_Y and $\widetilde{P}_{UY}$ is kept implicit for notational convenience.

Using a similar argument to the discussion following (199), we observe that $(\overline{\boldsymbol{u}}, \overline{\boldsymbol{X}}^{(j_1, j_2)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{UXY})$ if and only if $(\overline{\boldsymbol{X}}_u^{(j_u)}, \boldsymbol{y}_u(\overline{\boldsymbol{u}})) \in T^{n_u}(\widetilde{P}_{XY|U=u})$ for $u = 1, 2$. Thus, applying Lemma 3 with $Z_1(j_1) = \boldsymbol{X}_1^{(j_1)}$, $Z_2(j_2) = \boldsymbol{X}_2^{(j_2)}$, $\mathcal{A} = T_{\overline{\boldsymbol{u}}\boldsymbol{y}}^n(\widetilde{P}_{UXY})$, $\mathcal{A}_1 = T_{\boldsymbol{y}_1(\overline{\boldsymbol{u}})}^{n_1}(\widetilde{P}_{XY|U=1})$ and $\mathcal{A}_2 = T_{\boldsymbol{y}_2(\overline{\boldsymbol{u}})}^{n_2}(\widetilde{P}_{XY|U=2})$, we obtain

$$\mathbb{P}\Big[\bigcup_{j_1, j_2} \Big\{ (\overline{\boldsymbol{u}}, \overline{\boldsymbol{X}}^{(j_1, j_2)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{UXY}) \Big\} \Big] = c_0'' \min \Big\{ 1,$$

$$\min_{u=1,2} M_{1u} \mathbb{P}\Big[(\overline{\boldsymbol{X}}_u, \boldsymbol{y}_u(\overline{\boldsymbol{u}})) \in T^{n_u}(\widetilde{P}_{XY|U=u}) \Big], M_{11} M_{12} \mathbb{P}\Big[\bigcap_{u=1,2} \Big\{ (\overline{\boldsymbol{X}}_u, \boldsymbol{y}_u(\overline{\boldsymbol{u}})) \in T^{n_u}(\widetilde{P}_{XY|U=u}) \Big\} \Big] \Big\}, \tag{210}$$

where $c_0'' \in [\frac{1}{4}, 1]$. Recall from the proof of Lemma 3 that the four terms in the minimization correspond to the four subsets of $\{1, 2\}$.

Substituting (210) into (209), applying the properties of types in (A.9) and (A.10), and using the fact that the number of joint types is polynomial in n, we obtain

$$\lim_{n \to \infty} -\frac{1}{n} \log \overline{p}_{e,0} = \min_{P_{UXY} : P_{UX} = Q_{UX}} \min_{\widetilde{P}_{UXY} \in \mathcal{T}_0^{cc}(P_{UXY})} D(P_{UXY} \| Q_{UX} \times W)$$

$$+ \Big[I_{\widetilde{P}}(U; Y) + \Big[\max_{\mathcal{K} \subseteq \mathcal{U}, \mathcal{K} \neq \emptyset} \sum_{u \in \mathcal{K}} Q_U(u) \Big(I_{\widetilde{P}}(X; Y | U = u) - R_{1u} \Big) \Big]^+ - R_0 \Big]^+. \tag{211}$$

Similarly to the proof of Theorem 2, we have replaced the minimizations over types by minimizations over all distributions. Taking $P_{UXY} \to Q_{UX} \times W$, we obtain that the right-hand side of (211) is positive whenever (204) holds with strict inequality, thus completing the proof. ∎

The proof of Theorem 17 gives an exponentially tight analysis, yielding the exponent in (211). However, this does not prove that the resulting rate is ensemble-tight, since a subexponential decay of the error probability to zero is possible in principle. We can prove the ensemble tightness of the rate using (209)–(210), and dealing with the summation over P_{UXY} more carefully by following [6]. Instead of replacing the summation in (209) by a maximum, we restrict the summation to types which are ϵ-close to $Q_{UX} \times W$ in terms of ℓ_1 norm. The probability of $(\boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y})$ having such a joint type approaches one as $n \to \infty$ for any $\epsilon > 0$. Since c_0' and c_0'' in (206) and (210) are greater than or equal to $\frac{1}{4}$, the resulting error probability is lower bounded by $\frac{1}{16}(1 + o(1))$ for rates exceeding $R_0^* + \Delta(\epsilon)$, where R_0^* denotes the right-hand side of (204), and $\Delta(\epsilon) \to 0$ as $\epsilon \to 0$. The desired result follows by taking $\epsilon \to 0$.

B. Comparison to Standard Superposition Coding

In this subsection, we show that the conditions in (203)–(204) can be weakened to (173)–(174) upon identifying

$$R_1 = \sum_u Q_U(u) R_{1u}. \tag{212}$$

We show this to be the case for *any* auxiliary alphabet and input distribution (in contrast with Proposition 3, where we showed an analogous result to hold after the optimization of the parameters).

Proposition 4. *For any auxiliary alphabet $\mathcal{U}$ and input distribution Q_{UX}, the rate $\max_{R_0, R_{11}, \cdots, R_{1|\mathcal{U}|}} R_0 + \sum_u Q_U(u) R_{1u}$ given in Theorem 17 is at least as high as the rate $\max_{R_0, R_1} R_0 + R_1$ given in Theorem 14.*

Proof: We begin by weakening (204) to (174). We lower bound the right-hand side of (204) by replacing the maximum over $\mathcal{K}$ by the particular choice $\mathcal{K} = \mathcal{U}$, yielding the weakened condition

$$R_0 \le \min_{\widetilde{P}_{UXY} \in \mathcal{T}_0^{\mathrm{cc}}(Q_{UX} \times W)} I_{\widetilde{P}}(U; Y) + \left[I_{\widetilde{P}}(X; Y | U) - R_1 \right]^+, \tag{213}$$

where we have used (212) and the definition of conditional mutual information. We can weaken (213) to (174) using the chain rule for mutual information, and noting that (213) is always satisfied when the minimizing $\widetilde{P}_{UXY}$ satisfies $I_{\widetilde{P}}(U; Y) > R_0$.

Next, we show that highest value of R_1 permitted by the $|\mathcal{U}|$ conditions in (203), denoted by R_1^*, can be lower bounded by the right-hand side of (173). From (212) and (203), we have

$$R_1^* = \sum_u Q_U(u) I_{\widetilde{P}^*}(X; Y | U = u), \tag{214}$$

where $\widetilde{P}_{XY|U=u}^*$ is the distribution which achieves the minimum in (6) under $Q_{X|U=u}$. Defining the joint distribution $\widetilde{P}_{UXY}^*$ accordingly with $\widetilde{P}_U^* = Q_U$, we can write (214) as

$$R_1^* = I_{\widetilde{P}^*}(X; Y | U). \tag{215}$$

Therefore, we can lower bound R_1^* by the right-hand side of (173) provided that $\widetilde{P}_{UXY}^* \in \mathcal{T}_1^{\mathrm{cc}}(Q_{UX} \times W)$. The constraints $\widetilde{P}_{UX}^* = Q_{UX}$ and $\widetilde{P}_{UY}^* = P_{UY}$ in (164) are satisfied since we have chosen $\widetilde{P}_{UXY}^* = Q_U$, and since the constraints in (6) imply $\widetilde{P}_{X|U=u}^* = Q_{X|U=u}$ and $\widetilde{P}_{Y|U=u}^* = P_{Y|U=u}$ for all $u \in \mathcal{U}$. The constraint $\mathbb{E}_{\widetilde{P}^*}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]$ is satisfied since, from (6), we have $\mathbb{E}_{\widetilde{P}^*}[\log q(X,Y) | U = u] \ge \mathbb{E}_P[\log q(X,Y) | U = u]$ for each $u \in \mathcal{U}$. This concludes the proof. ∎

Intuitively, one can think of the gain of the refined SC ensemble as being due to a stronger dependence among the codewords. For standard SC, the codewords $\{\boldsymbol{X}^{(i,j)}\}_{j=1}^{M_1}$ are conditionally independent given $\boldsymbol{U}^{(i)}$, whereas for refined SC this is generally not the case. It should be noted, however, that the exponents for standard SC may be better, particularly at low to moderate rates. In particular, we noted in the proof of Theorem 17 that the type-1 and type-2 error events are equivalent to a single-user channel, but the corresponding block lengths are only n_1 and n_2. Thus, if either $Q_U(1)$ or $Q_U(2)$ is close to zero, the corresponding exponent will be small. These observations are analogous to the comparison of explicit time-sharing and coded time-sharing for the MAC, given in Section III-E.

C. Numerical Examples

In this subsection, we provide a number of examples comparing the two versions of superposition coding and the LM rate. Numerical results are obtained using the optimization software YALMIP [29]. We do not explicitly give values for Lapidoth's rate [6] (see Theorem 3), since for each example given, we found it to coincide with the SC rate (Theorem 14). This was done by taking a locally optimized input distribution for SC, and using it to choose the parameters for

the MAC using similar ideas to those given in Section IV-C (e.g. see (180)). In general, however, we required larger auxiliary alphabet sizes for the MAC, and we found its direct local optimization to be more difficult than that of SC.

1) Example 1: Sum Channel: Here we provide an example which can be seen as an analog of Lapidoth's parallel channel example [6] (see Section I-C).

Given two channels (W_1, W_2) respectively defined on the alphabets $(\mathcal{X}_1, \mathcal{Y}_1)$ and $(\mathcal{X}_2, \mathcal{Y}_2)$, the sum channel is defined to be the channel $W(y|x)$ with $|\mathcal{X}| = |\mathcal{X}_1| + |\mathcal{X}_2|$ and $|\mathcal{Y}| = |\mathcal{Y}_1| + |\mathcal{Y}_2|$ such that one of the two subchannels is used on each transmission [40]. One can similarly combine two metrics $q_1(x_1, y_1)$ and $q_2(x_2, y_2)$ to form a sum metric $q(x, y)$. Assuming without loss of generality that $\mathcal{X}_1$ and $\mathcal{X}_2$ are disjoint and $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint, we have

$$q(x, y) = \begin{cases} q_1(x_1, y_1) & x_1 \in \mathcal{X}_1 \text{ and } y_1 \in \mathcal{Y}_1 \\ q_2(x_2, y_2) & x_2 \in \mathcal{X}_2 \text{ and } y_2 \in \mathcal{Y}_2 \\ 0 & \text{otherwise,} \end{cases} \tag{216}$$

and similarly for $W(y|x)$. Let $\hat{Q}_1$ and $\hat{Q}_2$ be the distributions that maximize the LM rate in (6) on the respective subchannels. We set $\mathcal{U} = \{1, 2\}$, $Q_{X|U=1} = (\hat{Q}_1, \mathbf{0})$ and $Q_{X|U=2} = (\mathbf{0}, \hat{Q}_2)$, where $\mathbf{0}$ denotes the zero vector. We leave Q_U to be specified.

Combining the constraints $\widetilde{P}_{UX} = Q_{UX}$ and $\mathbb{E}_{\widetilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]$ in (163), it is straightforward to show that the minimizing $\widetilde{P}_{UXY}(u, x, y)$ in (204) only takes on non-zero values for (u, x, y) such that (i) $u = 1$, $x \in \mathcal{X}_1$ and $y \in \mathcal{Y}_1$, or (ii) $u = 2$, $x \in \mathcal{X}_2$ and $y \in \mathcal{Y}_2$. It follows that U is a deterministic function of Y under the minimizing $\widetilde{P}_{UXY}$, and hence

$$I_{\widetilde{P}}(U; Y) = H(Q_U) - H_{\widetilde{P}}(U|Y) = H(Q_U). \tag{217}$$

Therefore, the right-hand side of (204) is lower bounded by $H(Q_U)$. Using (202) and performing a simple optimization of Q_U, it follows that the rate $\log\left(e^{I_1^{\mathrm{LM}}(\hat{Q}_1)} + e^{I_2^{\mathrm{LM}}(\hat{Q}_2)}\right)$ is achievable, where I_ν^{LM} is the LM rate for subchannel ν. An analogous result has been proved in the setting of matched decoding using the known formula for channel capacity [40]. It should be noted that the LM rate of (W, q) can be strictly less than $\log\left(e^{I_1^{\mathrm{LM}}(\hat{Q}_1)} + e^{I_2^{\mathrm{LM}}(\hat{Q}_2)}\right)$ even for simple examples (e.g. binary symmetric subchannels).

2) Example 2: Zero Undetected Error Capacity: It was shown by Csiszár and Narayan [3] that two special cases of the mismatched capacity are the zero-undetected erasures capacity [41] and the zero-error capacity [42]. Here we consider the zero-undetected erasures capacity, defined to be the highest achievable rate in the case that the decoder is required to know with certainty whether or not an error has occurred. For any DMC, the zero-undetected erasures capacity is equal to the mismatched capacity under the decoding metric $q(x, y) = \mathbb{1}\{W(y|x) > 0\}$ [3].

We consider an example from [43], where $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, and the channel and metric are described by the entries of the matrices

$$\boldsymbol{W} = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.75 & 0.25 \\ 0.25 & 0 & 0.75 \end{bmatrix} \tag{218}$$

$$\boldsymbol{q} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \tag{219}$$

where x indexes the rows and y indexes the columns.

Using an exhaustive search, we found the optimized LM rate to be $R_{\mathrm{LM}}^* = 0.599$ bits/use, using the input distribution $Q = (0.449, 0.551, 0)$.[2] Furthermore, it was stated in [43] that the rate obtained by considering the second-order product

---

[2]The rate of 0.6128 bits/use reported in [43] appears to be erroneous.

Table I
ACHIEVABLE RATES (BITS/USE) FOR THE MISMATCHED CHANNEL IN (20)-(21).

Input Distribution	Refined SC	Standard SC
$Q_{UX}^{(1)}$	1.313	1.060
$Q_{UX}^{(2)}$	1.236	1.236
$Q_{UX}^{(\mathrm{LM})}$	1.111	1.111

of the channel and metric (see Section I-C) is equal to $R_{\mathrm{LM2}}^* = 0.616$ bits/use. Using local optimization techniques, we verified that this rate is achieved with $Q = (0, 0.250, 0, 0.319, 0, 0, 0, 0.181, 0.250)$, where the order of the inputs is $(0,0), (0,1), (0,2), (1,0), \cdots, (2,2)$.

The global optimization of (173)–(174) over $\mathcal{U}$ and Q_{UX} is difficult. However, setting $|\mathcal{U}| = 2$ and applying local optimization techniques using a number of starting points, we obtained an achievable rate of $R_{\mathrm{sc}}^* = 0.695$ bits/use, with $Q_U = (0.645, 0.355)$, $Q_{X|U=1} = (0.3, 0.7, 0)$ and $Q_{X|U=2} = (0, 0, 1)$. Thus, SC not only yields an improvement over the single-letter LM rate, but also over the two-letter version. We obtained the same rate and input distribution as that of standard SC when applying local optimization techniques to refined SC.

3) Example 3: We consider the channel and decoding metric described by the entries of the matrices

$$
\boldsymbol{W} = \begin{bmatrix} 0.99 & 0.01 & 0 & 0 \\ 0.01 & 0.99 & 0 & 0 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix} \tag{220}
$$

$$
\boldsymbol{q} = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0.05 & 0.15 & 1 & 0.05 \\ 0.15 & 0.05 & 0.5 & 1 \end{bmatrix}. \tag{221}
$$

We have intentionally chosen a highly asymmetric channel and metric, since such examples often yield larger gaps between the various achievable rates. Using an exhaustive search, we found the optimized LM rate to be $R_{\mathrm{LM}}^* = 1.111$ bits/use, which is achieved by the input distribution $Q_X^* = (0.403, 0.418, 0, 0.179)$.

Setting $|\mathcal{U}| = 2$ and applying local optimization techniques using a number of starting points, we obtained an achievable rate of $R_{\mathrm{rsc}}^* = 1.313$ bits/use, with $Q_U = (0.698, 0.302)$, $Q_{X|U=1} = (0.5, 0.5, 0, 0)$ and $Q_{X|U=2} = (0, 0, 0.528, 0.472)$. We denote the corresponding input distribution by $Q_{UX}^{(1)}$.

Applying similar techniques to the standard superposition coding rate in (173)–(174), we obtained an achievable rate of $R_{\mathrm{sc}}^* = 1.236$ bits/use, with $Q_U = (0.830, 0.170)$, $Q_{X|U=1} = (0.435, 0.450, 0.115, 0)$ and $Q_{X|U=2} = (0, 0, 0, 1)$. We denote the corresponding input distribution by $Q_{UX}^{(2)}$.

The achievable rates for this example are summarized in Table I, where $Q_{UX}^{(\mathrm{LM})}$ denotes the distribution in which U is deterministic and the X-marginal maximizes the LM rate. While the achievable rate of Theorem 17 coincides with that of Theorem 14 under $Q_{UX}^{(2)}$, the former is significantly higher under $Q_{UX}^{(1)}$. Both types of superposition coding yield a strict improvement over the LM rate.

Our parameters may not be globally optimal, and thus we cannot conclude from this example that refined SC yields a strict improvement over standard SC (and hence over Lapidoth's rate [6]) after optimizing $\mathcal{U}$ and Q_{UX}. However, as discussed in Section IV-C, improvements for a fixed set of random-coding parameters are still of significant interest due to the fact that global optimizations are infeasible in general.

D. Dual Expressions and General Alphabets

In this subsection, we present a dual expression for the rate given in Theorem 17 in the case that $|\mathcal{U}| = 2$, as well as extending the result to general alphabets $\mathcal{X}$ and $\mathcal{Y}$. Generalizations to the case that $|\mathcal{U}| > 2$ are possible, but cumbersome. It should be noted that the random-coding ensemble relies on $|\mathcal{U}|$ being finite.

With $\mathcal{U} = \{1, 2\}$, the condition in (204) is given by

$$R_0 \leq \min_{\widetilde{P}_{UXY} \in \mathcal{T}_0^{cc}(Q_{UX} \times W)} I_{\widetilde{P}}(U;Y) + \Big[\max \Big\{ Q_U(1)\big(I_{\widetilde{P}}(X;Y|U=1) - R_{11}\big), $$

$$Q_U(2)\big(I_{\widetilde{P}}(X;Y|U=2) - R_{12}\big), I_{\widetilde{P}}(X;Y|U) - R_1 \Big\} \Big]^+, \quad (222)$$

where

$$R_1 \triangleq \sum_u Q_U(u) R_{1u}. \quad (223)$$

Similarly to Lemma 7 for the MAC, the following lemma gives an an alternative expression for the condition in (204) which is more amenable to Lagrange duality techniques.

Lemma 10. *For any input distribution Q_{UX} and rates (R_0, R_{11}, R_{12}), the condition in (222) holds if and only if at least one of the following conditions hold:*

$$R_0 \leq \min_{\widetilde{P}_{UXY} \in \mathcal{T}_0^{cc}(Q_{UX} \times W)} I_{\widetilde{P}}(U;Y) + \Big[\max \Big\{ Q_U(1)\big(I_{\widetilde{P}}(X;Y|U=1) - R_{11}\big), I_{\widetilde{P}}(X;Y|U) - R_1 \Big\} \Big]^+ \quad (224)$$

$$R_0 \leq \min_{\widetilde{P}_{UXY} \in \mathcal{T}_0^{cc}(Q_{UX} \times W)} I_{\widetilde{P}}(U;Y) + \Big[\max \Big\{ Q_U(2)\big(I_{\widetilde{P}}(X;Y|U=2) - R_{12}\big), I_{\widetilde{P}}(X;Y|U) - R_1 \Big\} \Big]^+. \quad (225)$$

Proof: This is a special case of Lemma 6 with the following identifications:

$$f(\boldsymbol{z}) = I_{\widetilde{P}}(U;Y) \quad (226)$$

$$g(\boldsymbol{z}) = I_{\widetilde{P}}(X;Y|U) - R_1 \quad (227)$$

$$g_1(\boldsymbol{z}) = Q_U(1)\big(I_{\widetilde{P}}(X;Y|U=1) - R_{11}\big) \quad (228)$$

$$g_2(\boldsymbol{z}) = Q_U(2)\big(I_{\widetilde{P}}(X;Y|U=2) - R_{12}\big). \quad (229)$$

In this case, the condition in (31) holds with equality. ∎

Since the right-hand side of (203) is the LM rate, we can make use of the dual expression in (7). The following theorem gives the dual expressions for (224)–(225).

Theorem 18. *The conditions in (224)–(225) can respectively be written as*

$$R_0 \leq \sup_{s \geq 0, \rho_1 \in [0,1], \rho_2 \in [0,1], a(\cdot,\cdot)} \mathbb{E}\left[\log \frac{\big(q(X,Y)^{s'(u)} e^{a(U,X)}\big)^{\rho'(U)}}{\mathbb{E}\Big[\big(\mathbb{E}\big[q(\overline{X},Y)^{s'(\overline{U})} e^{a(\overline{U},\overline{X})} \,\big|\, \overline{U}\big]\big)^{\rho'(\overline{U})} \,\Big|\, Y\Big]} \right] - \sum_{u=1,2} \rho'(u) Q_U(u) R_{1u} \quad (230)$$

$$R_0 \leq \sup_{s \geq 0, \rho_1 \in [0,1], \rho_2 \in [0,1], a(\cdot,\cdot)} \mathbb{E}\left[\log \frac{\big(q(X,Y)^{s''(u)} e^{a(U,X)}\big)^{\rho''(U)}}{\mathbb{E}\Big[\big(\mathbb{E}\big[q(\overline{X},Y)^{s''(\overline{U})} e^{a(\overline{U},\overline{X})} \,\big|\, \overline{U}\big]\big)^{\rho''(\overline{U})} \,\Big|\, Y\Big]} \right] - \sum_{u=1,2} \rho''(u) Q_U(u) R_{1u}, \quad (231)$$

where

$$\rho'(1) = \rho_1, \ \ \rho'(2) = \rho_1 \rho_2, \ \ s'(1) = \rho_2 s, \ \ s'(2) = s \quad (232)$$

$$\rho''(1) = \rho_1 \rho_2, \ \ \rho''(2) = \rho_2, \ \ s''(1) = s, \ \ s''(2) = \rho_1 s \quad (233)$$

and $(U, X, Y, \overline{U}, \overline{X}) \sim Q_{UX}(u,x) W(y|x) Q_{UX}(\overline{u}, \overline{x})$.

Proof: See Appendix D. ∎

It is interesting to observe that the conditions in (230)–(231) bear a strong resemblance to the standard SC condition in (176). In particular, the latter can be recovered by setting $\rho_2 = 1$ in (230) or $\rho_1 = 1$ in (231), thus providing an alternative proof of Proposition 3 for the case that $|\mathcal{U}| = 2$.

In order to extend Theorem 18 to general alphabets, we again use cost-constrained coding. We consider the ensemble given in (196), with $P_{\boldsymbol{X}_u}$ given by

$$P_{\boldsymbol{X}_u}(\boldsymbol{x}_u) = \frac{1}{\mu_{u,n_u}} \prod_{i=1}^{n_u} Q_{X|U}(x_{u,i}|u) \mathbb{1}\{\boldsymbol{x}_u \in \mathcal{D}_{u,n_u}\}, \tag{234}$$

where

$$\mathcal{D}_{u,n_u} \triangleq \left\{ \boldsymbol{x}_u \ : \ \left| \frac{1}{n_u} \sum_{i=1}^{n_u} a_{u,l}(x_{u,i}) - \phi_{u,l} \right| \leq \frac{\delta}{n_u}, \, l = 1, \ldots, L_u \right\} \tag{235}$$

$$\phi_{u,l} \triangleq \mathbb{E}_Q\left[a_{u,l}(X) \,|\, U = u \right], \tag{236}$$

and where μ_{u,n_u}, $\{a_{u,l}\}$ and δ are defined analogously to (98)–(99), and n_u is defined in (195).

Theorem 19. *For any mismatched memoryless single-user channel W and input distribution Q_{UX} ($\mathcal{U} = 1, 2$), the rate $R = R_0 + \sum_{u=1,2} Q_U(u) R_{1u}$ is achievable for any (R_0, R_{11}, R_{12}) satisfying (203) (with I_{LM} defined in (7)) and at least one of (230)–(231), where each supremum is subject to $\mathbb{E}_Q[a(U, X)^2] < \infty$. Furthermore, the rate is achievable using cost-constrained coding in (234) with $L_1 = L_2 = 2$.*

Proof: The direct derivation of (230)–(231) is given in Appendix D. The choice $L_1 = L_2 = 2$ suffices since for $u = 1, 2$, one cost is required for (203) and another for (230)–(231). It suffices to let the cost functions for (230)–(231) coincide, since the theorem only requires that one of the two hold. ∎

VI. CONCLUSION

We have provided finite-length bounds, error exponents and achievable rates for mismatched decoding using three multiuser coding techniques. The bounds on the probability of a multiply-indexed union given in Section II-A are general, and may be valuable for obtaining finite-length bounds and error exponents for other network information theory settings. Interestingly, their utility is not restricted to suboptimal decoding rules; see [44] for an application to the matched relay channel.

Throughout the paper, we considered several variations of constant-composition random coding and cost-constrained random coding. Furthermore, we have discussed the number of costs required for the cost-constrained ensemble to match the performance of the constant-composition ensemble. From Lemma 5, we see that each dual optimization variable (e.g. $a_1(\cdot)$) corresponds to a primal constraint on a marginal (e.g. $P_{X_1} = Q_1$). Since the dual variables also correspond to cost functions, we can use this observation to relate the number of costs with the number of marginal constraints. For example, consider the exponents for the mismatched MAC (Theorems 2 and 4) and those for ML decoding (Theorems 8 and 9). In the primal domain, ML decoding yields two fewer constraints on the X_ν marginals for each $\nu = 1, 2$, due to the absence of the sets $\mathcal{T}_\nu^{\mathrm{cc}}$ for $\nu = 1, 2, 12$. Analogously, in the dual domain, the number of cost functions per user is reduced by two.

APPENDIX A
THE METHOD OF TYPES

In this section we give an overview of the method of types, and present the properties that are used throughout the paper. Wherever possible, we use standard notation (e.g. see [10]), but for completeness we include all relevant

definitions in this section. We do not provide proofs, since we only present well-known properties of simple variations thereof (e.g. see [10, Ch. 2], [11], [45]).

Let $\mathcal{X}$ be a finite alphabet. The type of a sequence $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ is its empirical distribution, defined by

$$\hat{P}_{\boldsymbol{x}}(x) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = x\}. \tag{A.1}$$

The set of all types on $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$. For a given $P_X \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(P_X)$ is the set of all sequences in $\mathcal{X}^n$ with type P_X,

$$T^n(P_X) \triangleq \left\{ \boldsymbol{x} \in \mathcal{X}^n \ : \ \hat{P}_{\boldsymbol{x}} = P_X \right\}. \tag{A.2}$$

For two finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, $\hat{P}_{\boldsymbol{xy}}(x,y)$, $\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and $T^n(P_{XY})$ are defined analogously. In this case, the type is also referred to as a joint type. The conditional type of $\boldsymbol{y}$ given $\boldsymbol{x}$ is defined by

$$\hat{P}_{\boldsymbol{y}|\boldsymbol{x}}(y|x) \triangleq \frac{\hat{P}_{\boldsymbol{xy}}(x,y)}{\hat{P}_{\boldsymbol{x}}(x)} \tag{A.3}$$

if $\hat{P}_{\boldsymbol{x}}(x) > 0$, and $\hat{P}_{\boldsymbol{y}|\boldsymbol{x}}(y|x) = \frac{1}{|\mathcal{Y}|}$ otherwise. The set of all conditional types on $\mathcal{Y}^n$ given $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{Y}|\mathcal{X})$. For any $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and sequence $\boldsymbol{x}$, the conditional type class $T_{\boldsymbol{x}}^n(P_{XY})$ is the set of all sequences $\boldsymbol{y}$ in $\mathcal{Y}^n$ such that $(\boldsymbol{x}, \boldsymbol{y})$ has joint type P_{XY},

$$T_{\boldsymbol{x}}^n(P_{XY}) = \left\{ \boldsymbol{y} \in \mathcal{Y}^n \ : \ \hat{P}_{\boldsymbol{xy}} = P_{XY} \right\}. \tag{A.4}$$

The key property of types which permits an exponentially tight analysis of certain quantities is that the number of (possibly joint or conditional) types is polynomial in n:

$$|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|-1} \tag{A.5}$$

$$|\mathcal{P}_n(\mathcal{Y}|\mathcal{X})| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|-1}. \tag{A.6}$$

For concreteness, we present the subsequent properties of types in the context of the multiple-access channel. However, the properties are also used in other settings throughout this paper. We fix the conditional input distributions $Q_1 \in \mathcal{P}(\mathcal{X}_1)$ and $Q_2 \in \mathcal{P}(\mathcal{X}_2)$, and let $Q_{1,n} \in \mathcal{P}_n(\mathcal{X}_1)$ and $Q_{2,n}(x) \in \mathcal{P}_n(\mathcal{X}_2)$ be the most probable types under Q_1 and Q_2 respectively. We define $\boldsymbol{X}_1$ and $\overline{\boldsymbol{X}}_1$ to be independent random variables drawn from a uniform distribution on the type class $T^n(Q_{1,n})$, and similarly for $\boldsymbol{X}_2$ and $\overline{\boldsymbol{X}}_2$ We define $\boldsymbol{Y}$ to be generated from $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ according to the conditional product distribution on $W(y|x_1, x_2)$.

We have the following:

- Letting $Q_1^n(\boldsymbol{x}_1) \triangleq \prod_{i=1}^{n} Q_1(x_{1,i})$, we have

$$\mathbb{P}[\boldsymbol{X}_1 = \boldsymbol{x}_1] \ \dot{\leq} \ Q_1^n(\boldsymbol{x}_1). \tag{A.7}$$

- For any $P_{X_1 X_2}$ such that $P_{X_1} = Q_{1,n}$ and $P_{X_2} = Q_{2,n}$,

$$\mathbb{P}\left[(\boldsymbol{X}_1, \boldsymbol{X}_2) \in T^n(P_{X_1 X_2}) \right] \ \dot{=} \ \exp\left(-n I_P(X_1; X_2) \right). \tag{A.8}$$

- For any $P_{X_1 X_2 Y}$ such that $P_{X_1} = Q_{1,n}$ and $P_{X_2} = Q_{2,n}$,

$$\mathbb{P}\left[(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \in T^n(P_{X_1 X_2 Y}) \right] \ \dot{=} \ \exp\left(-n D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \right). \tag{A.9}$$

- For $\nu = 1, 2$, if $\boldsymbol{y} \in T^n(P_Y)$, then

$$\mathbb{P}\left[(\overline{\boldsymbol{X}}_\nu, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_\nu Y}) \right] \ \dot{=} \ \exp\left(-n I_{\widetilde{P}}(X_\nu; Y) \right) \tag{A.10}$$

for any $\widetilde{P}_{X_\nu Y}$ such that $\widetilde{P}_{X_\nu} = Q_{\nu,n}$ and $\widetilde{P}_Y = P_Y$.

- If $\boldsymbol{y} \in T^n(P_Y)$, then

$$\mathbb{P}\big[(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1 X_2 Y})\big] \doteq \exp\big(-nD(\widetilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \widetilde{P}_Y)\big) \tag{A.11}$$

for any $\widetilde{P}_{X_1 X_2 Y}$ such that $\widetilde{P}_{X_1} = Q_{1,n}$, $\widetilde{P}_{X_2} = Q_{2,n}$ and $\widetilde{P}_Y = P_Y$.

APPENDIX B
CONVEX OPTIMIZATION PROOFS

A. Proof of Lemma 5

The Lagrangian [12, Sec. 5.1.1] of the optimization problem in (29) is given by

$$L = \sum_{z_1, z_2} \widetilde{P}_{Z_1 Z_2}(z_1, z_2) \bigg(\log \frac{\widetilde{P}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2)} - \log g(z_1, z_2) - \lambda \log f(z_1, z_2) \bigg)$$
$$+ \sum_{z_1} \mu_1(z_1)\big(P_{Z_1}(z_1) - \widetilde{P}_{Z_1}(z_1)\big) + \sum_{z_2} \mu_2(z_2)\big(P_{Z_2}(z_2) - \widetilde{P}_{Z_2}(z_2)\big) + \lambda\beta, \tag{B.1}$$

where $\lambda \geq 0$, $\mu_1(\cdot)$ and $\mu_2(\cdot)$ are Lagrange multipliers. Since the objective in (29) is convex and the constraints are affine, the optimal value is equal to L for some choice of $\widetilde{P}_{Z_1 Z_2}$ and the Lagrange multipliers satisfying the Karush-Kuhn-Tucker (KKT) conditions [12, Sec. 5.5.3].

We proceed to simplify (B.1) using the KKT conditions. Setting $\frac{\partial L}{\partial \widetilde{P}(z_1,z_2)} = 0$ yields

$$1 + \log \frac{\widetilde{P}_{Z_1 Z_2}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2) f(z_1, z_2)^\lambda g(z_1, z_2)} - \mu_1(z_1) - \mu_2(z_2) = 0, \tag{B.2}$$

and hence

$$\widetilde{P}_{Z_1 Z_2}(z_1, z_2) = P_{Z_1}(z_1) P_{Z_2}(z_2) f(z_1, z_2)^\lambda g(z_1, z_2) e^{-1 + \mu_1(z_1) + \mu_2(z_2)}. \tag{B.3}$$

Applying the constraint $\widetilde{P}_{Z_2} = P_{Z_2}$ yields

$$\sum_{\overline{z}_1} P_{Z_1}(\overline{z}_1) f(\overline{z}_1, z_2)^\lambda g(\overline{z}_1, z_2) e^{\mu_1(\overline{z}_1)} = e^{1 - \mu_2(z_2)}, \tag{B.4}$$

and hence

$$\mu_2(z_2) = 1 - \log \sum_{\overline{z}_1} P_{Z_1}(\overline{z}_1) f(\overline{z}_1, z_2)^\lambda g(\overline{z}_1, z_2) e^{\mu_1(\overline{z}_1)}. \tag{B.5}$$

Substituting (B.2) into (B.1) yields

$$L = -1 + \sum_{z_1} \mu_1(z_1) P_{Z_1}(z_1) + \sum_{z_2} \mu_2(z_2) P_{Z_2}(z_2) + \lambda\beta, \tag{B.6}$$

and applying (B.5) yields (30) with the supremum omitted. It follows that (30) is an upper bound to (29).

To obtain a matching lower bound, we make use of the log-sum inequality [46, Thm. 2.7.1] similarly to [4, Appendix A]. For any $\widetilde{P}_{Z_1 Z_2}$ satisfying the constraints in (29), we can lower bound the objective in (29) as follows:

$$\sum_{z_1, z_2} \widetilde{P}_{Z_1 Z_2}(z_1, z_2) \log \frac{\widetilde{P}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2) g(z_1, z_2)} \tag{B.7}$$

$$\geq \sum_{z_1, z_2} \widetilde{P}_{Z_1 Z_2}(z_1, z_2) \log \frac{\widetilde{P}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2) f(z_1, z_2)^\lambda g(z_1, z_2)} + \lambda\beta \tag{B.8}$$

$$= \sum_{z_1, z_2} \widetilde{P}_{Z_1 Z_2}(z_1, z_2) \log \frac{\widetilde{P}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2) f(z_1, z_2)^\lambda g(z_1, z_2) e^{\mu_1(\overline{z}_1)}} + \sum_{z_1} P_{Z_1}(z_1) \mu_1(z_1) + \lambda\beta, \tag{B.9}$$

where (B.8) holds for any $\lambda \geq 0$ due to the constraint $\mathbb{E}_{\tilde{P}}[\log f(Z_1, Z_2)] \geq \beta$, and (B.9) holds for any $\mu_1(\cdot)$ by a simple expansion of the logarithm. Applying the log-sum inequality, we can lower bound (B.9) by the objective in (30). Since $\lambda \geq 0$ and $\mu_1(\cdot)$ are arbitrary, the proof is complete.

B. Proof of Lemma 6

We define the functions

$$\Phi_0(\boldsymbol{z}) \triangleq f(\boldsymbol{z}) + \big[g(\boldsymbol{z})\big]^+ \tag{B.10}$$

$$\Phi_1(\boldsymbol{z}) \triangleq f(\boldsymbol{z}) + \Big[\max\big\{g_1(\boldsymbol{z}), g(\boldsymbol{z})\big\}\Big]^+ \tag{B.11}$$

$$\Phi_2(\boldsymbol{z}) \triangleq f(\boldsymbol{z}) + \Big[\max\big\{g_2(\boldsymbol{z}), g(\boldsymbol{z})\big\}\Big]^+. \tag{B.12}$$

Since $f(\cdot)$, $g(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ are convex by assumption, it follows that $\Phi_0(\cdot)$, $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ are also convex.

We wish to show that

$$\min_{\boldsymbol{z} \in \mathcal{D}} \max\big\{\Phi_1(\boldsymbol{z}), \Phi_2(\boldsymbol{z})\big\} = \max\Big\{\min_{\boldsymbol{z} \in \mathcal{D}} \Phi_1(\boldsymbol{z}), \min_{\boldsymbol{z} \in \mathcal{D}} \Phi_2(\boldsymbol{z})\Big\}. \tag{B.13}$$

We define the regions

$$\mathcal{R}_1 = \big\{\boldsymbol{z} : \Phi_1(\boldsymbol{z}) > \Phi_0(\boldsymbol{z})\big\} \tag{B.14}$$

$$\mathcal{R}_2 = \big\{\boldsymbol{z} : \Phi_2(\boldsymbol{z}) > \Phi_0(\boldsymbol{z})\big\}. \tag{B.15}$$

The key observation is that $\mathcal{R}_1$ and $\mathcal{R}_2$ are disjoint. To see this, we observe from (B.10)–(B.12) that any $\boldsymbol{z} \in \mathcal{R}_1 \cap \mathcal{R}_2$ must satisfy $g_1(\boldsymbol{z}) > g(\boldsymbol{z})$ and $g_2(\boldsymbol{z}) > g(\boldsymbol{z})$. When combined with (31), these imply $g_1(\boldsymbol{z}) < 0$ and $g_2(\boldsymbol{z}) < 0$, and it follows from (B.10)–(B.12) that $\Phi_0(\boldsymbol{z}) = \Phi_1(\boldsymbol{z}) = \Phi_2(\boldsymbol{z})$, in contradiction with the assumption that $\boldsymbol{z} \in \mathcal{R}_1 \cap \mathcal{R}_2$. Thus, $\mathcal{R}_1 \cap \mathcal{R}_2$ is empty, which implies that $g_1(\boldsymbol{z})$ and $g_2(\boldsymbol{z})$ cannot simultaneously be the unique maximizers in (B.11) and (B.12) respectively. Combining this observation with (B.10), we obtain

$$\Phi_0(\boldsymbol{z}) = \min\big\{\Phi_1(\boldsymbol{z}), \Phi_2(\boldsymbol{z})\big\}. \tag{B.16}$$

To prove (B.13), we use a proof by contradiction. Let the left-hand side and right-hand side be denoted by f^* and $\tilde{f}^*$ respectively. It is easily seen that $f^* \geq \tilde{f}^*$ always holds, so we assume that $f^* > \tilde{f}^*$. Let $\boldsymbol{z}_1^*$ and $\boldsymbol{z}_2^*$ minimize Φ_1 and Φ_2 respectively on the right-hand side of (B.13), so that

$$\tilde{f}^* = \max\big\{\Phi_1(\boldsymbol{z}_1^*), \Phi_2(\boldsymbol{z}_2^*)\big\}. \tag{B.17}$$

The assumption $f^* > \tilde{f}^*$ implies that

$$\Phi_2(\boldsymbol{z}_1^*) > \Phi_1(\boldsymbol{z}_1^*) \tag{B.18}$$

$$\Phi_1(\boldsymbol{z}_2^*) > \Phi_2(\boldsymbol{z}_2^*). \tag{B.19}$$

Next, we define

$$\Phi_0'(\lambda) \triangleq \Phi_0\big(\lambda \boldsymbol{z}_1^* + (1-\lambda)\boldsymbol{z}_2^*\big) \tag{B.20}$$

$$\Phi_1'(\lambda) \triangleq \Phi_1\big(\lambda \boldsymbol{z}_1^* + (1-\lambda)\boldsymbol{z}_2^*\big) \tag{B.21}$$

$$\Phi_2'(\lambda) \triangleq \Phi_2\big(\lambda \boldsymbol{z}_1^* + (1-\lambda)\boldsymbol{z}_2^*\big), \tag{B.22}$$

for $\lambda \in [0, 1]$. Since any convex function is also convex when restricted to a straight line [12, Section 3.1.1], it follows

that Φ_0', Φ_1' and Φ_2' are convex in λ. From (B.18)–(B.19), we have

$$\Phi_2'(1) > \Phi_1'(1) \tag{B.23}$$

$$\Phi_1'(0) > \Phi_2'(0). \tag{B.24}$$

Since Φ_1' and Φ_2' are convex, they are also continuous, and it follows that the two must intersect somewhere in $(0,1)$, say at λ^*. Therefore,

$$\Phi_0'(\lambda^*) = \min\left\{\Phi_1'(\lambda^*), \Phi_2'(\lambda^*)\right\} \tag{B.25}$$

$$= \max\left\{\Phi_1'(\lambda^*), \Phi_2'(\lambda^*)\right\} \tag{B.26}$$

$$\geq \min_{\boldsymbol{z}\in\mathcal{D}} \max\left\{\Phi_1(\boldsymbol{z}), \Phi_2(\boldsymbol{z})\right\} \tag{B.27}$$

$$= f^*, \tag{B.28}$$

where (B.25) follows from (B.16), and (B.27) follows from (B.21)–(B.22). Finally, we have the following contradiction:

- Combining (B.28) with the assumption that $f^* > \tilde{f}^*$, we have

$$\Phi_0'(\lambda^*) > \tilde{f}^* = \max\{\Phi_1'(1), \Phi_2'(0)\}. \tag{B.29}$$

- From (B.16), we have $\Phi_0'(\lambda) = \min\{\Phi_1'(\lambda), \Phi_2'(\lambda)\}$, and it follows from (B.23)–(B.24) that $\Phi_0'(1) = \Phi_1'(1)$ and $\Phi_0'(0) = \Phi_2'(0)$. Hence, using the convexity of Φ_0' and Jensen's inequality, we have

$$\Phi_0'(\lambda^*) \leq \lambda^*\Phi_1'(1) + (1-\lambda^*)\Phi_2'(0) \tag{B.30}$$

$$\leq \max\{\Phi_1'(1), \Phi_2'(0)\}. \tag{B.31}$$

APPENDIX C
MULTIPLE-ACCESS CHANNEL PROOFS

A. Proof of Theorem 4

Each of the four expressions are derived using a similar analysis, so we focus on $E_{r,12,1}^{\mathrm{cc}}$. We claim that the expression in (76) is equivalent to

$$E_{r,12,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) = \max_{\rho_1\in[0,1],\rho_2\in[0,1]} \hat{E}_{0,12,1}^{\mathrm{cc}}(\boldsymbol{Q}, \rho_1, \rho_2) - \rho_1(R_1 + \rho_2 R_2), \tag{C.1}$$

where

$$\hat{E}_{0,12,1}^{\mathrm{cc}}(\boldsymbol{Q}, \rho_1, \rho_2) \triangleq \min_{P_{X_1 X_2 Y}\in\mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1 X_2 Y}\in\mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1 X_2 Y})}$$
$$D(P_{X_1 X_2 Y}\|Q_1\times Q_2\times W) + \rho_1\big(I_{\widetilde{P}}(X_1;Y) + \rho_2 I_{\widetilde{P}}(X_2;X_1,Y)\big). \tag{C.2}$$

This follows by using (111) to write

$$E_{r,12,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) = \min_{P_{X_1 X_2 Y}\in\mathcal{S}^{\mathrm{cc}}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1 X_2 Y}\in\mathcal{T}_{12}^{\mathrm{cc}}(P_{X_1 X_2 Y})}$$
$$D(P_{X_1 X_2 Y}\|Q_1\times Q_2\times W) + \Big[I_{\widetilde{P}}(X_1;Y) + \big[I_{\widetilde{P}}(X_2;X_1,Y) - R_2\big]^+ - R_1\Big]^+ \tag{C.3}$$

and using a similar argument to [19, Sec. IV] as follows. We first apply $[\alpha]^+ = \max_{0\leq\rho_1\leq1}\rho_1\alpha$ to the outer $[\cdot]^+$ quantity in (C.3). The resulting objective is linear in ρ_1 and jointly convex in $(P_{X_1 X_2 Y}, \widetilde{P}_{X_1 X_2 Y})$, so we can apply Fan's minimax theorem [30] to interchange the maximization and minimization. Applying the same argument to the remaining $[\cdot]^+$ term yields (C.1)–(C.2).

We define the sets

$$\mathcal{V}_1(P_{X_1X_2Y}, \hat{P}_{X_1Y}) \triangleq \Bigg\{ \widetilde{P}_{X_1X_2Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \widetilde{P}_{X_2} = P_{X_2},$$

$$\widetilde{P}_{X_1Y} = \hat{P}_{X_1Y}, \mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \Bigg\} \tag{C.4}$$

$$\mathcal{V}_2(P_{X_1X_2Y}) \triangleq \Big\{ \hat{P}_{X_1Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{Y}) : \hat{P}_{X_1} = P_{X_1}, \hat{P}_Y = P_Y \Big\}. \tag{C.5}$$

It follows that $\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}^{cc}(P_{X_1X_2Y})$ (see (51)) if and only if $\widetilde{P}_{X_1X_2Y} \in \mathcal{V}_1(P_{X_1X_2Y}, \hat{P}_{X_1Y})$ for some $\hat{P}_{X_1Y} \in \mathcal{V}_2(P_{X_1X_2Y})$. We can therefore replace the minimization over $\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}^{cc}(P_{X_1X_2Y})$ in (C.2) with minimizations over $\hat{P}_{X_1Y} \in \mathcal{V}_2(P_{X_1X_2Y})$ and $\widetilde{P}_{X_1X_2Y} \in \mathcal{V}_1(P_{X_1X_2Y}, \hat{P}_{X_1Y})$.

We prove the theorem by performing the minimization in several steps, and performing multiple applications of Lemma 5. Each such application will yield an overall optimization of the form $\sup \min \sup\{\cdot\}$, and we will implicitly use Fan's minimax theorem [30] to obtain an equivalent expression of the form $\sup \sup \min\{\cdot\}$. Thus, we will leave the optimization of the dual variables until the final step.

Step 1: We first consider the minimization of the term $I_{\widetilde{P}}(X_1; X_2, Y)$ over $\widetilde{P}_{X_1X_2Y}$ when $P_{X_1X_2Y} \in \mathcal{S}^{cc}(\boldsymbol{Q})$ and $\hat{P}_{X_1Y} \in \mathcal{V}_2(P_{X_1X_2Y})$ are fixed, and thus all of the terms in the objective in (C.2) other than $I_{\widetilde{P}}(X_1; X_2, Y)$ are fixed. The minimization is given by

$$F_1 \triangleq \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{V}_1(P_{X_1X_2Y}, \hat{P}_{X_1Y})} I_{\widetilde{P}}(X_1; X_2, Y). \tag{C.6}$$

Applying Lemma 5 with $P_{Z_1} = P_{X_2}$, $P_{Z_2} = \hat{P}_{X_1Y}$ and $\mu_1(\cdot) = a_2(\cdot)$, we obtain the dual expression

$$F_1 = -\sum_{x_1, y} \hat{P}_{X_1Y}(x_1, y) \log \sum_{\overline{x}_2} P_{X_2}(\overline{x}_2) q(x_1, \overline{x}_2, y)^s e^{a_2(\overline{x}_2)}$$

$$+ s \sum_{x_1, x_2, y} P_{X_1X_2Y}(x_1, x_2, y) \log q(x_1, x_2, y) + \sum_{x_2} P_{X_2}(x_2) a_2(x_2). \tag{C.7}$$

Step 2: After Step 1, the overall objective (see (C.2)) is given by

$$D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times W) + \rho_1 \big(I_{\hat{P}}(X_1; Y) + \rho_2 F_1 \big), \tag{C.8}$$

where we have replaced $I_{\widetilde{P}}(X_1; Y)$ by $I_{\hat{P}}(X_1; Y)$ due to the constraint $\widetilde{P}_{X_1Y} = \hat{P}_{X_1Y}$ in (C.4). Since the only terms involving $\hat{P}_{X_1Y}$ are $I_{\hat{P}}(X_1; Y)$ and the first term in the summation of (C.7), we consider the minimization

$$F_2 \triangleq \min_{\hat{P}_{X_1Y} \in \mathcal{V}_2(P_{X_1X_2Y})} I_{\hat{P}}(X_1; Y) - \rho_2 \sum_{x_1, y} \hat{P}_{X_1Y}(x_1, y) \log \sum_{\overline{x}_2} P_{X_2}(\overline{x}_2) q(x_1, \overline{x}_2, y)^s e^{a_2(\overline{x}_2)}. \tag{C.9}$$

Applying Lemma 5 with $P_{Z_1} = P_{X_1}$, $P_{Z_2} = P_Y$ and $\mu_1(\cdot) = a_1(\cdot)$, we obtain

$$F_2 = -\sum_y P_Y(y) \log \sum_{\overline{x}_1} P_{X_1}(\overline{x}_1) \left(\sum_{\overline{x}_2} P_{X_2}(\overline{x}_2) q(\overline{x}_1, \overline{x}_2, y)^s e^{a_2(\overline{x}_2)} \right)^{\rho_2} e^{a_1(\overline{x}_1)} + \sum_{x_1} P_{X_1}(x_1) a_1(x_1). \tag{C.10}$$

Step 3: From (C.7), (C.8) and (C.10), the overall objective is now given by

$$F_3 \triangleq D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times W) + \rho_1 F_2 + \rho_1 \rho_2 \sum_{x_1, x_2, y} P_{X_1X_2Y}(x_1, x_2, y) \log q(x_1, x_2, y)^s e^{a_2(x_2)}. \tag{C.11}$$

By expanding the divergence and writing $a_\nu(x_\nu) = \log e^{a_\nu(x_\nu)}$ ($\nu = 1, 2$), (C.11) can equivalently be expressed as

$$F_3 = -\sum_{x_1, x_2, y} P_{X_1X_2Y}(x_1, x_2, y) \log \frac{g_3(x_1, x_2, y)}{P_{X_1X_2Y}(x_1, x_2, y)}, \tag{C.12}$$

where

$$g_3(x_1, x_2, y) \triangleq Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)\left(\sum_{\overline{x}_1} P_{X_1}(\overline{x}_1)\left(\frac{\sum_{\overline{x}_2} P_{X_2}(\overline{x}_2)q(\overline{x}_1, \overline{x}_2, y)^s e^{a_2(\overline{x}_2)}}{q(x_1, x_2, y)^s e^{a_2(x_2)}}\right)^{\rho_2}\frac{e^{a_1(\overline{x}_1)}}{e^{a_1(x_1)}}\right)^{\rho_1}.$$

(C.13)

From (C.2), we must minimize F_3 over $P_{X_1 X_2 Y} \in \mathcal{S}(\boldsymbol{Q})$. We first minimize over $P_{Y|X_1 X_2}$ for a fixed $P_{X_1 X_2}$. Rather than using Lagrange duality, we can simply write

$$P_{X_1 X_2 Y}(x_1, x_2, y) = P_{X_1 X_2}(x_1, x_2)P_{Y|X_1 X_2}(y|x_1, x_2)$$

(C.14)

and apply Jensen's inequality to (C.12) to obtain

$$F_3 \geq -\sum_{x_1, x_2, y} P_{X_1 X_2}(x_1, x_2)\log\sum_y \frac{g_3(x_1, x_2, y)}{P_{X_1 X_2}(x_1, x_2)}.$$

(C.15)

It follows that the minimizing $P_{Y|X_1 X_2}$ is the one which makes $\frac{g_3(x_1, x_2, y)}{P_{X_1 X_2 Y}(x_1, x_2, y)}$ independent of y for any given x_1 and x_2, so that equality holds in (C.15).

Step 4: Using (C.15) and the definition of g_3, we can write the overall objective as

$$F_4 \triangleq I_P(X_1; X_2) - \mathbb{E}_P[\log g_4(X_1, X_2)],$$

(C.16)

where

$$g_4(x_1, x_2) \triangleq \sum_y W(y|x_1, x_2)\left(\sum_{\overline{x}_1} P_{X_1}(\overline{x}_1)\left(\frac{\sum_{\overline{x}_2} P_{X_2}(\overline{x}_2)q(\overline{x}_1, \overline{x}_2, y)^s e^{a_2(\overline{x}_2)}}{q(x_1, x_2, y)^s e^{a_2(x_2)}}\right)^{\rho_2}\frac{e^{a_1(\overline{x}_1)}}{e^{a_1(x_1)}}\right)^{\rho_1}.$$

(C.17)

Applying Lemma 5 with $P_{Z_1} = Q_2$, $P_{Z_2} = Q_1$ and $\mu_1(\cdot) = a_2'(\cdot)$, we obtain

$$F_4 = -\sum_{x_1} Q_1(x_1)\log\sum_{x_2} Q_2(x_2)g_4(x_1, x_2)e^{a_2'(x_2)} + \sum_{x_2} Q_2(x_2)a_2'(x_2).$$

(C.18)

Step 5: From (C.18), we have

$$L_4 = -\sum_{x_1} Q_1(x_1)\log\left(e^{a_1'(x_1) - \phi_1'}\sum_{x_2} Q_2(x_2)g_4(x_1, x_2)e^{a_2'(x_2) - \phi_2'}\right)$$

(C.19)

$$\geq -\log\sum_{x_1, x_2} Q_1(x_1)Q_2(x_2)g_4(x_1, x_2)e^{a_2'(x_2) - \phi_2'}e^{a_1'(x_1) - \phi_1'},$$

(C.20)

where (C.19) follows for any $a_1'(\cdot)$ with $\phi_\nu' \triangleq \mathbb{E}_{Q_\nu}[a_\nu'(X_\nu)]$ ($\nu = 1, 2$) by expanding the logarithm, and (C.20) follows from Jensen's inequality. By choosing $a_1'(\cdot)$ to make the argument to the logarithm in (C.19) independent of x_1, we obtain equality in (C.20). Thus, the proof is concluded by substituting the expression for g_4 in (C.17) into (C.20), and taking the supremum over $s \geq 0$, $a_1(\cdot)$, $a_2(\cdot)$, $a_1'(\cdot)$ and $a_2'(\cdot)$.

APPENDIX D
REFINED SUPERPOSITION CODING PROOFS

A. Proof of Theorem 18

We focus on showing the equivalence of (224) and (230), since identical arguments apply to (225) and (231). The primal expression is written in terms of a minimization over $\widetilde{P}_{UXY}$. It is convenient to split this distribution into three distributions: $\widetilde{P}_{UY}$, $\hat{P}_{XY} \triangleq \widetilde{P}_{XY|U=1}$ and $\hat{\hat{P}}_{XY} \triangleq \widetilde{P}_{XY|U=2}$. Using an identical argument to the start of Appendix C, we can write the right-hand side of (224) as

$$\sup_{\rho_1 \in [0,1], \rho_2 \in [0,1]} \min_{\widetilde{P}_{UY}, \hat{P}_{XY}, \hat{\hat{P}}_{XY}} I_{\widetilde{P}}(U; Y) + \rho_1 Q_U(1)I_{\hat{P}}(X; Y) + \rho_1\rho_2 Q_U(2)I_{\hat{\hat{P}}}(X; Y) - \rho_1 R_{11} - \rho_1\rho_2 R_{12}.$$

(D.1)

Defining $P_{UXY} = Q_{UX} \times W$, the minimization is subject to the constraints (i) $\widetilde{P}_U = Q_U$, (ii) $\hat{P}_X = Q_{X|U=1}$, (iii) $\hat{P}_X = Q_{X|U=2}$, (iv) $\widetilde{P}_Y = P_Y$, (v) $\hat{P}_Y = \widetilde{P}_{Y|U=1}$, (vi) $\hat{P}_Y = \widetilde{P}_{Y|U=2}$, (vii) $Q_U(1)\mathbb{E}_{\hat{P}}[\log q(X,Y)] + Q_U(2)\mathbb{E}_{\hat{P}}[\log q(X,Y)] \geq \mathbb{E}_P[\log q(X,Y)]$.

Similarly to Appendix C, we apply the minimization in several steps, using Fan's minimax theorem [30] after each step so that the supremum over the dual variables can be left until the end. To avoid repetition, we provide less detail than the amount given in Appendix C.

Step 1: For given joint distributions $\widetilde{P}_{UY}$ and $\hat{P}_{XY}$, the minimization $F_1 \triangleq \min_{\hat{P}_{XY}} I_{\hat{P}}(X;Y)$ subject to the constraints (iii), (vi) and (vii) has a dual expression given by

$$F_1 = -F_{1,1} + F_{1,2} + F_{1,3} - sQ_U(1)F_{1,4}, \tag{D.2}$$

where

$$F_{1,1} \triangleq \sum_y \widetilde{P}_{Y|U}(y|2) \log \sum_{\overline{x}} Q_2(\overline{x}) q(\overline{x}, y)^{sQ_U(2)} e^{a_2(\overline{x})} \tag{D.3}$$

$$F_{1,2} \triangleq \sum_{x_2} Q_2(x) a_2(x) \tag{D.4}$$

$$F_{1,3} \triangleq s \sum_{x,y} P_{XY}(x,y) \log q(x,y) \tag{D.5}$$

$$F_{1,4} \triangleq \sum_{x,y} \hat{P}_{XY}(x,y) \log q(x,y), \tag{D.6}$$

and where $s \geq 0$ and $a_2(\cdot)$ are dual variables.

Step 2: For a given joint distribution $\widetilde{P}_{UY}$, the minimization $F_2 \triangleq \min_{\hat{P}_{XY}} I_{\hat{P}}(X;Y) - s\rho_2 Q_U(2)F_{1,4}(\hat{P}_{XY})$ subject to (ii) and (v) has a dual expression given by

$$F_2 = F_{2,1} - F_{2,2}, \tag{D.7}$$

where

$$F_{2,1} \triangleq \sum_x Q_1(x) a_1(x) \tag{D.8}$$

$$F_{2,2} \triangleq \sum_y \widetilde{P}_{Y|U}(y|1) \log \sum_{\overline{x}} Q_1(\overline{x}) q(\overline{x}, y)^{s\rho_2 Q_U(2)} e^{a_1(\overline{x})}, \tag{D.9}$$

and where $a_1(\cdot)$ is a dual variable.

Step 3: The minimization $F_3 \triangleq \min_{\widetilde{P}_{UY}} I_{\widetilde{P}}(U;Y) - \rho_1 Q_U(1)F_{2,2} - \rho_1 \rho_2 Q_U(2)F_{1,1}$ subject to (i) and (iv) can equivalently be expressed as

$$F_3 = I_{\widetilde{P}}(U;Y) - \sum_u \rho'(u) \sum_y \widetilde{P}_{UY}(u,y) \log \sum_{\overline{x}} Q_u(\overline{x}) q(\overline{x}, y)^{s'(u)} e^{a(u,\overline{x})} \tag{D.10}$$

using the definitions in (232) along with $a(u,x) = a_u(x)$. The dual expression is thus given by

$$F_3 = -\sum_y P_Y(y) \log \sum_{\overline{u}} Q_U(\overline{u}) \left(\sum_{\overline{x}} Q_{\overline{u}}(\overline{x}) q(\overline{x}, y)^{s'(\overline{u})} e^{a(\overline{u},\overline{x})} \right)^{\rho'(u)} e^{b(\overline{u})} + \sum_u Q_U(u) b(u), \tag{D.11}$$

where $b(\cdot)$ is a dual variable.

Step 4: The final objective is given by $F_3 + \rho_1 Q_U(1)F_{2,1} + \rho_1\rho_2 Q_U(2)\big(F_{1,2} + F_{1,3}\big)$. Applying some algebraic manipulations yields the dual expression

$$
\sum_{u,x,y} P_{UXY}(u,x,y) \log \frac{\left(q(x,y)^{s'(u)} e^{a(u,x)}\right)^{\rho'(u)} e^{b(u)}}{\sum_{\overline{u}} Q_U(\overline{u})\left(\sum_{\overline{x}} Q_{\overline{u}}(\overline{x}) q(\overline{x},y)^{s'(\overline{u})} e^{a(\overline{u},\overline{x})}\right)^{\rho'(\overline{u})} e^{b(\overline{u})}}.
\tag{D.12}
$$

The variable $b(u)$ can be factored into $a(u,x)$ without affecting the final dual optimization, and can thus be removed from the numerator and denominator in (D.12). This concludes the proof.

B. Proof of Theorem 19

We focus on the derivation of (230), since (231) can be derived similarly. The ideas used in the derivation are similar to those for the MAC (see the proof of Theorem 6), but the details are more involved.

Applying Lemma 1 to the union in (206), with $Z_1(i) = \boldsymbol{X}_1^{(1,i)}$ and $Z_2(j) = \boldsymbol{X}_2^{(1,j)}$, we obtain the non-asymptotic RCU bound

$$
\overline{p}_{e,0} \leq \mathbb{E}\left[\min\left\{1, (M_0-1)\mathbb{E}\left[\min\left\{1, M_{11}\mathbb{E}\left[\min\left\{1, M_{12}\mathbb{P}\left[\frac{q^n(\overline{\boldsymbol{X}},\boldsymbol{Y})}{q^n(\boldsymbol{X},\boldsymbol{Y})} \geq 1 \,\Big|\, \overline{\boldsymbol{X}}_1\right]\right\} \,\Big|\, \overline{\boldsymbol{U}}\right]\right\} \,\Big|\, \boldsymbol{U},\boldsymbol{X},\boldsymbol{Y}\right]\right\}\right].
\tag{D.13}
$$

Using (199), Markov's inequality, and $\min\{1,\alpha\} \leq \alpha^\rho$ ($\rho \in [0,1]$), we obtain[3]

$$
\overline{p}_{e,0} \leq (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1\rho_2})^{\rho_0} \sum_{\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2} P_{\boldsymbol{U}}(\boldsymbol{u}) P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) \sum_{\boldsymbol{y}} W^n(\boldsymbol{y}|\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2))\left(\sum_{\overline{\boldsymbol{u}}} P_{\boldsymbol{U}}(\overline{\boldsymbol{u}})\right.
$$
$$
\left. \times \left(\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1)\left(\frac{q^{n_1}\big(\overline{\boldsymbol{x}}_1,\boldsymbol{y}_1(\overline{\boldsymbol{u}})\big)}{q^{n_1}\big(\boldsymbol{x}_1,\boldsymbol{y}_1(\boldsymbol{u})\big)}\right)^{\rho_2 s}\right)^{\rho_1}\left(\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)\left(\frac{q^{n_2}\big(\overline{\boldsymbol{x}}_2,\boldsymbol{y}_2(\overline{\boldsymbol{u}})\big)}{q^{n_2}\big(\boldsymbol{x}_2,\boldsymbol{y}_2(\boldsymbol{u})\big)}\right)^s\right)^{\rho_1\rho_2}\right)^{\rho_0},
\tag{D.14}
$$

where $s \geq 0$ and $\rho_1, \rho_2 \in [0,1]$ are arbitrary. Using the definition of the cost-constrained ensemble in (234)–(236), we obtain

$$
\overline{p}_{e,0} \;\dot{\leq}\; (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1\rho_2})^{\rho_0} \sum_{\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2} P_{\boldsymbol{U}}(\boldsymbol{u}) P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) \sum_{\boldsymbol{y}} W^n(\boldsymbol{y}|\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2))\left(\sum_{\overline{\boldsymbol{u}}} P_{\boldsymbol{U}}(\overline{\boldsymbol{u}})\right.
$$
$$
\left. \times \left(\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1)\left(\frac{q^{n_1}\big(\overline{\boldsymbol{x}}_1,\boldsymbol{y}_1(\overline{\boldsymbol{u}})\big)}{q^{n_1}\big(\boldsymbol{x}_1,\boldsymbol{y}_1(\boldsymbol{u})\big)}\right)^{\rho_2 s} \frac{e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)}}{e^{a_1^{n_1}(\boldsymbol{x}_1)}}\right)^{\rho_1}\left(\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)\left(\frac{q^{n_2}\big(\overline{\boldsymbol{x}}_2,\boldsymbol{y}_2(\overline{\boldsymbol{u}})\big)}{q^{n_2}\big(\boldsymbol{x}_2,\boldsymbol{y}_2(\boldsymbol{u})\big)}\right)^s \frac{e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)}}{e^{a_2^{n_2}(\boldsymbol{x}_2)}}\right)^{\rho_1\rho_2}\right)^{\rho_0},
\tag{D.15}
$$

where for $u = 1,2$, $a_u(\cdot)$ is one of the $L_u = 2$ cost functions in (235), and $a_u^{n_u}(\boldsymbol{x}_u) \triangleq \sum_{i=1}^{n_u} a_u(x_{u,i})$. For each $(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y})$, we write the argument to the summation over $\boldsymbol{y}$ in (D.15) as a product of two terms, namely

$$
T_1 \triangleq W^n(\boldsymbol{y}|\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2)) q^{n_1}(\boldsymbol{x}_1,\boldsymbol{y}_1(\boldsymbol{u}))^{-\rho_1\rho_2 s} e^{-\rho_1 a_1^{n_1}(\boldsymbol{x}_1)} q^{n_2}(\boldsymbol{x}_2,\boldsymbol{y}_2(\boldsymbol{u}))^{-\rho_1\rho_2 s} e^{-\rho_1\rho_2 a_2^{n_2}(\boldsymbol{x}_2)}
\tag{D.16}
$$
$$
T_2 \triangleq \left(\sum_{\overline{\boldsymbol{u}}} P_{\boldsymbol{U}}(\overline{\boldsymbol{u}})\left(\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) q^{n_1}(\overline{\boldsymbol{x}}_1,\boldsymbol{y}_1(\overline{\boldsymbol{u}}))^{\rho_2 s} e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)}\right)^{\rho_1}\left(\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) q^{n_2}(\overline{\boldsymbol{x}}_2,\boldsymbol{y}_2(\overline{\boldsymbol{u}}))^s e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)}\right)^{\rho_1\rho_2}\right)^{\rho_0}.
\tag{D.17}
$$

---

[3]In the case of continuous alphabets, the summations should be replaced by integrals.

Since $P_{\boldsymbol{X}_u}$ is upper bounded by a subexponential prefactor times Q^{n_u} for $u = 1, 2$ (see Proposition 1), we have

$$\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) q^{n_1}(\overline{\boldsymbol{x}}_1, \boldsymbol{y}_1(\overline{\boldsymbol{u}}))^{\rho_2 s} e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)} \overset{\cdot}{\leq} \prod_{i=1}^{n_1} \sum_{\overline{x}_1} Q_1(\overline{x}_1) q(\overline{x}_1, y_{1,i}(\overline{\boldsymbol{u}}))^{\rho_2 s} e^{a_1(\overline{x}_1)} \tag{D.18}$$

$$\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) q^{n_2}(\overline{\boldsymbol{x}}_2, \boldsymbol{y}_2(\overline{\boldsymbol{u}}))^s e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)} \overset{\cdot}{\leq} \prod_{i=1}^{n_2} \sum_{\overline{x}_2} Q_2(\overline{x}_2) q(\overline{x}_2, y_{2,i}(\overline{\boldsymbol{u}}))^s e^{a_2(\overline{x}_2)}, \tag{D.19}$$

where for $u = 1, 2$, $y_{u,i}(\overline{\boldsymbol{u}})$ is the i-th entry of $\boldsymbol{y}_u(\overline{\boldsymbol{u}})$. Using the definitions in (232) along with $a(u, x) \triangleq a_u(x)$, we thus obtain

$$\left(\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) q^{n_1}(\overline{\boldsymbol{x}}_1, \boldsymbol{y}_1(\overline{\boldsymbol{u}}))^{\rho_2 s} e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)} \right)^{\rho_1} \left(\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) q^{n_2}(\overline{\boldsymbol{x}}_2, \boldsymbol{y}_2(\overline{\boldsymbol{u}}))^s e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)} \right)^{\rho_1 \rho_2} \tag{D.20}$$

$$\overset{\cdot}{\leq} \left(\prod_{i=1}^{n_1} \sum_{\overline{x}_1} Q_1(\overline{x}_1) q(\overline{x}_1, y_{1,i}(\overline{\boldsymbol{u}}))^{\rho_2 s} e^{a_1(\overline{x}_1)} \right)^{\rho_1} \left(\prod_{i=1}^{n_2} \sum_{\overline{x}_2} Q_2(\overline{x}_2) q(\overline{x}_2, y_{2,i}(\overline{\boldsymbol{u}}))^s e^{a_2(\overline{x}_2)} \right)^{\rho_1 \rho_2} \tag{D.21}$$

$$= \prod_{i=1}^{n} \left(\sum_{\overline{x}} Q(\overline{x}|\overline{u}_i) q(\overline{x}, y_i)^{s'(\overline{u}_i)} e^{a(\overline{u}_i, \overline{x})} \right)^{\rho'(\overline{u}_i)}. \tag{D.22}$$

Hence, and using the fact that $P_{\boldsymbol{U}}(\boldsymbol{u}) \overset{\cdot}{\leq} Q_U^n(\boldsymbol{u})$ (see (A.7)), we obtain

$$T_2 \overset{\cdot}{\leq} \prod_{i=1}^{n} \left(\sum_{\overline{u}} Q_U(\overline{u}) \left(\sum_{\overline{x}} Q(\overline{x}|\overline{u}_i) q(\overline{x}, y_i)^{s'(\overline{u}_i)} e^{a(\overline{u}_i, \overline{x})} \right)^{\rho'(\overline{u}_i)} \right)^{\rho_0}. \tag{D.23}$$

A similar argument (without the need for the $\overset{\cdot}{\leq}$ steps) gives

$$T_1 = \prod_{i=1}^{n} \sum_{y} W(y|x_i) q(x_i, y)^{-\rho'(u_i) s'(u_i)} e^{-\rho'(u_i) a(u_i, x_i)}, \tag{D.24}$$

where we have used the fact that $W^n(\boldsymbol{y}|\Xi(\boldsymbol{u}, \boldsymbol{x}_1, \boldsymbol{x}_2)) = W^{n_1}(\boldsymbol{y}_1(\boldsymbol{u})|\boldsymbol{x}_1) W^{n_2}(\boldsymbol{y}_2(\boldsymbol{u})|\boldsymbol{x}_2)$. Substituting (D.23) and (D.24) into (D.15), we obtain

$$\overline{p}_{e,0} \overset{\cdot}{\leq} (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1 \rho_2})^{\rho_0}$$

$$\times \sum_{\boldsymbol{u}, \boldsymbol{x}} P_{\boldsymbol{UX}}(\boldsymbol{u}, \boldsymbol{x}) \prod_{i=1}^{n} \sum_{y} W(y|x_i) \left(\sum_{\overline{u}} Q_U(\overline{u}) \left(\sum_{\overline{x}} Q(\overline{x}|\overline{u}_i) \left(\frac{q(\overline{x}, y_i)}{q(x_i, y_i)} \right)^{s'(\overline{u}_i)} \frac{e^{a(\overline{u}_i, \overline{x})}}{e^{a(u_i, x_i)}} \right)^{\rho'(\overline{u}_i)} \right)^{\rho_0}, \tag{D.25}$$

where

$$P_{\boldsymbol{UX}}(\boldsymbol{u}, \boldsymbol{x}) \triangleq \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2} P_{\boldsymbol{U}}(\boldsymbol{u}) P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) \mathbb{1}\{\boldsymbol{x} = \Xi(\boldsymbol{u}, \boldsymbol{x}_1, \boldsymbol{x}_2)\} \tag{D.26}$$

If $P_{\boldsymbol{UX}}$ were i.i.d. on Q_{UX}, then (D.25) would yield an error exponent which is positive when (230) holds with strict inequality, by taking $\rho_0 \to 0$ similarly to Theorem 5. The same can be done in the present setting by upper bounding $P_{\boldsymbol{UX}}$ by a subexponential prefactor times Q_{UX}^n, analogously to (D.18)–(D.19). More precisely, we have

$$P_{\boldsymbol{UX}}(\boldsymbol{u}, \boldsymbol{x}) \overset{\cdot}{\leq} \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2} P_{\boldsymbol{U}}(\boldsymbol{u}) Q_{X|U=1}^{n_1}(\boldsymbol{x}_1) Q_{X|U=2}^{n_2}(\boldsymbol{x}_2) \mathbb{1}\{\boldsymbol{x} = \Xi(\boldsymbol{u}, \boldsymbol{x}_1, \boldsymbol{x}_2)\} \tag{D.27}$$

$$= P_{\boldsymbol{U}}(\boldsymbol{u}) Q_{X|U}^n(\boldsymbol{x}|\boldsymbol{u}) \tag{D.28}$$

$$\overset{\cdot}{\leq} Q_U^n(\boldsymbol{u}) Q_{X|U}^n(\boldsymbol{x}|\boldsymbol{u}) \tag{D.29}$$

$$= Q_{UX}^n(\boldsymbol{u}, \boldsymbol{x}). \tag{D.30}$$

This concludes the proof.

REFERENCES

[1] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, MIT, 1983.

[2] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.

[3] 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.

[4] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[5] V. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1889–1902, Nov. 1995.

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

[7] 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.

[8] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," submitted to *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1303.6166].

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

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

[11] R. Gallager, "Fixed composition arguments and lower bounds to error probability," http://web.mit.edu/gallager/www/notes/notes5.pdf.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[13] A. Somekh-Baruch, "A general formula for the mismatch capacity," http://arxiv.org/abs/1309.7964.

[14] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *Arch. Elek. Über.*, vol. 47, no. 4, pp. 228–239, 1993.

[15] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[16] A. Dyachkov, "Random constant composition codes for multiple access channels," *Prob. Contr. Inf. Theory*, vol. 13, pp. 357–369, 1984.

[17] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, March 1985.

[18] J. Pokorny and H. Wallmeier, "Random coding bound and codes produced by permutations for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 741–750, Nov. 1985.

[19] Y. Liu and B. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, March 1996.

[20] A. Nazari, A. Anastasopoulos, and S. S. Pradhan, "Error exponent for multiple-access channels: Lower bounds," arXiv:1010.1303v1 [cs.IT].

[21] D. de Caen, "A lower bound on the probability of a union," *Discrete Mathematics*, vol. 169, pp. 217–220, 1997.

[22] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Ensemble-tight error exponents for mismatched decoders," in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.

[23] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel Aviv University, 2003.

[24] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[25] H. Yagi, "Finite blocklength bounds for multiple access channels with correlated sources," in *Int. Symp. Inf. Theory and Apps.*, 2012, pp. 377–381.

[26] P. Elias, "Coding for two noisy channels," in *Third London Symp. Inf. Theory*, 1955.

[27] A. G. D'yachkov, "Bounds on the average error probability for a code ensemble with fixed composition," *Prob. Inf. Transm.*, vol. 16, no. 4, pp. 3–8, 1980.

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

[29] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004.

[30] K. Fan, "Minimax theorems," *Proc. Nat. Acad. Sci.*, vol. 39, pp. 42–47, 1953.

[31] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[32] G. Poltyrev, "Random coding bounds for some broadcast channels," *Prob. Peredachi Inf.*, vol. 19, no. 1, pp. 9–20, 1983.

[33] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[34] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements and connections," submitted to *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1307.6679].

[35] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Cost-constrained random coding and applications," in *Inf. Theory and Apps. Workshop*, San Diego, CA, Feb. 2013.

[36] R. Gallager, "Capacity and coding for degraded broadcast channels," *Prob. Peredachi Inf.*, vol. 10, no. 3, pp. 3–14, 1974.

[37] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, Jan. 1977.

[38] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 670–679, Nov. 1980.

[39] J. Scarlett and V. Y. F. Tan, "Second-order asymptotics for the Gaussian MAC with degraded message sets," submitted to *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1310.1197].

[40] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, July and Oct. 1948.

[41] M. S. Pinsker and A. Sheverdjaev, "Zero error capacity with erasure," *Prob. Inf. Transm.*, vol. 6, no. 1, pp. 20–24, 1970.

[42] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[43] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 55–62, 1996.

[44] V. Y. F. Tan, "Error exponents for the relay channel," http://arxiv.org/abs/1304.3553.

[45] G. Poltyrev, "Random coding bounds for discrete memoryless channels," *Prob. Inf. Transm.*, vol. 18, no. 1, pp. 9–21, 1982.

[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2001.